Deepika Singh

Machine Learning based Smart Home System in Ambient Assisted Living

DOCTORAL THESIS

to achieve the university degree of

Doktor der technischen Wissenchaften (Dr. techn.)

submitted to

Graz University of Technology

THESIS SUPERVISOR
Prof. Dr. Andreas Holzinger
Graz University of Technology

Graz, Austria, June 2021

Abstract

During the past years, there has been rapid development in Ambient Assisted Living technologies and Smart Home solutions for the growing aging population. A smart home system consists of many different components that interact with each other and can provide services that are context-aware and personalized. In order to develop a flexible, scalable, and acceptable smart home solution, certain improvements are required in different areas such as meeting end-user requirements, activity recognition, privacy aspect, and interaction with the system. In this thesis, we explored different ways to further improve the performances of the major components of a smart home system that includes activity recognition, privacy-preserving, and dialogue systems. The research in this thesis has been conducted in five main areas. First, we proposed a privacy-enabled smart home framework in an open-source software platform. Next, we performed different user studies and interviews with the older people and end-users to understand their needs better, their perception and attitudes towards smart home applications such as activity monitoring, data sharing, privacy notions, and views regarding voice assistants. Third, we investigated deep learning algorithms on smart home sensor datasets for both single and multiple resident activity recognition. Using deep learning models, we achieved promising results on raw sensor datasets, in comparison to traditional machine learning methods. Furthermore, we investigated various class imbalance techniques on smart home datasets to address the class imbalance problem. Fourth, we proposed and implemented a privacy-preserving mechanism for data transformation and anonymization that allows sharing of encoded data instead of original data with third parties and thus, protect users' sensitive information. In the last part of this thesis, we provided an overview of the dialogue system and presented a new human-annotated dialogue dataset with a benchmark metrics using deep learning models to evaluate the quality of dialogue replies for the given context. Overall, the thesis addresses the challenges in existing research and components of smart home system and the results obtained will potentially help developers and service providers in designing practical and adaptable smart home solutions.

Keywords: Ambient Assited Living, Smart Homes, Activity Recognition, Privacy Preservation, Dialogue Systems, Deep Learning

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the used sources.

Date	Signature

Acknowledgements

The work presented in this doctoral thesis has received funding from two sources. From September 2016 to October 2019, the doctoral research was funded by the European Commission's 7th Framework Programme FP7-PEOPLE Marie Curie Actions Initial Training Network ACROSSING project, under grant agreement 676157. The work was principally carried out at Biomedical Systems, Center for Health and Bioresources, AIT Austrian Institute of Technology GmbH (Wiener Neustadt, Austria), and two short stays abroad. From November 2019 to December 2020, the work was carried out at Institut für Medizinische Informatik, Statistik und Dokumentation, Medical University Graz (Graz, Austria). During the course of over four years, I had this unique opportunity to work under the supervision of great experts from both industry and academia.

First of all, I would like to express deep gratitude to my supervisor, Prof. Andreas Holzinger, for his patient guidance, enthusiastic encouragement, and useful critiques of my research work. His continuous encouragement kept my motivation high and provided me the right guidance to sail through this journey. His enthusiasm and energy towards research have always been a great source of inspiration for me. Besides my supervisor, I would like to thank Dr. Johannes Kropf and Dr. Sten Hanke for their supervision during my stay at AIT; their expertise in the field of AAL provided me the right path to start my research work; furthermore, their advice and assistance kept my progress on schedule. I must also acknowledge the extended support provided in terms of computational resources even after leaving AIT.

I would like to thank all my fellow Early Stage Researchers of the project for having great collaborations and exchange of knowledge throughout the project. I would also like to thank my colleagues at AIT for being so welcoming and for making my time memorable. Thanks to Prof. Liming Chen and Easter Veera, my short stays at DMU (Leicester, UK) and Zuyderland (Sittard, Netherlands) respectively have been extremely successful.

Last but not least, I am extremely indebted to my parents for their unconditional love and belief in me. Their support and encouragement are beyond words. I can not imagine finishing this work without the immense support, love, and motivation from my husband, Jitendra. Having already earned his Ph.D. three years ago, he understood my ups and downs and was always been, my mentor. Lastly, this thesis would have not been possible without the support of my adorable son Devaditya (Adi). He will be 3.5 months old at the time of defense and I feel so proud and blessed to have him in my Ph.D. journey.

Contents

Abbreviations

AAL	**A**mbient **A**sisted **L**iving
AI	**A**rtificial **I**ntelligence
AMT	**A**mazon **M**echanical **T**urk
ASR	**A**utomatic **S**peech **R**ecognizer
BERT	**B**idirectional **E**ncoder **R**epresentations
from **T**ransformers	
CNN	**C**onvolutional **N**eural **N**etwork
CRF	**C**onditional **R**andom **F**ield
DM	**D**ialog **M**anager
DRS	**D**ialog **R**esponse **S**election
DST	**D**ialog **S**tate **T**racker
ECA	**E**mbodied **C**onversational **A**gent
GRU	**G**ated **R**ecurrent **U**nit
HAN	**H**ierarchical **A**ttention **Networks**
HAR	**H**uman **A**ctivity **R**ecognition
HCI	**H**uman **C**omputer **I**nteraction
HMM	**H**idden **M**arkov **M**odel
HUMOD	**Hu**man Annotated **D**ialogues **D**ataset
IOT	**I**nternet **O**f **T**hings
LSTM	**L**ong **S**hort **T**erm **M**emory
NLP	**N**atural **L**anguage **P**rocessing
OOV	**O**ut **O**f **V**ocabulary
RNN	**R**ecurrent **N**eural **N**etwork

SH	**S**mart **H**ome
UI	**U**ser **I**nterface

CHAPTER 1

Introduction

1.1 Background

Ambient assisted living technologies have gained immense interest in recent years for supporting independence and quality of life to the aging population. Current global population trends show that the population above 65 years is growing faster than all other age groups. According to the reports of the United Nations, population over 65 years will increase to 16.0% by 2050 from 9.3% in the present date [1]. The report predicts that one in four persons in North America and Europe could be aged 65 or over by 2050. Rapidly aging populations have social and economic implications around the globe. With the resulting changes, the health care sector is facing several challenges from the shortage of caregivers and skilled workers, to increased demand for home care and eventually increase in health care costs. Innovative new technologies are needed to assist the needs of the elderly and empower them to continue to lead an independent and comfortable life.

Ambient assisted living technologies (AAL) uses different information and communication technologies, stand-alone devices, and smart home solutions with various functionalities such as social participation, fall detection, robotic support, and health monitoring which enable individuals to remain active and lead an independent life. Among all AAL solutions, smart home (SH) has gained a lot of interest due to its versatile applications in the area of the Internet of Things (IoT). A smart home provides an ambient assisted living environment where installed IoT devices have the capability to interact with each other and with the resident. These devices include smart appliances such as television, refrigerator, lighting and heating control, sensors recognizing environment stimuli such as movements, security cameras, and voice assistant, which are interconnected with standardized communication protocol [2]. Various factors need to be considered to develop a successful AAL solution which includes privacy and data protection; interoperability and communication between devices, robustness, and meeting the end-user needs and requirements.

Various research projects have been designed and implemented successfully for developing ambient assisted living technologies such as Cogni Win project introduced a personalized adaptive interface and learning assistant for older people [3]. RelaxedCare project allows easy integration of innovative off-the-shelf products, which helps in the acceptance of technology in daily living [4]. Agewell [5] project provide a personalized virtual assistant to support an active and healthy life for older people and Active@Home [6] focused on physical, cognitive, and social aspects, aims at promoting physical activity at home and foster fall prevention. The current research is a part of the European Union Horizon 2020 MSCA ITN ACROSSING project [7], with an aim to develop an open Smart Home (SH) technology infrastructure by interlinking disciplines from sensor technology and integration, context inferences, and interaction, to service infrastructures, and considering key principles of social impact, security and privacy. The overall project focuses on four scientific objectives (SO) and four applications demonstrators (APP). This research is a part of one of the scientific objective of ACROSSING, led by AIT Austrian Institute of Technology GmbH which focuses on open SH platform and service infrastructure consisting of three major areas of SH that is improving human activity recognition; addressing privacy concerns, and developing dialogue systems, with an emphasis on a framework that is extensible, adaptive and scalable for real-world applications.

1.2 Motivation

Ambient Assisted Living solution means a whole repertoire of solutions with the aim to make everyday life easier for the elderly and people with disabilities. The smart AAL solution can be recognized and accepted when they are easy to operate, monitor ubiquitously, improve comfort, provide privacy and safety and integrate discreetly into the everyday life of people in the need of care. The recent trend in the smart home market shows that there has been major growth in smart home devices, health monitoring products, and smart HVACR (heating, ventilation, air conditioning, and refrigeration) systems. The penetration rate of smart home technology is on the rise in each segment i.e. smart appliances, home monitoring, control and connectivity, energy management, and security. The smart home market was valued at USD 64.60 billion in 2019 and is expected to reach USD 246.42 billion by 2025. This trend is certainly set to continue into the next decade as industry analysis states a compound annual growth rate (CAGR) of 25% between the period 2020 and 2025 [8]. The term ”smart” gets misused a fair amount in home technology as just connecting home appliances to the internet does not make devices smart, a better term, in this case, is ”connected home”. The operation of these devices is primarily based on the analysis of sensor data, use of machine learning, natural language processing, and other technologies that are capable of learning and decision-making. Therefore, one of the major trends today is the increasing use of machine learning and artificial intelligence (AI) technology in the smart home.

Although the idea of smart home technology has shown immense interest and projected huge growth, there are still barriers and gaps that need to overcome. The key barriers are the cost of devices which are more than the traditional device, lack of interest, benefits are unclear to the consumers, also sometimes the system does not meet the end-user requirements, lack of standards and communication protocols that enable devices to interact with each other, data privacy concerns and lack of trust in providers related to privacy risk-aware data sharing. These barriers and gaps contribute to a weak value proposition for residents to invest in smart homes. Previous works have introduced various smart home frameworks that were focused only on a particular component of the framework and on resolving specific issues in smart home such as activity recognition module, privacy-preserving SH framework, an open framework with standard communication protocols, or adaptive user interface framework. However, a complete smart home framework requires the integration of the majority of components. The three primary components of the SH framework are mainly home monitoring component, also known as activity recognition with ubiquitous sensors; privacy-preserving data management components for protecting residents’ sensitive data, and interaction component or dialogue manager for communication within the smart home. Each component of the system plays an important role in the smart home environment and interaction among them will help in developing a robust and secure smart home system. In this research, the following challenges and issues were identified and addressed:

- End-user needs and the requirement for developing effective and adaptable smart home solutions.
- Smart home framework that consists of all major components i.e. activity recognition module, privacy module, and dialogue manager for interaction.
- Deep learning methodologies for developing each module of an SH framework

1.3 Objectives, Approaches and Contribution

The main goal of this thesis is to develop a framework and associated technologies for an adaptable and extensible SH system. This research seeks to investigate five major aspects of building an AAL system. The Ph.D. work was performed with the following objectives which also outline the adopted novel approaches and major scientific outcomes:

1. *A holistic framework for smart home system integrating different components*

   This work first proposed a smart home framework integrating each module and sensor components of the system using an open-source middleware. The framework combines all the major components of a smart home system and provides a scalable and extensible solution for different AAL based applications.

2. *Identification of users' perception towards AAL technologies, their needs, and concerns towards monitoring and data sharing*

   The objective of this work is to identify end-user requirements and their perceptions towards smart home and AAL technologies in order to develop an adaptable solution. In this regard, interviews and online surveys were conducted with the older people, caregivers, and end-users of the system. The findings of the study contribute immensely in understanding the needs, concerns, and privacy preferences of the end-users towards home monitoring, data sharing, and voice assistant or interaction system.

3. *Methodologies for improving human activity recognition with ubiquitous sensing*

   Until recently traditional machine learning approaches such as Naive Bayes, Hidden Markov Models, and Conditional Random Fields were used for activity recognition and require feature engineering for accurate classification. Therefore, the objective of this work is to explore and implement deep learning algorithms for human activity recognition, which do not require handcrafting of features and performs well on raw sensor data. In addition, the advantage of deep learning algorithms is their scalability to a large number of datasets. In this work, we were the first to investigate deep learning algorithms such as Long short term memory (LSTM) and Convolutional neural network (CNN) on smart home sensor dataset. Furthermore, the research explores data imbalance problems on multiple resident settings using different class imbalance approaches and deep learning algorithms.

4. *The development of data transformation and anonymization methodologies for privacy preserving*

   This objective aims to create a mechanism that is able to provide flexible and adaptable anonymization and data transformation. These mechanisms are especially useful in cases where a specific subset of the data should be communicated and other sets should not. Since in a smart home setting, a large amount of personal and sensitive data of users are collected, therefore it is crucial to develop a privacy-preserving component that can provide accurate analysis with minimal distortion of the data. In order to meet this objective, we developed a deep learning method that enables to share encoded data instead of raw or original data with third parties. The third parties can decode the encoded data to view the information as per user preferences. The method offers flexible data anonymization and obfuscation for multiple stakeholders at different access levels without sharing sensitive data.

5. *A novel dialogue dataset and methodologies for developing a dialogue systems*

   Direct natural interaction as dialogue in smart home through a user interface, is another important component for smart home users, especially for older people. The objective of this work is to develop an open domain dialogue dataset with human ratings and diverse human replies for the given dialogue histories. The availability of such a dataset opens various research possibilities and the development of dialogue systems for real-life applications. In addition, a new method to train dialogue managers is presented in this work. The proposed method may help in Natural language processing (NLP) tasks where textual data can be processed in visual format or where dialogue management has low language variation but more choosing the right information to present user.

1.4 Outline of the thesis

Figure 1.1 presents the structure of the thesis and reminder of this thesis is organized as follows:

- Chapter 2: Literature Review

  This chapter discusses the existing research on ambient assisted living technologies and smart homes. Following that state of the art approaches in activity recognition, privacy-preserving mechanisms and dialogue systems are also discussed. Each section of the chapter outlines the issues and challenges in the existing research and provides a good understanding of how it can be addressed.

- Chapter 3: Smart Home Framework with HOMER

  Chapter 3 presents the proposed privacy enabled smart home framework with a voice assistant. The proposed framework shows how different modules can be integrated within a HOMER middleware. HOMER is an open-source software platform that provides integration of various modules as per the user requirements and provides flexibility and scalability to the system. Each component of the system architecture and interaction among them are discussed in detail. Besides, it also presented two use case scenarios where the framework can be applicable.

The content of this chapter is based on the following publication:

***Singh, D.**, Psychoula, I., Merdivan, E., Kropf, J., Hanke, S., Sandner, E., Chen, L. and Holzinger, A., 2020. Privacy-Enabled Smart Home Framework with Voice Assistant. In Smart Assisted Living (pp. 321-339). Springer, Cham.*[9]

Deepika Singh is the main author of the publication and developed the idea to present a system architecture for a smart home. HOMER middleware was developed at AIT GmbH, therefore discussion with Johannes Kropf, Sten Hanke, and Emanual Sandner was really helpful regarding how individual modules can be integrated. The manuscript was written by Deepika Singh. The other authors contributed to the scientific discussion and revision of the manuscript.

- Chapter 4: Users' perception towards AAL and smart home technologies

  It focuses on identifying users' smart home requirements and their perception towards AAL technologies. The chapter includes three publications and each publication contributes in finding smart home requirements from the perceptive of older people, upcoming smart home users, and their privacy concerns towards the Internet of Things. The first publication of this chapter focuses on identifying older people, caregivers, and professionals' needs in smart home and their perceptions of existing AAL technologies. The second publication is more focused on the needs and concerns of future smart home users and the third publication provides insights into the privacy issues from the perceptive of end-users from all around the world.

  The content of this chapter is based on the following publication:

  ***Singh, D.**, Kropf, J., Hanke, S. and Holzinger, A., 2017, August. Ambient assisted living technologies from the perspectives of older people and professionals. In International*

Cross-Domain Conference for Machine Learning and Knowledge Extraction (pp. 255-266). Springer, Cham.[10]

Deepika Singh designed the user studies, questionnaires, interviewed the participants, acquired and analyzed the data. The manuscript was written by Deepika Singh and the paper was presented at CD-MAKE 2017 conference in Regio di Calabria, Italy. The other authors contributed to the revision of the manuscript.

***Singh, D.**, Psychoula, I., Kropf, J., Hanke, S. and Holzinger, A., 2018, July. Users' perceptions and attitudes towards smart home technologies. In International Conference on Smart Homes and Health Telematics (pp. 203-214). Springer, Cham.*[11]

Deepika Singh designed the user studies, online questionnaires, acquired and analyzed the data. The manuscript was written by Deepika Singh and the paper was presented at ICOST 2018 conference in Singapore. The paper received the best paper award at the conference. The other authors contributed to the revision of the manuscript.

Psychoula, I., **Singh, D.****, *Chen, L., Chen, F., Holzinger, A. and Ning, H., 2018, October. Users' Privacy Concerns in IoT Based Applications. In 2018 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) (pp. 1887-1894). IEEE.*[12]

Deepika Singh and Ismini Psychoula equally contributed in the publication and designed the user studies, online questionnaires, acquired and analyzed the data. The manuscript was written by Deepika Singh and Ismini Psychoula. The paper was presented at the IEEE SmartWorld conference 2018 in Guangzhou, China. The other authors contributed to the scientific discussion and revision of the manuscript.

- Chapter 5: Human activity recognition in smart home

  This chapter presents the work in human activity recognition using deep learning methodologies. The work comprises three publications in human activity recognition for both single and multiple residents in the smart home domain.

* denotes equal contribution.

The content of this chapter is based on the following publication:

Singh, D., *Merdivan, E., Psychoula, I., Kropf, J., Hanke, S., Geist, M. and Holzinger, A., 2017, August. Human activity recognition using recurrent neural networks. In International Cross-Domain Conference for Machine Learning and Knowledge Extraction (pp. 267-274). Springer, Cham.*[13]

Deepika Singh proposed the idea and both Deepika Singh and Erinc Merdivan performed the experiments and evaluation on the sensor dataset. The manuscript was written and presented at CD-MAKE 2017 in Regio di Calabria, Italy by Deepika Singh. The other authors contributed to the scientific discussion and revision of the manuscript.

Singh, D., *Merdivan, E., Hanke, S., Kropf, J., Geist, M. and Holzinger, A., 2017. Convolutional and recurrent neural networks for activity recognition in smart environment. In Towards integrative machine learning and knowledge extraction (pp. 194-205). Springer, Cham.*[14]

Deepika Singh initiated the work and both Deepika Singh and Erinc Merdivan performed the experiments and evaluation on the sensor dataset. The manuscript was written by Deepika Singh. The other authors contributed to the scientific discussion and revision of the manuscript.

Singh, D, *Merdivan, E., Kropf, J. and Holzinger, A. Handling Imbalanced Data in Deep Learning for Multiple Resident Activity Recognition. Submitted to IEEE Transactions on Neural Networks and Learning Systems (2021).*

Deepika Singh designed and performed all the experiments. The manuscript was written by Deepika Singh and the other authors contributed to the scientific discussion and revision of the manuscript.

- Chapter 6: Privacy preservation in smart home using deep learning

  Chapter 6 introduces the privacy-preserving deep learning mechanism for flexible anonymization and data sharing in AAL. The chapter includes two publications. The first publication describes the mechanism that was developed and its evaluation on the simulated dataset and the extension of this work was a part of the second publication mentioned.

  The content of this chapter is based on the publication:

Psychoula, I., Merdivan, E., ***Singh, D.****, Chen, L., Chen, F., Hanke, S., Kropf, J., Holzinger, A. and Geist, M., 2018, March. A deep learning approach for privacy preservation in assisted living. In 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) (pp. 710-715). IEEE.*[15]

Deepika Singh, Ismini Pyschoula, and Erinc Merdivan worked together in designing and implementation of experiments. Data generation was conducted by Ismini Psychoula. The manuscript was written by both Ismini Psychoula and Deepika Singh. Ismini Psychoula presented the paper at PERCOM 2018, Athens, Greece. The other authors contributed to the scientific discussion and revision of the manuscript.

Psychoula, I., ***Singh, D.****, Merdivan, E., Chen, L, 2021, January. Privacy Preservation with Autoencoder based De-Identification and Differential Privacy. Submitted*

Deepika Singh, Ismini Pyschoula, and Erinc Merdivan worked together in designing the experiments. The first part of the experiments was performed by Deepika Singh and Erinc Merdivan. And, the experiments on Differential privacy was performed by Ismini Psychoula. The manuscript was mainly written by Ismini Psychoula and Deepika Singh partially contributed in some sections of the manuscript. The other authors contributed to the scientific discussion and revision of the manuscript.

- Chapter 7: Dialogue systems

  Chapter 7 focuses on dialogue systems in the smart home domain. The chapter consists of two publications. The first publication presents the developed human-annotated dialogue dataset for conversational agents and the second publication presents the overview of the existing methods for training dialogue manager and proposed a new method to train dialogue manager.

  The content of this chapter is based on the following publications:

  Merdivan, E.,* ***Singh, D.****, Hanke, S., Kropf, J., Holzinger, A. and Geist, M., 2020. Human annotated dialogues dataset for natural conversational agents. Applied Sciences, 10(3), p.762.*[16]

  Deepika Singh and Erinc Merdivan equally contributed to this work. The development of the website, data collection, and analysis was mainly performed by Deepika Singh. The experiments on the dataset were carried out together with Erinc Merdivan. The manuscript was mainly written by Deepika Singh together with continuous discussion with Erinc Merdivan. The remaining authors contributed to the revision of the manuscript.

Merdivan, E., **Singh, D.***, Hanke, S. and Holzinger, A., 2019. Dialogue systems for intelligent human computer interactions. Electronic Notes in Theoretical Computer Science, 343, pp.57-71.*[17]

Deepika Singh and Erinc Merdivan equally contributed to this work. The designing of the method is performed together with Erinc Merdivan. Erinc Merdivan implemented the experiments of this work and the manuscript was written by Deepika Singh. The other authors contributed to the scientific discussion and revision of the manuscript.

- Chapter 8: Conclusion and Future work

  It summarises the studies that have been conducted in the thesis. First, it provides an overview of the entire research project and the main research contributions. It concludes with discussions of the results and directions for future research in this area.

1.5 Additional publications

Followings are the additional publications during my doctoral studies which are not included in this thesis:

- *Machado, E., **Singh, D.**, Cruciani, F., Chen, L., Hanke, S., Salvago, F., Kropf, J. and Holzinger, A., 2018, March. A conceptual framework for adaptive user interfaces for older adults. In 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) (pp. 782-787). IEEE.*

- *Saranti, A., Streit, S., Müller, H., **Singh, D.**, Holzinger, A., 2020, September. Towards Visual Concept Learning and Reasoning: On Insights into Representative*

* denotes equal contribution.

Approaches. In 25th International Symposium on Methodologies for Intelligent Systems, ISMIS 2020.

- ***Singh, D.***, *Kropf, J., 2017, March. Activity Recognition using context aware techniques in Smart Home domain. In MCAA General Assembly and Conference, Salamanca, Spain. (Poster).*

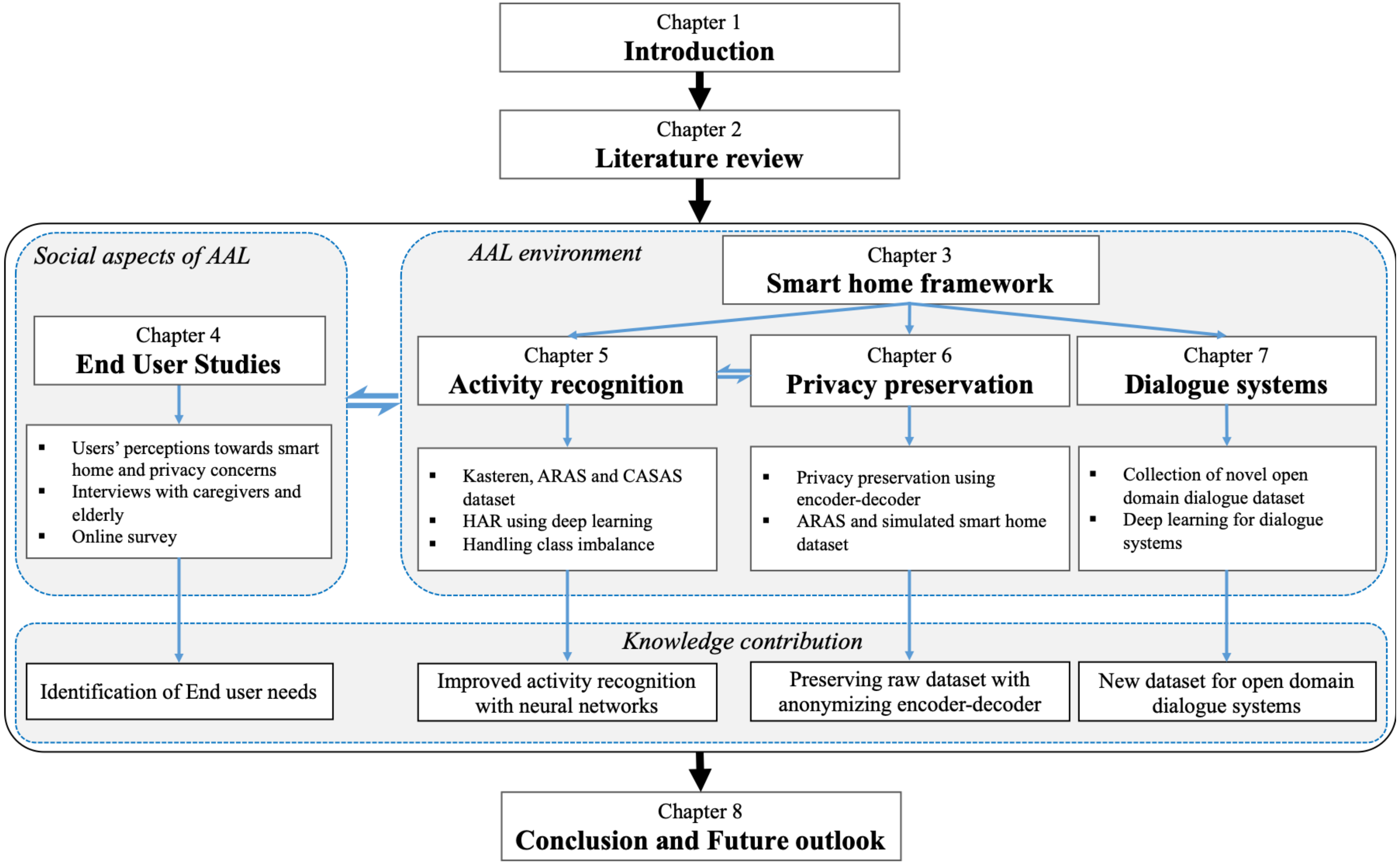


FIGURE 1.1: Structure of the thesis

CHAPTER 2

Literature Review

2.1 Ambient Assisted Living and Smart Homes

The increasing aging population brings many challenges to society and the healthcare systems. Such a demographics shift has led to the introduction of Ambient and Assisted Living technologies (AAL). AAL technologies involve the use of information and communication technologies (ICT), smart home technologies, and stand-alone devices in a users' daily living and working environment enabling people to stay active longer and provide comfort, safety and independent life in old age [18]. It provides supportive home environments to the residents by integrating actuators, sensors, smart interfaces, and artificial intelligence technologies [19]. The basis of AAL lies in the following: assistive technologies for individuals with disabilities; universal design approaches for accessibility, usability, and acceptability of interactive technologies; and the emerging ambient intelligence (AI) computing paradigm, which provides intelligent, unobtrusive, and ubiquitous assistance [20]. In recent years, there has been rapid growth in AAL research which is partly due to an increase in Ambient Assisted Living Joint Programme, which provides funding to many research projects aimed at developing novel and innovative AAL solutions across Europe [21]. Various AAL projects are conducted to enhance the quality of life of older people with a range of health-monitoring and emergency-alert devices to ensure safety and security in their home environment. These projects are categorized into physical health systems, mental health systems, enhanced user interfaces, support in daily tasks, security systems, and standardization of AAL systems [22]. Some of the existing projects that promote physical health are CASAS project [23] in which a smart home is designed to monitor activity, behavior change and provide prompts and reminders using sensors and actuators; Dem@Care project [24] that collects information on activity, physiology, and lifestyle through sensors in order to better assist older adults with dementia; vAssist [21] helps in providing voice-controlled telehealth, telemonitoring, and communication services and Fearless [21] tracks risks and behavioral changes visually and acoustically and contact care providers in emergencies. AAL projects that support mental health are Companion Able [25] combines robotic and AI technology into a friendly robot that promotes cognitive stimulation and health management for older adults with mild cognitive impairment and Host [21] offers easy-to-use technologies with digital self-serve systems and connections to support systems. Various systems were developed to support daily living tasks such as Assistant project supports navigation ability for older people using a smartphone, GPS, and special user interface; Bank4Elder project assist in banking systems through the user interface, and MobileSage project encourages travel and transportation by providing instruction from an agent using smartphone [21]. AALiance 2 [26] and UniversAAL [27] projects were designed with an aim to support the standardization of AAL technology by creating recommendations and guidelines for standardization, universality, coordination, and regulation of these technologies.

Research and development in smart home technology have gained attention around the world and different smart home projects have been conducted in previous years. Smart home project in Boulder, Colorado uses the adaptive control of home environment (ACHE) system and neural networks to observe and predict the resident lifestyle and needs [28]. The aim of MavHome project [29] was to improve the comfort of its resident while reducing the operation costs and behaves as a rational agent to guess the mobility patterns and device usages of the residents. The GatorTech Smart House project [30] was developed in Florida which consists of several smart devices including sensors and actuators which are designed to optimize the comfort and safety of older people. Sixteen smart homes were built in Ireland within the Great Northern Heaven Smart Home project [31] to detect the behavioral patterns in the Activities of Daily Living (ADL) for older people. The concept of the smart home has shown to be a feasible and cost-effective projects to help older people to improve their comfort and live independently in their private spaces for as long as possible. However, a smart home presents some challenges that need to be overcome. One of the major hurdles for deploying AAL systems in real-world is technology acceptance by the older people [32] [33]. Lack of perceived benefits [34] [35], usability problems [36] [37] and self-efficacy [33] [38] are other barriers in technology adoption among older people. The challenges with existing smart home technology are high cost, the cumbersome installation process of different sensors or devices and no open smart home architecture has been developed so far that enable interoperability, extensibility, and which can be configurable to different application scenarios. Data integration is another issue in the smart home system, a large amount of data generated are not properly handled by the existing AI modules. In addition, legal and privacy issues need to be considered while data processing, storing, and sharing in smart home domain [39].

In order to develop an effective and adaptable smart home system that can address requirements of the older people, caregivers, and end-users, their suggestions and inputs are much needed. Different studies have been performed in previous works with older people regarding their opinions and suggestions [40] [41]. The participants showed interest in the AAL technologies and the need for a smart home for living independently and better quality of life [42]. However, the most important aspect is the acceptance of technology by the older people, caregivers, professionals, and other end-users [43]. Therefore, we conducted interviews with the elderly, caregivers, and professionals to understand their needs and perspective towards AAL technologies. Furthermore, online surveys were conducted which focuses on various factors such as users' attitudes towards monitoring; privacy preferences and data sharing; and views regarding voice assistant such as robots or communication interface.

2.2 Activity Recognition

Human activity recognition is one of the most crucial characteristics in the AAL system especially in the smart home domain, which is capable of understanding the behavior, routines, and needs of the users. Several works have been performed in previous years on sensor-based activity recognition as they can better address the sensitive issues in the assisted living environment such as privacy, ethics, and obtrusiveness than conventional vision-based approaches. Sensor-based activity monitoring is classified into two main categories: wearable sensor-based activity monitoring and ambient sensor-based activity monitoring [44]. Wearable sensor-based activity monitoring is mainly driven by pervasive and mobile computing and ambient sensing is mainly used in smart environment applications. Wearable sensors are generally positioned directly on a human body and can be embedded into wristwatches, clothes, eyeglasses, or mobile devices. Various research work has been conducted using accelerometer sensors for monitoring body movements [45] [46]; GPS sensors for monitoring location-based activities [47] [48] and Biosensors to measure vital signs such as heart rate, blood pressure, ECG, EEG, and respiratory information [49]. However, wearable sensor-based activity monitoring has limitations which include the size of the devices and ability to wear them all the time to be monitored, battery life, acceptability of device by older people and wearable sensors are not suitable for monitoring complex physical activities and activities that involve interaction with the environment. Ambient sensor-based activity recognition has an advantage over wearable sensor-based activity monitoring as ambient sensors are attached to objects within an environment and activities are monitored through user-object interaction. Therefore, ambient sensing is more suitable for intelligent environment based applications and widely used in the smart home paradigm.

Numerous approaches have been used in previous works for activity modeling and recognition which includes data-driven approaches and knowledge-driven approaches [44]. Data-driven activity modeling is classified into the generative and discriminative approaches. Naive Bayes, Hidden Markov Model (HMM), and Dynamic Bayesian networks (DBNs) are the most popular generative approach for activity recognition. These classifiers are simple and yield good accuracy, but they have some limitations. Naive Bayes classifiers do not explicitly model temporal information, which is most important in activity recognition. HMM classifier is not capable of capturing long-range or transitive dependencies of the sensor observations due to its strict assumptions on the observations. In discriminative approach, Nearest Neighbour (NN) classifier was investigated for activity recognition from accelerometer data [50]. However, the decision trees method outperformed the simple nearest neighbor approach, where the training data is partitioned into subsets as per activity labels and a set of rules based on features of the training data. Other classifiers that are used in discriminative approach are Support Vector Machine (SVM) [51] and Conditional Random Field [52] for activity recognition. Both generative and discriminative approaches require large datasets for training the models and thus sometimes suffers from data scarcity problem. Knowledge-driven activity modeling avoids the problems of data-driven activity modeling that are a requirement of a large amount of data and the inflexibility issue that each model needs to be computationally trained. Knowledge-driven approaches use rich domain knowledge and heuristics for activity modeling and pattern recognition and knowledge structures can be represented in different forms such as rules, schemas, and networks. These approaches are further classified into three categories: mining-based approach, logic-based approach, and ontology based approach. The weakness of these approaches is the inability to represent fuzziness and uncertainty. Besides, logical activity models are seen as one model fits all, thus inflexible for adoption to different users' activity patterns. Existing works proposed hybrid approaches for activity recognition since both symbolic and statistical methods have some limitations. One of the work uses Markov Logic Networks (MLN), which is a probabilistic first-order logic approach [53] in which weights can be learned for each probabilistic formula by repetitively optimizing a pseudo-likelihood measure. These weights are the confidence value of the formula. The addition of deterministic formulas to probabilistic ones is performed to express deterministic knowledge about the domain of interest. A hybrid technique was proposed in [54] to recognize anomalies at a fine-grained level by integrating supervised learning and symbolical reasoning approach on a smart home sensor dataset. The benefit of using probabilistic description logic is that it defines complex knowledge-based constraints that can capture the uncertainty of sensor measurements. Therefore, by learning the weights of these constraints, the model combines useful features of knowledge-based and data-driven methods and can improve the recognition accuracy. These approaches still require a labeled dataset for activity recognition. Another work [55] presents an unsupervised method which combines data and knowledge-driven methods in which a general ontology model represent domain knowledge and augment a range of learning techniques to facilitate the unsupervised pattern recognition of user activities.

Conventional pattern recognition approaches have made immense progress in human activity recognition. However, there are still drawbacks to these methods. In most activity recognition tasks, the features are always extracted via heuristics and hand-crafted way, which relies on human experience and domain knowledge [56]. Human knowledge may be helpful in certain tasks, but for more general tasks and environments, it will take a longer time to build a successful activity recognition system. Furthermore, only shallow features can be learned by these conventional pattern recognition approaches. These shallow features refer to statistical information including mean, frequency, and variance, etc., which can only be used to recognize low-level activities such as running or walking and difficult to infer high level or context-aware activities [57]. Real-time activity recognition requires robust and incremental learning models whereas existing pattern recognition approaches mainly focuses on learning from static data. Deep learning models overcome these limitations by learning high-level features automatically through neural networks instead of manual handcrafting, thus achieved promising performance in many areas such as natural language processing, visual object recognition and logic reasoning [58]. Deep learning networks are also more feasible in performing unsupervised [59] and incremental learning. In addition, a trained deep learning model on a large-scale labeled dataset can be transferred to new tasks where there are few or no labels. Therefore, in this thesis, we focused on deep learning approaches and were the first to apply different deep learning models for activity recognition on smart home datasets. Also, an application of deep learning models in combination with class imbalance techniques is explored to handle imbalanced datasets in multiple resident settings.

2.3 Privacy Preserving data management

There has been immense growth in the research and development of AAL systems over the past few years to provide personalized health services to older people. Apart from older people who are the main stakeholders, there are other users of AAL systems as well such as caregivers, doctors, pharmacists, hospitals, family members, or researchers. Different smart devices are used in an ambient environment such as sensors for activity monitoring and user behavior analysis, wearable to monitor vital signs, and voice-controlled appliances for interaction and assistance. These devices collect and use sensitive personal data of the user to provide various services. The continuous data monitoring and collection may cause security and privacy concerns and risks to the users of the system. In addition, there can be a possibility of information leakage and risk of data breaches [60]. Hence, there is a need for new techniques which can provide transparency, user control to the system and ensure that individual privacy requirements are met. In order to develop such a system, it is crucial to understand the users' thoughts about privacy implications and their preferences on data sharing. Various studies have been performed to understand users' requirements, their concerns, and perceptions towards smart home [61]. The participants of the surveys showed interest in assistive technologies and the need for smart home technology for independent living, comfort, safety and better quality of life [42]. Several other studies were performed to understand the users' privacy concerns and their attitudes towards data sharing [62] [63]. An interesting finding in the studies is that the privacy concerns of the users reduce when smart devices are used for medical purposes such as in case of emergency location tracking or health monitoring. In addition, privacy concerns are varied according to the individual needs. Older adults those who are physically active and can perform most of their daily tasks by themselves, do not like to be monitored all the time whereas older adults who are dependent on others and suffering from serious health problems, would like to be monitored and have no privacy concerns. Existing surveys have already investigated the impact of the location of data collection, type of data collected and purpose of collection [64], however, it would be interesting to identify privacy concerns in heterogeneous scenarios that use different types of data collection and formats. Such factors will help in developing privacy-enable solutions suitable to a variety of contexts in IoT environments.

Privacy-enhancing technologies are technologies that are designed to protect users' privacy by minimizing personal data use, maximizing data security, and offer additional levels of protection besides just reckoning on laws and policies. Several approaches have been proposed to address the privacy concerns of the users which include privacy and context awareness, information manipulation, access control, and data anonymization. Anonymization methods are used to obscure the identity of a user within a dataset of different users, such that individual users are unidentifiable and indistinguishable. Some of these methods are k-anonymity, t-closeness and l-diversity. Among these, the most popular method is k-anonymity. K-anonymity method is achieved by suppressing (deleting an attribute value from the data and replacing it with a random value that matches any possible attribute value) or generalizing the attributes in the data, which means that an attribute is replaced with a less specific but semantically consistent value [65]. In most of the cases K-anonymity is able to prevent identity disclosure but in some cases, it may fail to shield against attribute disclosure. l-diversity method is another commonly used anonymization technique, which is developed to address the weakness of K-anonymity that does not guarantee privacy against adversaries that use background knowledge or where data lack diversity. In l-diversity, the anonymization conditions are satisfied if, for each group of records sharing a combination of key attributes, there are at least l -"well-represented" values for each confidential attribute [66]. The drawback of this method is that it depends on the range of sensitive attributes. If such attributes do not have numerous different values, then in that scenario artificial data will be inserted. Insertion of artificial data will improve privacy but may cause problems in the data analysis, thus wrecking the data utility. Besides, this method is vulnerable to skewness and similarity attacks, therefore it cannot always prevent attribute disclosure. Another method named t-closeness [67] was proposed to address the problems of l-diversity.

This method requires the distribution of the sensitive attributes in an equivalent class to be close to the distribution of the attribute in the overall table, which in turn means that the distance between the two distributions should be no more than a specified threshold t. In [67], it is mentioned that *t*-closeness limits the amount of useful information which is released, however, no computational procedure to enforce this property is given so far.

Previous works have used traditional machine learning methods for privacy preservation. The most popular among them is the Differential privacy-preserving method [68] which allows the gathering of statistics from a dataset while protecting the information of individual records. The differential privacy method estimates how much noise is required to add to the dataset to achieve privacy guarantees. Some of its applications include principal component analysis, boosting, support vector machines, and continuous data processing. The advantage of this method is that it can be used through database queries to authorized third parties instead of publishing whole datasets. A limitation of this method is in the case of inference tasks where a single record is fully exposed. Since the differential privacy method is only applied to the query results and does not modify the original dataset, in such a scenario a data controller is able to identify the individuals in a dataset. Furthermore, the differential privacy method is not as effective in datasets with strong correlations such as medical data, mobile or social datasets [69] [70]. Secure multi-party computation (SMC) [71] is another privacy-preserving technique that aims at protecting intermediate steps of computation when multiple parties perform collaborative machine learning. Some of its applications are Naive Bayes, linear regression, decision trees, k-means clustering, and association rules.

In recent works, deep learning models have been used for privacy-preserving. In [72], proposed a privacy-preserving system which allows participants to train independently on their datasets and selectively share models key parameters during training. An alternative method [73] computes the privacy loss in each model to select parameter updates and ensure differential privacy for the stochastic gradient descent algorithm. Private Aggregation of Teacher Ensembles (PATE) [74] is another deep learning approach which partitions sensitive training data to train independently an ensemble of machine learning models on each data partition. The model prediction on test data is by collectively voting on one of the possible labels. In addition, Homomorphic Encryption [75] together with neural networks have been used for data encryption but currently, it is not computationally feasible to use Fully Homomorphic Encryption (FHE) [76] and then perform classification on the encrypted data. The existing works have used Partially Homomorphic Encryption (PHE) to encrypt data, however, it only allows additions or multiplications over data which can be an issue for classification tasks, since the use of activation functions is no longer possible in PHE and they have to be approximated with low degree polynomials. Unlike existing approaches, the thesis will attempt to address some of the gaps in privacy-preserving with a mechanism for encoding big data which is applicable on heterogeneous raw data and allows the classification on encoded data without exposing any records. The work in the thesis presents a method which is flexible and allows the use of activation functions in classification tasks while not allowing access to the sensitive dataset.

2.4 Dialogue Systems

A dialogue is a verbal or written communication between two or more people or groups of people. With the advancement in artificial intelligence technologies, dialogue can also be conducted between a person and a machine. In recent years, dialogue systems have emerged as a new type of human-machine interaction systems. In AAL and smart home domain, dialogue systems or user interfaces also plays a crucial role. Communication with the users using natural language eases the interaction, especially for older people who are having issues with these technologies due to disregard or disabilities and also during emergency situations such as falls or serious health conditions. Dialogue systems are also known as conversational agents that can communicate with the human in spoken or textual form. The main components of the dialogue system are Automatic speech recognition (ASR), Natural language understanding (NLU), Dialogue Manager (DM), Knowledge base, Response generator, and Text-to-speech synthesizer. In ASR, the user's input is recognized automatically and transformed into text and could be omitted if the user's input is in form of text. NLU identifies user intents and extracts information that the dialogue manager can process. Dialogue manager is the fundamental component of the dialogue system and consists of two modules, Dialogue State Tracker (DST) and Dialogue Response Selection (DRS). DST keeps track of the dialogue state that is required by DRS in order to select the next response. DRS's next response is feedback given to DST to update its current dialogue state. The knowledge base component contains information that is used by the dialogue manager. Response generator generates the output and then text to speech synthesizer converts the output of response generator into a speech form [77] [78].

Dialogue systems are built for various purposes and have been divided into two main groups: non-goal driven dialogue systems and goal-driven dialogue systems [79]. The first non-goal oriented dialogue system was ELIZA [80] and PARRY [81], which were based on simple text parsing rules that mimic by persistently rephrasing statements or asking questions. Neither of the two systems used data-driven methods. Later works [82], begin to use data-driven methods and proposed modeling dialogue as a stochastic sequence of discrete symbols (words) using Markov chains. In the past few years, neural network architecture trained on large-scale corpora has been explored. Neural networks models have shown promising results for several non-goal oriented dialogue tasks [83] [84] [85] [86]. However, these models require sufficiently large corpora to achieve these results. Non-task oriented dialogue systems or chatbots are designed based on two major approaches [87] [88]: (i) generative models such as sequence-to-sequence methods, that generate actual responses during the conversation; and (ii) retrieval-based models, which learn to select responses from the current conversation from a repository. Preliminary work on goal-driven dialogue systems was based on deterministic hand-crafted rules couple with a learned speech recognition model. For example, the SUNDAL system was designed to provide timetable information for trains and airplanes, and airline reservations [89] [90]. Later, machine learning methods were used to bridge the gap between text and speech, considering uncertainty associated with the outputs of the speech recognition model [91]. Goal-oriented dialogue systems are designed for specific domains and tasks. In these systems, the dialogue agent interacts with users in a limited manner to acquire information for completing the task and the performance of the system can be evaluated based on the completion of the task. The dialogue system first understands the user message, represents it as an internal state, then takes actions according to the policy with respect to the dialogue state, and lastly, the action is transformed into natural language. Though natural language understanding is handled by statistical models, the majority of deployed dialogue systems still uses handcrafted rules or manual features for the state and action space representations, intent detection, and slot filling, which makes the system expensive, time-consuming and limits its usage to other domains [92]. In recent works, deep learning models have been developed to reduce these issues by learning feature representations in a high dimensional distributed way and achieved remarkable improvements. Furthermore, end-to-end task-oriented dialogue systems are designed, that can extend the state space representation in the conventional pipeline systems and generalize dialogues outside the annotated task-specific corpora. A detailed overview of deep learning-based researches on dialogue systems is presented in [92] [93].

In the last few years, several publicly available conversational dialogue dataset has been released to train dialogue managers. A major challenge while training dialogue manager is a lack of benchmark metrics that can be used to measure and compare performances of different dialogue manager methodologies. Also, existing research on dialogue datasets lacks in the publicly available dataset with human annotations on the quality of dialogue–reply pairs to develop such metrics. Therefore, in this thesis, we developed a Human Annotated Movie Dialogue Dataset (HUMOD) with human ratings on dialogue-reply pairs together with diverse human replies for given dialogue histories. In addition, we proposed a new method to train dialogue manager.

CHAPTER 3

Smart Home Framework with Home Event Recognition System (HOMER)

3.1 Introduction

In this chapter, we present a conceptual framework of a smart home system with integration of all the major components in Home Event Recognition System (HOMER) middleware. HOMER is an open and flexible OSGi based software platform, developed at AIT Austrian Institute of Technology GmbH, Vienna, Austria. The middleware aims at the integration of different home automation systems and consequential event and situation recognition for Ambient Assisted Living (AAL) and smart home applications. In the proposed SH framework, we present an integration of three major modules i.e. activity recognition and occupancy detection; privacy-preserving data management, and voice assistant with two different use cases. Each module of the framework plays a crucial in monitoring residents as well as protecting their sensitive and personal information through the privacy module, together with a voice assistant for interaction within the smart home domain. This privacy-enabled smart home architecture allows users to have control over the monitoring of their daily activities and flow of information in a secure manner to preserve sensitive and personal information that users do want to share with anyone. The chapter presents a detailed description of each component of the HOMER middleware and its integration with different SH modules such that the system will be scalable, flexible, and adaptable according to the different applications.

The content of this chapter is based on the publication:

Singh, D., Psychoula, I., Merdivan, E., Kropf, J., Hanke, S., Sandner, E., Chen, L. and Holzinger, A., 2020. Privacy-Enabled Smart Home Framework with Voice Assistant. In Smart Assisted Living (pp. 321-339). Springer, Cham.

Contribution: Deepika Singh is the main author of the publication and developed the idea to present a system architecture for a smart home. HOMER middleware was developed at AIT GmbH, therefore discussion with Johannes Kropf, Sten Hanke, and Emanual Sandner was really helpful regarding how individual modules can be integrated.

The manuscript was written by Deepika Singh. The other authors contributed to the scientific discussion and revision of the manuscript.

3.2 Publication I: Privacy-Enabled Smart Home Framework with Voice assistant

Chapter 16
Privacy-Enabled Smart Home Framework with Voice Assistant

Deepika Singh, Ismini Psychoula, Erinc Merdivan, Johannes Kropf, Sten Hanke, Emanuel Sandner, Liming Chen and Andreas Holzinger

Abstract Smart home environment plays a prominent role in improving the quality of life of the residents by enabling home automation, health care and safety through various Internet of Things (IoT) devices. However, a large amount of data generated by sensors in a smart home environment heighten security and privacy concerns among potential users. Some of the data can be sensitive as it contains information about users' private activities, location, behavioural patterns and health status. Other concerns of the users are towards the distribution and sharing of data to third parties. In this chapter, we propose privacy-enabled smart home framework consisting of three major components: activity recognition and occupancy detection, privacy-preserving data management and voice assistant. The proposed platform includes unobtrusive sensors for multiple occupancy detection and activity recognition. The privacy-enabled voice assistant performs interaction with smart home. We also present a detailed description of system architecture with service middleware.

Keywords Smart home · Activity recognition · Occupancy detection · Privacy-preserving data management · Dialogue manager

D. Singh (✉) · E. Merdivan · J. Kropf · E. Sandner
AIT Austrian Institute of Technology, Wiener Neustadt, Austria
e-mail: deepika.singh@ait.ac.at

I. Psychoula · L. Chen
School of Computer Science and Informatics, De Montfort University, Leicester, UK

E. Merdivan
CentraleSupelec, Metz, France

S. Hanke
FH JOANNEUM Gesellschaft mbH, Graz, Austria

D. Singh · A. Holzinger
Holzinger Group, HCI-KDD, Institute for Medical Informatics/Statistics, Medical University Graz, Graz, Austria

F. Chen et al. (eds.), *Smart Assisted Living*, Computer Communications and Networks, https://doi.org/10.1007/978-3-030-25590-9_16

16.1 Introduction

In recent years, smart home technology along with Internet of Things (IoT) devices has gained a lot of attention due to its various applications. It has experienced rapidly growing presence in the households of end-users. A smart home comprises of sensors such as motion sensors, thermostat, passive infrared sensors, energy tracking switches, smart lights, shades, cameras and voice assistants, which communicate with each other and collect data to monitor user activities. The data collected from these embedded sensors and smart devices provide numerous services to the residents such as safety and guidance features by user behaviour monitoring, activity recognition and fall detection; home automation by controlling lights, doors, windows, temperature and energy consumption; and security with alarms, lock/unlock of doors and monitoring of outsiders in the absence of resident.

Various devices such as tablets, personal computers and smartphones are being used for communication and interaction with users in a smart home environment. Recent works have shifted towards dialogue systems in the form of voice assistant. The integration of the dialogue system in a smart home can provide a natural and convenient way of interaction with a user. Dialogue systems are categorized mainly into two groups [5] as task-oriented systems and non-task-oriented systems (also known as chatbots). The aim of task-oriented systems is to assist users to complete specific tasks by understanding the inputs from the user such as restaurant bookings, ticket bookings or information searching. Non-task-oriented dialogue systems can communicate with humans on open domains, and thus, they are preferred in real-world applications. In a smart home environment, task-oriented dialogue systems can assist users in performing various tasks depending on their needs. For example, a medication reminder can checks if a user has taken medicine as prescribed by their doctors, and further, recommend exercises or any other physical activity when sensors in the home detect that a user is in a sedentary state for a considerably longer time. It could also make an appointment with a doctor on behalf of the user. A dialogue system can also be customized according to the user's preferences.

A user's interaction with a dialogue system and sensors installed in the home environment generate a large amount of data. Such data usually contain private and sensitive information such as users' profile, locations, pictures, daily activities and network access information, which leads to privacy, and security concerns for the users. These raised concerns include "who manage and regulate these devices?", "where the data is stored?" and "who has access to the data?" [41], as well as how do machine learning systems reach the decisions they do and how fair and non-biased are the algorithms used in them? [37, 38].

Previous studies have introduced various smart home frameworks with network-level security and privacy in IoT devices. A three-party architecture with Security Management Provider (SMP) has been proposed, which identifies and blocks the threats and attacks at the network level [35]. Passive network adversary can infer private activities of a user from smart home traffic rates even when devices use encryption, this problem has been discussed in [3] by introducing stochastic traffic padding technique. Different frameworks have been designed for user activity recognition [15, 21], occupancy detection [40], adaptive user interaction [17] and dialogue systems [24] in smart home environments. Most of the existing frameworks focus only on one particular component of the system, for example, only activity recognition module, adaptive interface or privacy in smart home IoT devices. Therefore, in this work we presented a framework, which combines three major components of the smart home system, i.e. user occupancy detection, privacy-preserving data management and dialogue system for user interaction. The proposed framework focuses on the development of technological solutions, which can monitor the users and environment while keeping their privacy intact.

The remainder of this chapter is organized as follows: the related work of each component of the framework is presented in Sect. 16.2. Section 16.3 gives an overview of the proposed framework. System architecture of the proposed framework with a detailed description of individual component is explained in Sect. 16.4. Section 16.5 presents the two different scenarios based on the proposed framework. The last section includes the conclusion and future work.

16.2 Related Work

In this section, we review related work focusing on the three major components of the framework, which are occupancy detection and activity recognition, privacy preservation in data management and dialogue systems.

16.2.1 Activity Recognition and Occupancy Detection Module

Human activity recognition has an important role in a smart home system by learning high-level knowledge about the user's daily living activities from sensor data. The monitoring systems can be categorized mainly into three categories: (1) camera-based, (2) wearable devices and (3) binary and continuous sensors. Camera-based monitoring is not preferred by the users as it is considered privacy invasive. Wearable devices provide good accuracy in the case of personalized systems and do not raise as many privacy concerns, and however, such systems are not practical to use while monitoring long-term activities. As a result, the most preferred solution for activity recognition is using unobtrusive sensors in a smart home. Early work on activity recognition have used data mining methods and knowledge-driven approaches, e.g. Chen et al. [6], Okeyo et al. [23]. In the recent work, deep learning approaches have contributed tremendously towards activity recognition systems [32, 33] and tend to overcome the limitations of conventional pattern recognition approaches. In deep neural networks, the features can be learned automatically instead of using manually handcrafted features. The deep generative networks can also make use of unlabelled data for model training as in activity recognition systems; it is not feasible to have labelled data most of the time [37, 38].

Occupancy detection and estimation plays a major role in reducing energy consumption by controlling heating, ventilation and air conditioning (HVAC) and lighting systems in the smart environment. It also helps in the detection of intruders, abnormal events such as falls and monitoring activities of multiple residents inside a home environment. Different methodologies on occupancy estimation have been presented and discussed which include occupancy detection using the camera, passive infrared sensor (PIR), ultrasonic sensor, radio frequency signals (RFID), fusion of sensors and using wireless local area network (WLAN), Bluetooth and Wi-fi [2]. Camera-based occupancy estimation is accurate in comparison with other methods but due to privacy concerns, they are not preferable in real-time applications. PIR and ultrasonic sensors are economical but can only detect the presence/absence of an occupant as it produces a binary output. Use of RFID tags is not feasible in real-life situations as occupants have to carry RFID all the time and it has privacy issues as well. Occupancy detection through WLAN, Wi-fi and Bluetooth technologies is not applicable in large buildings due to its high positive/negative detection of occupants. Use of multiple sensors such as CO_2, PIR, temperature, humidity, light and motion sensors for occupancy detection provides accurate results and can be easily applied in real-time applications. Therefore, we prefer data generated from fusion of multiple sensors for occupancy detection in the proposed smart home framework.

16.2.2 Explainable AI in Smart Home Systems

As can be seen in previous sections, the success of machine learning in activity recognition and advancement of smart home technology has been undeniable. However, there are still concerns about adopting these learning techniques in practice. One of the reasons is that with these techniques it is quite difficult to understand what goes on inside an algorithm and why it gave the result it did. This difficulty of explaining the reasoning behind the generated result makes the decisions provided by these systems hard to be trusted by the end-users. A lot of recent work has been dedicated to making machine learning models explainable [12, 13]. There are two main aims of work on interpretability in the literature: transparency and post hoc interpretation [1]. Transparent design reveals how a model functions, while a Post hoc Explanation explains why a black-box model behaved that way [19].

Explainability enables the privacy risk discussions that need to happen when decisions are being made in the development and design stages of a machine learning model. When the decisions that the model makes will affect individuals based just on the model's output, there is the possibility for bias and wrong decision-making. Many machine learning models include multi-layer neural networks, which follow an internal process in which outcomes may not be able to be understood in a mathematical way even by experts in the field. Due to this, multi-layer features, model performs better. This is often referred to as the accuracy vs. interpretability trade-off which is also known as the "black-box" problem. Some of the solutions to this problem are following the steps of the algorithm and describe with more details about what the model is doing. Another way is to perform risk management assessments that evaluate the model, the data and the output and the risk to the individual if the decision reached is not correct which will also affect the privacy of the individual. Bias can often occur based on the selection process of a model's features, and the weighting their assigned, during the implementation of the system. The decisions made when defining categories and establishing the relationships between them affect both the accuracy of the model and the outputs that will be produced; especially when the systems are making recommendations that affect individuals.

According to the General Data Protection Regulation [11], one of the requirements is the "right of explanation" which is a right for all individuals to obtain "meaningful explanations of the logic involved" when "automated (algorithmic) individual decision-making", including profiling, takes place. Explainable AI will assist in making the decisions of machine learning algorithms more understandable and in this way help make the privacy implications of data sharing clearer to the end-users. Explainable AI is essential to build trust and increase the acceptance of smart home systems. Any inherent bias either in the data sets, in the weighting of various features or choosing which ones to use, has real-world results that must be fair and explainable by the system.

16.2.3 Privacy in Data Management

Generally speaking, data collected in smart home environments usually go through the following steps: first data collection, followed by data aggregation and finally data mining and data analytics. Although these steps enable a lot of services to be provided to the smart home users, they pose a lot of privacy challenges. For example, in the smart home environment if an adversary obtains data about the occupancy of the residents, he can infer the pattern of when users are inside or out of the house which could lead to theft or other damage to the users. Privacy-preserving mechanisms are necessary in these scenarios to protect the sensitive data and the privacy of the users.

Existing privacy-preserving mechanisms include anonymity, encryption, perturbation and privacy-preserving data analytics techniques such as differential privacy [9] and homomorphic encryption [22]. Recent techniques make use of deep learning mechanisms to provide privacy-preserving probabilistic inference [25], privacy-preserving speaker identification [26] and computing on encrypted data [4, 39]. While the previous work on privacy-enabled frameworks has focused on issues like assisting users with mobile app permissions [16], protecting user location data [14] and privacy-aware video streaming [8], the concerns and preferences of the users are not always taken into account.

Most of the time users realize what they want from a specific technology only after they have started using it. In most privacy surveys, the users make comments like the following: “I want to decide what types of data a service or application gets access to” and “I should get to decide who my data is shared with” [28]. Users have significant reservations about continuous ongoing monitoring [34] in which sensors collect measurements and can store and analyse all the sensor data in the cloud. Users prefer to have control over how their data are used. For example, if data from door sensors and light sensors are being combined to detect room occupancy, the users might want to be aware that there is this possibility of detection since most sensors do not appear too intrusive on their own, but when combined they can reveal a lot of information. Also, users may wish to give access to their data, but may prefer to do so anonymously, for example, as medical data for research. So it is essential for privacy-enabled frameworks to support anonymization. However, most of the time users often find it difficult to understand privacy controls. That is why privacy-enabled voice assistant can be beneficial in helping users manage their privacy preferences and offer them more control over which data they share and at what granularity. In this work, we extend beyond existing privacy preservation techniques and propose a framework that allows users to control the flow of information in a privacy-preserving manner.

16.2.4 Dialogue Systems

Lately, there has been tremendous growth in embodied conversational agents (ECAs) using natural language processing (NLP) techniques for developing intelligent dialogue components. NLP-based systems have gained a lot of interest in human–machine interactions for multi-modal interfaces. They are preferred widely for natural and spoken language understanding (NLU or SLU), natural language generation (NLG) and dialogue processing (DP) tasks.

A complete dialogue solution consists of various components which include automatic speech recognition (ASR), natural language interpreter, dialogue state tracker (DST), dialogue response selection (DRS) and text-to-speech (TTS) component [31], where each component has a specific task associated with it. Existing work on training dialogue managers includes rule-based methods; sequence-to-sequence methods; reinforcement learning (RL)-based methods; and hierarchical reinforcement learning (HRL) methods. In rule-based methods, a set of rules are defined by humans to train dialogue managers which make these systems robust and easy to use. However, it is not possible to use such systems for real-life applications. Sequence-to-sequence methods are used to transform a given sequence from one domain to another and are widely used in translation tasks such as English to French and performed well with less or none natural processing of sentences [36].

Reinforcement learning and hierarchical reinforcement learning have gained a lot of popularity due to their astonishing results in certain gaming tasks such as Atari, chess and GO [20] which outperform the human score but require a large amount of training data which are impractical to collect and simulate due to complexity and variety in human dialogues. HRL-based methods on dialogue management outperform the standard RL methods, but there are still unresolved issues. Namely, those issues are in handling deeper hierarchies; designing a reward for lower level hierarchies; and also automatically dividing the complex task into simpler sub-tasks which are possible with deeper hierarchies than few levels of hierarchy.

16.3 Proposed Framework

This section presents different modules of the smart home framework. An overview of the proposed smart sensing framework is presented in Fig. 16.1. The hardware includes an integrated sensor network with devices that can monitor multiple occupants and the home environment, a gateway and local storage device to collect the data, along with a smart interactive user interface that enables the communication with the user and the controlling of the smart home. In the following subsections, the main elements of the proposed framework are described.

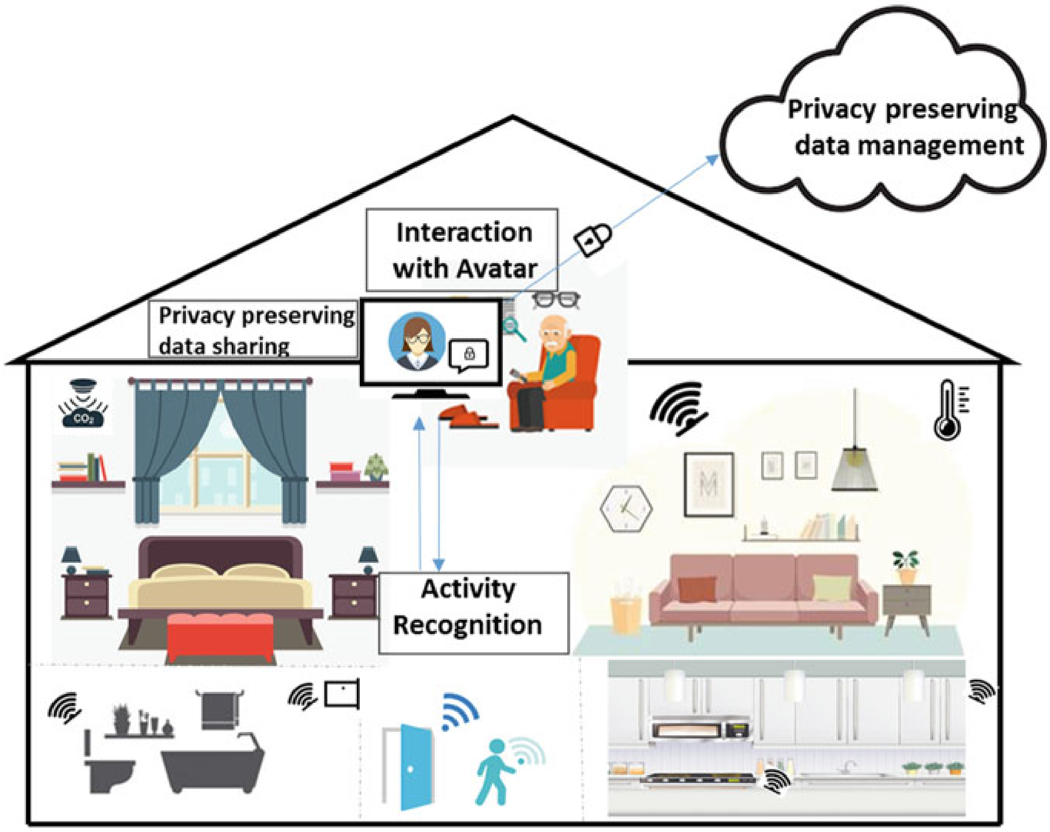


Fig. 16.1 Overview of the proposed framework

16.3.1 Activity Recognition and Occupancy Detection Module

The activity recognition module monitors daily living activities of the resident such as walking, sleeping, cooking, bathing and watching TV by using the unobtrusive sensor data from the home environment. The aim is to provide health care, comfort, safety and security to residents by detecting abnormal events and providing home automation services. The occupancy detection module of the framework identifies the resident's location and a number of occupants in home by using the data obtained from the sensors. The data obtained are then analysed and used to train machine learning algorithms for classification of various activities, resident behaviour analysis and detection of multiple occupancy in smart home.

One aspect often regarded is explainability in classification systems. Since many smart systems currently deploy deep learning models that are often treated as black box and can be hard to explain. Recent advances in explainability of deep learning systems will be also deployed in our proposed framework. Local Interpretable Model-Agnostic Explanations (LIME) [30] is widely used model-agnostic method to explain the model decision. It uses the trained model and by perturbing the input to the model it observes, how the prediction of model changes. LIME can be applied to image, categorical data or text. In our framework, since LIME is model-agnostic, we can deploy it for all classification tasks regardless of model specifications. Another method we will deploy is to explain our smart systems, which will be diving higher-level classification into smaller classification along with object and surrounding detection, which can give insight into why a certain decision is made. For example, if cooking is the main activity that is detected by our smart system, our system will also show low-level activities that are detected can be related to cooking. These low-level activities can be tap water is flowing, chopping sound detected, the oven is on, objects detected can be a knife, pan and potatoes and location of the user which can be the kitchen. These low-level activities can shed light, while cooking activity was detected also in the case or wrong classification of activity or not being able to classify an activity, again these low-level activities can explain the reason to the user. This methodology can be applied to all different part of the system whenever a decision is made by a smart algorithm.

16.3.2 Privacy-Enabled Voice Assistant Module

The privacy-enabled voice assistant allows the development of a more privacy-aware smart home that cannot only detect user activities and protect data. But it can also interact with the end-user in a meaningful way through a dialogue manager, which can learn and take into account their individual preferences not only about daily living activities but also about privacy. This includes interactions such as sometimes alerting users about data sharing they may not feel comfortable with, refining models of their user's preferences over time and at times prompting the users to carefully consider the possible implications of some of their privacy decisions.

This module is able to learn the user's preferences and help them manage their smart home and privacy settings across a wide range of devices and sensors without the need for frequent interruptions. In addition, with the privacy-preserving data management module, the users can automatically anonymize the collected data at the desired granularity and distribute or upload to the cloud. Different subsets of the data that are required by a third-party application or according to the access level of the receiver can be shared without providing access to the completely original data set.

16.4 System Architecture

The section presents a detailed description of the integration of each module of the framework and how the interaction and communication between them are performed. In the proposed framework, the Home Event Recognition System (HOMER) [10] acts as a middleware to integrate each module and sensor components of the system. HOMER is an open-source platform based on the Apache Karaf OSGi framework and enables modularity by encapsulating its functionalities in terms of OSGi bundles. The framework is flexible, extendable and adaptable to new components/modules. The components are in the form of OSGi bundles that can be remotely installed and updated without rebooting the system.

Figure 16.2 shows the system architecture of the proposed framework. The detailed description of each component is explained below.

16.4.1 Sensor Data and Pre-processing

The input to the system is the raw sensor data of the user and the environment from the smart home. The sensors used are motion sensors, door, window sensors, light sensor, temperature, humidity, passive infrared (PIR), pressure mats and CO_2 sensors. The data obtained are represented by a sequence of time-value pairs such as $\langle t_n, v_n \rangle$, $\langle t_{n+1}, v_{n+1} \rangle$ represents two consecutive pairs, where v_n spans the time interval $\left[t_n, t_{n+1}\right)$. The sampling rate varies on per sample basis depending on the sensors. Therefore, it is important to perform data pre-processing as it has a significant impact on the performance of machine learning models.

The data pre-processing includes removal of noisy data and outliers using data filtering techniques and handling of missing/incomplete values, which can be performed using interpolation methods to maintain completeness and consistency in time series. In the smart home systems, it is also very important to check if the data received are from the right sensor of a wireless sensor network. After the data

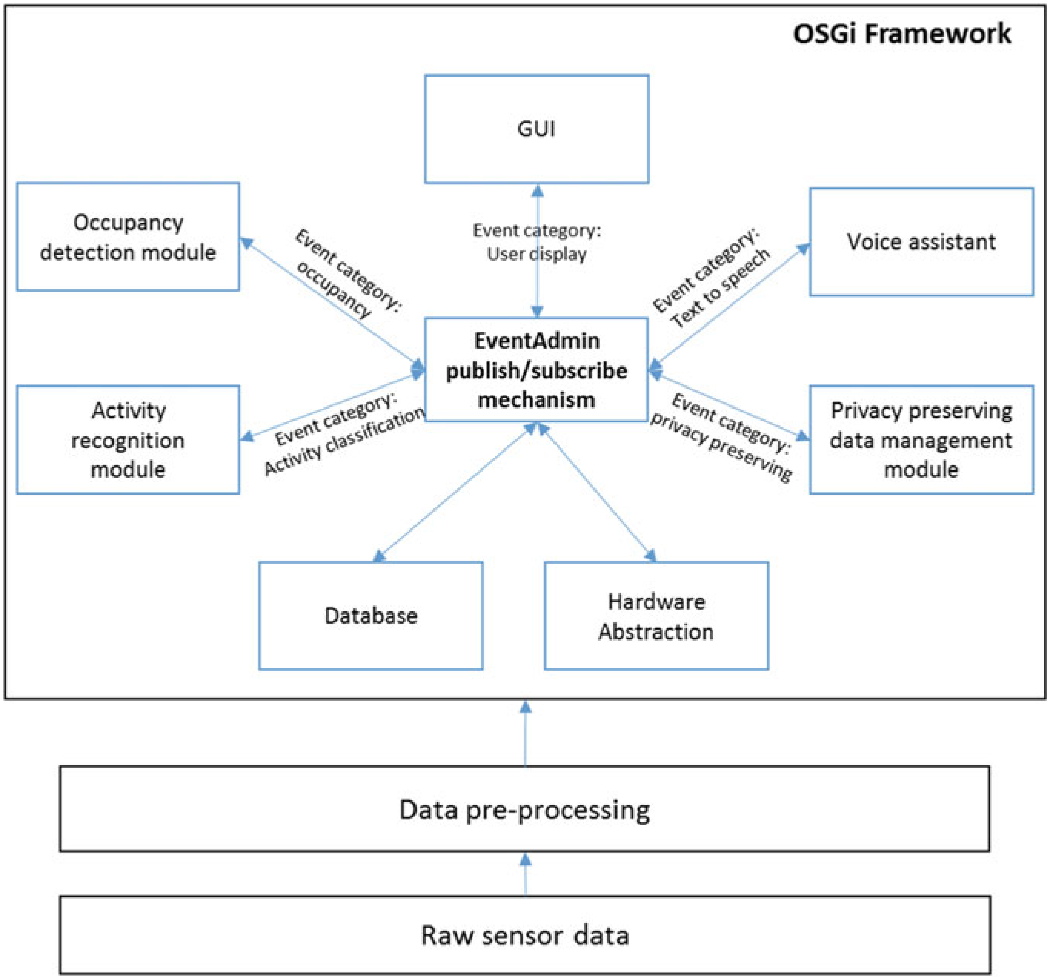


Fig. 16.2 System architecture

pre-processing is performed, the data are sent to the hardware abstraction bundle of OSGi framework.

The OSGi framework of HOMER ensures modularity and flexibility, which facilitates the parallel deployment of a bundle in the framework. Each bundle is separated and provides different functionalities. Each module of the OSGi framework is termed as bundles. As shown in Fig. 16.2, it consists of bundles, i.e. hardware abstraction, database, activity recognition, occupancy detection, data management, voice assistant and graphical user interface (GUI) bundles.

16.4.2 Hardware Abstraction

The hardware abstraction acts as an intermediate layer for harmonization of incoming sensor data from various sensor networks. It provides a mapping of non-standardized devices to the appropriate notation in ISO 11073 specification. The ISO/IEEE 11073

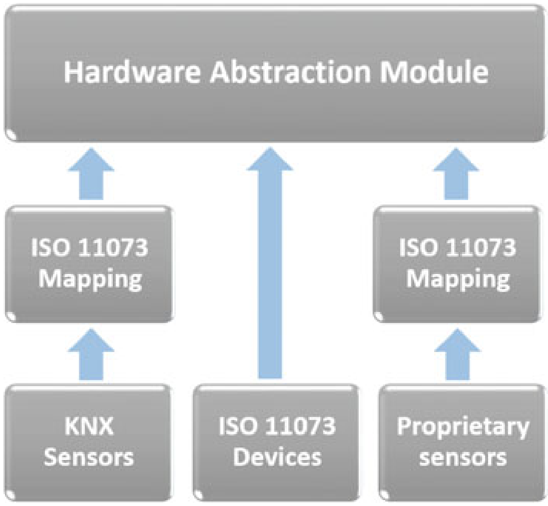


Fig. 16.3 Layout of hardware abstraction module [10]

standards enable the system to exchange sensor data between different medical devices and systems analysing these data. The home automation sensors are covered in the Independent Living Activity Hub specialization ISO/IEEE 11073-10471 [10]. This module provides the possibility of integration of non-intrusive off-the-self devices from different domains for data acquisition. The layout of the hardware abstraction module consisting of different bundles is shown in Fig. 16.3.

A hardware abstraction bundle is used as an abstraction layer between HOMER and different kinds of hardware. This enables hardware to distribute events generated by devices or to get commands for actuation. This way all types of hardware can use the same events for different devices.

16.4.3 Database

The database bundle stores all the sensor data in a standardized format. It contains information about the HouseID, RoomID, SensorID, sensor type, actuator ID, actuator and message type such as "ON" or "OFF". A data access interface provides the functionality to query the database for information retrieval or persistence.

16.4.4 Event Admin

All the bundles of the framework are interconnected with the Event Admin. The Event Admin is supplied by the OSGi framework and provides communication between the bundles by sending and receiving asynchronous events. It is based on a subscribe/publish mechanism in which bundles needs to subscribe to Event Admin in order to communicate and send/receive events within the framework. Each event is defined by a topic, which specifies the type of event and all the topics of the

```
{
    "topic": "hbs/acrossing/event/occupancy",
    "properties": {
        "amount": "2",
        "SOURCE_ID": "occupancy.detection.module",
        "event.topics": "hbs/acrossing/event/occupancy",
        "TIMESTAMP": 1552310208854
    }
}
```

```
{
    "topic": "hbs/acrossing/event/Privacy",
    "properties": {
        "LANGUAGE_CODE": "en",
        "SOURCE_ID": "privacy.module",
        "event.topics": hbs/acrossing/event/Privacy ",
        "TIMESTAMP": 1552310208870,
        "OUTPUT_DISPLAY_TEXT": "Meeting with John at 2pm at Bistro.",
        "OUTPUT_TEXT_TO_UI": " Meeting at 2pm."
    }
}
```

```
{
    "topic": "hbs/acrossing/event/Activity",
    "properties": {
        "LANGUAGE_CODE": "en",
        "SOURCE_ID": "Activityrecognition.module",
        "event.topics": hbs/acrossing/event/Activity ",
        "TIMESTAMP": 1552310208865,
    }
}
```

```
{
    "topic": "hbs/acrossing/event/Voiceassistant",
    "properties": {
        "LANGUAGE_CODE": "en",
        "SOURCE_ID": "voice.module",
        "event.topics": hbs/acrossing/event/Voiceassisant ",
        "TIMESTAMP": 1552310208870,
        "OUTPUT_DISPLAY_TEXT": "Meeting with John at 2pm at Bistro.",
        "OUTPUT_TEXT_TO_SPEECH": " Meeting at 2pm."
    }
}
```


Fig. 16.4 Example of event admin service specification

framework use the same prefix for identification followed by a category identifier. An event can also consist of one or more properties. A property is consisted of a key and its corresponding value. Examples of events are shown in Fig. 16.4.

Each of the event topics is linked to a specific category. The event categories in the proposed framework are occupancy, activity, privacy and voice assistant. A category consists of various event topics together with a message and set of properties, which enables communication and interaction between the different bundles.

16.4.5 Activity Recognition

The activity recognition bundle of the framework gathers processed and clean data from the database and performs a classification of the different activities of the user. In order to train the classifier, we prefer deep learning models such as long short-term memory (LSTM) and convolutional neural network (CNN) as they have the capability to learn features automatically through the network instead of manual handcrafting, which is a time-consuming and costly process. The Event Admin receives the classification output from the bundle and sends it to other bundles whenever it is required.

16.4.6 Occupancy Detection

The occupancy detection bundle uses a fusion of multiple non-intrusive sensors which have shown to be higher in accuracy, detection and occupancy estimation [7]. We prefer CO_2 sensors, motion sensors, PIR, temperature, light, door/window sensors installed in a home environment for occupancy estimation.

The occupancy detection can be seen as a multi-class problem with the aim to estimate the number of occupants in a home, and in order to train classifiers, recurrent neural network-based models have outperformed the existing standardized machine learning models. Since the data set can be unbalanced, we prefer F-score results to predict the performance of the classifiers. In case of multi-class problem, a macro-average method can be used, which is the average of precision and recall of each class label and then calculates the F-score.

16.4.7 Privacy-Preserving Data Management

The privacy-preserving data management bundle is used to sanitize the data before other modules access them. It anonymizes the data and removes personally identifiable information based on common personal information attributes as described in the Europe General Data Protection Regulation (GDPR). User preferences towards privacy have also taken into consideration, which we analysed according to the user studies [29].

In previous work [28], the LSTM encoder–decoder approach has been implemented where the encoder network takes the input sequence and maps it to an encoded representation (vector of fixed dimensionality). Then, the decoder network uses the encoded representation to generate output sequence. This makes the model a lock and key analogy, where only the correct key or decoder will able to access the resources behind the lock (encoder). The results on simulated smart home data set showed that method is able to learn privacy operations such as disclosure, generalization and deletion of data; thus generating different data views for data receivers in an encrypted way.

16.4.8 Voice Assistant

The voice assistant bundle handles the interaction between the user and the smart home in an efficient and comfortable way. The system consists of sequences of processes in which input to the system user's utterance and in return produce spoken output. A general architecture of a spoken dialogue system described in [27] consists of various components: first, the automatic speech recognition (ASR) component converts the raw audio input into a sequence of words (or the n-best results), which is then forwarded to a natural language understanding (NLU) component to extract the semantics of the utterance. A formal representation of semantics, generally a structured set of attribute-value pairs is used by the dialogue manager (DM) to decide the next actions to take according to the employed dialogue strategy. The action performed by DM is the generation of text from user utterance and transformation of text into speech by the text-to-speech engine. Other actions of DM are interaction with the back-end system or any kind of processing required by the application.

In this proposed framework, the dialogue manager module handles the interaction between the user and the smart home. In previous work, dialogue managers were developed using rule-based systems in order to be robust. The drawbacks of rule-based dialogue systems are that they require a human expert to design rules, which is time-consuming and makes the system static. Recently, a new model has been developed in a dialogue setting where dialogues are train with images instead of words [18]. This method handles words efficiently on which dialogue manager is not trained on and do not require pre-processing the data. We prefer this image based approach to train the dialogue manager.

16.4.9 Graphical User Interface (GUI)

The graphical user interface bundle provides graphical representation of the smart home architecture, which gives information about the user activities, statistics and analysis of activity data from the activity recognition module. It also provides information about user daily agenda/meetings and notifies the user for the upcoming event. The GUI of the system can be in the form of an avatar or mobile application on tablet, PCs or smartphone. All the information is presented in an abstract manner (for example: replacing the private activities such as bathing and sleeping with some unique notation; and anonymizing the personal information such as the name of the person with whom the meeting is scheduled, address and contact details) according to the privacy preferences of the user.

16.5 Example Use Cases of the Proposed Framework

The following section presents two different scenarios to illustrate how all the modules of the framework are interacting with each other in order to provide assistance to the resident.

16.5.1 Scenario 1

The first scenario examines the case where there is a reminder for a meeting that needs to be communicated to the user. In homes with multiple residents, it is a common phenomenon that there are different privacy preferences or events that people do not want to be shared. Therefore, in this instance, the framework makes use of the occupancy detection module to detect if the user is alone before communicating the details of the upcoming meeting. As a first step, the sensors installed in the home environment collect the data over a continuous period. Specifically, the sensors that were selected give data about the resident's location, the number of residents (one or more than one) and daily activities performed by the residents. Then, the dialogue manager module and privacy management module select the appropriate information

Table 16.1 Meeting reminder

Description	The use case describes the steps associated with the meeting reminder service. The voice assistant reminds the user of an upcoming meeting and depending on the circumstances reveals only specific information
Actors	User and voice assistant
Pre-conditions	The voice assistant has knowledge of the occupancy status and the events in the calendar of the user
Post-conditions	The user is reminded of an upcoming appointment without excess information being given in the presence of other people
Action sequence	1. The service starts with the framework detecting the occupancy status 2. If more than one person is present in the room, the voice assistant enables the privacy module, which hides not essential information 3. The voice assistant reminds the user of the upcoming meeting but does not mention the location, the identity of the other participants and the purpose of the meeting
Alternative sequence	1. The framework detects the occupancy status and the user is alone 2. The voice assistant reveals all the related information for the upcoming meeting
Requirements	The voice assistant has access to the smart home sensor data and the user's agenda

that needs to be communicated based on the result of the occupancy detection. Lastly, the voice assistant uses the dialogue manager to communicate the reminder to the user. Table 16.1 shows the procedure and interaction followed by the voice assistant.

16.5.2 Scenario 2

The second scenario examines the case of having guests in the smart home. There are many cases in which guests are not comfortable if they feel they are being monitored and they would like to be informed about the type of monitoring that takes place. User preferences show the users' expectations in regard to how their data should be managed by the smart home. The voice assistant can meet these preferences partially or completely depending on service policies and other user preferences existing in the same space. For example, user preferences in this scenario are:

Preference 1: Do not share the occupancy status of my house when guests are present.
Preference 2: Do not share my location with anyone.

The framework is able to deal with these privacy preferences by first detecting the presence through the occupancy detection module. The dialogue manager prompt user informs multiple occupancy and inquires user about their preferences towards monitoring. If the user disagrees with monitoring, then the privacy-preserving data management module applies user preferences in the privacy policy of the module and

Table 16.2 Guests privacy

Description	The use case describes the steps associated with the guest service. The voice assistant adjusts his privacy module to allow for the guests' privacy preferences
Actors	User, guests and voice assistant
Pre-conditions	The voice assistant has knowledge of the occupancy status
Post-conditions	The guests are informed about the presence of sensors and their privacy preferences are taken into account for the data collection and processing
Action sequence	1. The framework detecting the occupancy status and the presence of guests in the smart home 2. The voice assistant makes them aware of the sensors and that data collection occurring in the smart home and asks them if they agree with the monitoring 3. The guests do not feel comfortable with monitoring and they would like to have their data removed 4. The framework annotates the occupancy data during the time the guests were in the smart home and they are obfuscated before being processed by third parties 5. The voice assistant informs the guests that their data will be removed and turns itself off (not listening mode) until the occupancy module detects that the guests have left
Alternative sequence	1. The guests reply that they do not mind the monitoring as long as their identity remains hidden. 2. The framework annotates the data that was collected during that time as data that should be anonymized before shared with third parties
Requirements	Voice assistant has access to the smart home data of the resident and multiple occupancy

removes or anonymizes the data for some time until the user agrees with monitoring again. Table 16.2 presents the procedure and implementation of scenario 2. In both cases, the privacy-enabled voice assistant is able to handle the different privacy preferences of the users through its dialogue manager and the private data management module. This is a feature that has become more important lately and should be incorporated in all frameworks, as people become more aware of privacy and potential risks to it they have higher privacy expectations for the systems they use.

16.6 Conclusion and Prospects

The tremendous growth in smart home technology has led to improved home care through daily monitoring and automation. The data collected in the home environment raise various privacy and security concerns among the users. Therefore, in this chapter, we presented a secure smart home, which not only detects daily living activities of the user but also protects data and performs an interaction with the user in a meaningful way through the dialogue manager. The chapter presents the major components of the framework: activity recognition and occupancy detection; privacy-preserving data management and dialogue manager and propose methodologies to develop these modules. A system architecture of the proposed framework describes the communication and interaction between the modules in the HOMER middleware. One of the innovative aspects of the proposed framework is privacy-enabled voice assistant, which learns and takes into account users' preferences while sharing information and anonymizes the sensitive and personal information of the user depending upon the situation. In future work, implementation of the proposed framework will be performed and major focus will be on the integration of the individual modules in the OSGi framework so that system will be scalable, flexible and adaptable according to the different applications.

Acknowledgements This work has been funded by the European Union Horizon2020 MSCA ITN ACROSSING project (GA no. 616757). The authors would like to thank the members of the project's consortium for their valuable inputs.

References

1. Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on Explainable Artificial Intelligence (XAI). IEEE Access 6:52138–52160
2. Ahmad J, Larijani H, Emmanuel R, Mannion M, Javed A (2018) Occupancy detection in non-residential buildings—a survey and novel privacy preserved occupancy monitoring solution. Appl Comput Inform. https://doi.org/10.1016/j.aci.2018.12.001
3. Apthorpe N, Huang DY, Reisman D, Narayanan A, Feamster N (2018) Keeping the smart home private with smart(er) IoT traffic shaping. arXiv preprint arXiv:181200955
4. Bos JW, Lauter K, Naehrig M (2014) Private predictive analysis on encrypted medical data. J Biomed Inform 50:234–243
5. Chen H, Liu X, Yin D, Tang J (2017) A survey on dialogue systems: recent advances and new frontiers. ACM SIGKDD Explor Newsl 19(2):25–35
6. Chen L, Nugent C, Okeyo G (2014) An ontology-based hybrid approach to activity modeling for smart homes. IEEE Trans Hum Mach Syst 44(1):92–105
7. Chen Z, Jiang C, Xie L (2018) Building occupancy estimation and detection: a review. Energy Build 169:260–270
8. Das A, Degeling M, Wang X, Wang J, Sadeh N, Satyanarayanan M (2017) Assisting users in a world full of cameras: a privacy-aware infrastructure for computer vision applications. In: 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW), IEEE, pp 1387–1396
9. Dwork C (2011) Differential privacy. In: Encyclopedia of cryptography and security, pp 338–340
10. Fuxreiter T, Mayer C, Hanke S, Gira M, Sili M, Kropf J (2010) A modular platform for event recognition in smart homes. In: The 12th IEEE international conference on e-health networking, applications and services, IEEE, pp 1–6
11. GDPR (2018) General Data Protection Regulation (GDPR) final text neatly arranged. [online] Available at: https://www.gdpr-info.eu/. Accessed 16 Apr 2019
12. Holzinger A, Biemann C, Pattichis CS, Kell DB (2017) What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:171209923
13. Holzinger A, Kieseberg P, Weippl E, Tjoa AM (2018) Current advances, trends and challenges of machine learning and knowledge extraction: from machine learning to explainable AI.

In: International cross-domain conference for machine learning and knowledge extraction. Springer, Cham, pp 1–8

14. Jia R, Dong R, Sastry SS, Sapnos CJ (2017) Privacy-enhanced architecture for occupancy-based HVAC control. In: 2017 ACM/IEEE 8th international conference on cyber-physical systems (ICCPS), IEEE, pp 177–186
15. Jung Y (2017) Hybrid-aware model for senior wellness service in smart home. Sensors 17(5):1182
16. Liu B, Andersen MS, Schaub F, Almuhimedi H, Zhang SA, Sadeh N, Agarwal Y, Acquisti A (2016) Follow my recommendations: a personalized privacy assistant for mobile app permissions. In: Twelfth symposium on usable privacy and security (SOUPS 2016), pp 27–41
17. Machado E, Singh D, Cruciani F, Chen L, Hanke S, Salvago F, Kropf J, Holzinger A (2018) A conceptual framework for adaptive user interfaces for older adults. In: 2018 IEEE international conference on pervasive computing and communications workshops (PerCom Workshops), IEEE, pp 782–787
18. Merdivan E, Vafeiadis A, Kalatzis D, Henke S, Kropf J, Votis K, Giakoumis D, Tzovaras D, Chen, L, Hamzaoui R, Geist M (2018) Image-based natural language understanding using 2D convolutional neural networks. arXiv preprint arXiv:181010401
19. Mittelstadt B, Russell C, Wachter S (2018) Explaining explanations in AI. arXiv preprint arXiv:181101439
20. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S (2015) Human-level control through deep reinforcement learning. Nature 518(7540):p529
21. Monteriù A, Prist M, Frontoni E, Longhi S, Pietroni F, Casaccia S, Scalise L, Cenci A, Romeo L, Berta R, Pescosolido L (2018) A smart sensing architecture for domestic monitoring: methodological approach and experimental validation. Sensors 18(7):p2310
22. Naehrig M, Lauter K, Vaikuntanathan V (2011) Can homomorphic encryption be practical? In: Proceedings of the 3rd ACM workshop on cloud computing security workshop, ACM, pp 113–124
23. Okeyo G, Chen L, Wang H (2014) Combining ontological and temporal formalisms for composite activity modelling and recognition in smart homes. Future Gener Comput Syst 39:29–43
24. Park Y, Kang S, Seo J (2018) An efficient framework for development of task-oriented dialog systems in a smart home environment. Sensors 18(5):1581
25. Pathak M, Rane S, Sun W, Raj B (2011) Privacy preserving probabilistic inference with hidden Markov models. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 5868–5871
26. Pathak MA, Raj B (2013) Privacy-preserving speaker verification and identification using gaussian mixture models. IEEE Trans Audio Speech Lang Process 21(2):397–406
27. Pieraccini R, Huerta J (2005) Where do we go from here? Research and commercial spoken dialog systems. In: 6th SIGdial workshop on discourse and dialogue
28. Psychoula I, Merdivan E, Singh D, Chen L, Chen F, Hanke S, Kropf J, Holzinger A, Geist M (2018) A deep learning approach for privacy preservation in assisted living. In: 2018 IEEE international conference on pervasive computing and communications workshops (PerCom workshops), IEEE, pp 710–715
29. Psychoula I, Singh D, Chen L, Chen F, Holzinger A, Ning H (2018) Users' privacy concerns in IoT based applications. In: 2018 IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), IEEE, pp 1887–1894
30. Ribeiro MT, Singh S, Guestrin C (2016) Why should i trust you? Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, ACM, pp 1135–1144
31. Serban IV, Lowe R, Henderson P, Charlin L, Pineau J (2015) A survey of available corpora for building data-driven dialogue systems. arXiv preprint arXiv:151205742

32. Singh D, Merdivan E, Hanke S, Kropf J, Geist M, Holzinger A (2017) Convolutional and recurrent neural networks for activity recognition in smart environment. In: Towards integrative machine learning and knowledge extraction. Springer, Cham, pp 194–205
33. Singh D, Merdivan E, Psychoula I, Kropf J, Hanke S, Geist M, Holzinger A (2017) Human activity recognition using recurrent neural networks. In: International cross-domain conference for machine learning and knowledge extraction. Springer, Cham, pp 267–274
34. Singh D, Psychoula I, Kropf J, Hanke S, Holzinger A (2018) Users' perceptions and attitudes towards smart home technologies. In: International conference on smart homes and health telematics. Springer, Cham, pp 203–214
35. Sivaraman V, Gharakheili HH, Vishwanath A, Boreli R, Mehani O (2015) Network-level security and privacy control for smart-home IoT devices. In: 2015 IEEE 11th international conference on wireless and mobile computing, networking and communications (WiMob), IEEE, pp 163–167
36. Sutskever I, Vinyals O, Le, QV (2014) Sequence to sequence learning with neural networks. In: Advances in neural information processing systems, pp 3104–3112
37. Wang D, Yang Q, Abdul A, Lim BY (2019) Designing theory-driven user-centric explainable AI. In: Proceedings of the SIGCHI conference on human factors in computing systems CHI, vol 19
38. Wang J, Chen Y, Hao S, Peng X, Hu L (2019) Deep learning for sensor-based activity recognition: a survey. Pattern Recogn Lett 119:3–11
39. Xie P, Bilenko M, Finley T, Gilad-Bachrach R, Lauter K, Naehrig M (2014) Crypto-nets: neural networks over encrypted data. arXiv preprint arXiv:14126181
40. Yang J, Zou H, Jiang H, Xie L (2018) Device-free occupant activity sensing using WiFi-enabled IoT devices for smart homes. IEEE Internet Things J 5(5):3991–4002
41. Zheng S, Apthorpe N, Chetty M, Feamster N (2018) User perceptions of smart home IoT privacy. Proc ACM Hum-Comput Interact 2(CSCW):200

CHAPTER 4

Users' Perceptions towards AAL and Smart Home Technologies

4.1 Introduction

In recent years, there has been a significant advancement in AAL and smart home technologies, however, they have still not found their way into the care homes and daily life of the people. One of the major reasons is that such systems lack in addressing the end-user requirements of specifically older people, caregivers, and health care professionals. Acceptability of AAL systems including smart home relies on understanding end-users' perceptions, their specific needs, and concerns related to it. Therefore, we conducted interviews with older people at different caregiving organizations to perceive their requirements and concerns towards smart home. The first interviews were conducted at Zuyderland Medisch Zentrum, in Sittard-Herleen, Netherlands for 4 weeks, and the second interviews were performed at AKTIOS elderly care units, Greece for around 10 days. In addition to it, we conducted online surveys with participants from all around the globe to understand the attitudes and perceptions of the future smart home. The chapter includes three research studies, among them the first publication highlights the problems and needs of care home residents, caregivers, and professionals. The second publication presents the findings of the online survey focusing on users' attitudes towards indoor and outdoor monitoring and data sharing. Since privacy regarding data sharing is also one of the reasons that prevent the adoption of a smart home system, therefore the third publication focuses on identifying privacy concerns and issues from the perspective of the participants of the survey. Hence, these studies will help in the development of an acceptable smart home solution and providing better IoT services with minimized privacy risks to the public.

The content of this chapter is based on the publications:

Singh, D., Kropf, J., Hanke, S. and Holzinger, A., 2017, August. Ambient assisted living technologies from the perspectives of older people and professionals. In International Cross-Domain Conference for Machine Learning and Knowledge Extraction (pp. 255-266). Springer, Cham.

Contribution: Deepika Singh designed the user studies, questionnaires, interviewed the participants, acquired and analyzed the data. The manuscript was written by Deepika Singh and paper was presented at CD-MAKE 2017 conference in Regio di Calabria, Italy. The other authors contributed to the revision of the manuscript.

Singh, D., Psychoula, I., Kropf, J., Hanke, S. and Holzinger, A., 2018, July. Users' perceptions and attitudes towards smart home technologies. In International Conference on Smart Homes and Health Telematics (pp. 203-214). Springer, Cham.

Contribution: Deepika Singh designed the user studies, online questionnaires, acquired and analyzed the data. The manuscript was written by Deepika Singh and paper was presented at ICOST 2018 conference in Singapore. The paper received the best paper award at the conference. The other authors contributed to the revision of the manuscript.

Psychoula, I.*, **Singh, D.***, Chen, L., Chen, F., Holzinger, A. and Ning, H., 2018, October. Users' Privacy Concerns in IoT Based Applications. In 2018 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) (pp. 1887-1894). IEEE.

Contribution: Deepika Singh and Ismini Psychoula equally contributed in the publication and designed the user studies, online questionnaires, acquired and analyzed the data. The manuscript was written by Deepika Singh and Ismini Psychoula. The paper was presented at IEEE SmartWorld conference 2018 in Guangzhou, China. The other authors contributed to the scientific discussion and revision of the manuscript.

4.2 Publication II: Ambient Assisted Living Technologies from the Perspectives of Older People and Professionals

Ambient Assisted Living Technologies from the Perspectives of Older People and Professionals

Deepika Singh[1(✉)], Johannes Kropf[1], Sten Hanke[1], and Andreas Holzinger[2]

[1] AIT Austrian Institute of Technology, Wiener Neustadt, Austria
deepika.singh@ait.ac.at

[2] Holzinger Group, HCI-KDD, Institute for Medical Informatics/Statistics, Medical University Graz, Graz, Austria

Abstract. Ambient Assisted Living (AAL) and Ambient Intelligence technologies are providing support to older people in living an independent and confident life by developing innovative ICT-based products, services, and systems. Despite significant advancement in AAL technologies and smart systems, they have still not found the way into the nursing home of the older people. The reasons are manifold. On one hand, the development of such systems lack in addressing the requirements of the older people and caregivers of the organization and the other is the unwillingness of the older people to make use of assistive systems. A qualitative study was performed at a nursing home to understand the needs and requirements of the residents and caregivers and their perspectives about the existing AAL technologies.

Keywords: Smart homes · Ambient assisted living · Independent living · Ageing · Quality of life

1 Introduction

The aging population is the one of the major concerns of the world due to its direct socioeconomic implications. According to the United Nations' report, the number of people aged 65 years and above is expected to grow from 901 million in 2015 to 1.4 billion in 2030 [1,2]. According to the European population projections, it is expected that by 2040, one third of the elderly population will be aged 80 years and above. This demographic changes in the population is due to declines in fertility rate, continuous increases in the life expectancy and the retirement of the baby-boom generation [1,3].

Over aging leads to many problems ranging from basic functional disabilities to severe health problems e.g. osteoarthritis, diabetes, depression, chronic obstructive pulmonary diseases and dementia. In addition to the medical problems, the fact of being dependent on family members and/or care providers for their daily activities, cause embarrassment, social inactivity and poor nutrition etc.

A. Holzinger et al. (Eds.): CD-MAKE 2017, LNCS 10410, pp. 255–266, 2017.
DOI: 10.1007/978-3-319-66808-6_17

For long-term care and continuous assessment of physical and mental health in the older people there has been an increasing demand of nursing homes in the last decade [4,5]. However, this does not solve their problems completely; thus technology is the tool which can provide them an independent and happier life and at the same time accurate and timely personal care by the nursing home staff.

To overcome some of the mentioned problems, there has been a rapid development in ambient assisted living technologies (AAL) in Europe [6]. Different AAL solutions towards home monitoring, fall detection, social interaction have been developed using machine learning techniques. Technologies such as smart homes, assistive robots, mobile and wearable sensors have gained a lot of attention, but there are still many challenges that need to be addressed [7]. The concept of the smart home is a promising and cost-effective way of improving home care for the older people and the disabled in a non-obtrusive way, allowing greater independence, maintaining good health and preventing social isolation. There have been major advancements in the area of smart homes research that is enriching the home environment with technology (e.g., sensors and interconnected devices in a network) [8]. The most popular smart home projects are MavHome (Managing an Adaptive Versatile Home) [9], GatorTech [10], CASAS smart home [11], EasyLiving project [12] etc. The design of smart home depends on user requirement and living lifestyle of the resident; several approaches are proposed to identify activities of daily living of the resident of the home [13]. Although, the detection of a specific activity depends on the selection of appropriate set of sensors, data processing techniques and effective algorithms to understand daily lifestyle and human behavior [13,14]. The development of such systems lack in addressing the requirements of the older people and caregivers of the organization. One possibility to overcome this issue is proposed by applying extreme usability methods [15]. Additionally, various machine learning approaches have been used to develop an intelligent system for activity recognition; most common methods are Naive Bayes [16], Hidden Markov Model [17], and Conditional Random Field [18] classifiers. Despite the number of machine learning approaches, the accuracy of activity recognition is still not robust and unable to deal with uncertainty. There exist lot of challenges that need to be addressed in implementing effective solutions for older people [19].

To develop adaptable smart home technologies according to the specific needs of the older residents, their suggestions and inputs are much needed. A number of studies have been carried out where older people are participating in providing suggestions and opinions about the technologies [20]. Residents and family members indicated that they feel safe and had an overall positive attitude towards devices and installation of sensors [21] knowing that someone is monitoring them for their wellbeing [22]. The parameters of quality of life are usually evaluated by the older people on the basis of social contacts, dependency, health, material circumstances and social comparisons [23]. Certainly, AAL solutions and assistive technologies have a positive impact on different dimensions of health and quality of life. The needs and problems of older people can be addressed by applying appropriate solutions which influence the physical, mental and social dimensions of quality of life [24]. Nevertheless, the most important aspect is the acceptance of technology/devices by the older people and professionals [25].

Within the framework of this study, we wanted to know the different perspectives of the older people and care giving professionals of a particular nursing home. From the older people point of view we were interested in the following things: (i) how comfortable they are with the existing technologies (sensors, cameras, robots); any (ii) specific need they would require assistance or (iii) activities which they could imagine to be monitored; (iv) their problems and fears. From the caregiver point of view we were interested in the following: which data/information would be useful for them to provide accurate and timely attention to the patients. The study is aimed to understand the needs better so that more useful AAL solutions can be developed considering the inputs from the residents and caregivers.

2 Methodology

2.1 Planning

The research activity was planned at Zuyderland nursing home situated in Sittard-Geleen (The Netherlands) for 4 weeks. This nursing home is a part of Zuyderland Medical and Health care Group in the province of Limburg, Netherlands. The nursing home has a total of 273 residents and it is divided into small scale living apartments (48 residents), elderly care apartments (100 residents) and 3 apartments blocks for independent living (125 residents). The residents from the elderly care apartments and independent living apartment were contacted to participate in the study.

The research activities to be performed were divided for each week such that in the first week we collected information about the facilities which are provided to the patients in the hospital and the nursing home. The interviews with the residents and professionals of the hospital and the nursing home were conducted in the second week. The next two weeks were spent on transcribing the answers of the interview questions and analysis of the data obtained from the interviews. The aim of the study was to acquire knowledge about the daily living lifestyle of the old people and to investigate the needs, problems they face. In addition, it was also beneficial to know from the opinions of residents and caring staff about various existing ICT solutions and ambient assisted technologies including Smart Care Home. The caregivers had prior knowledge of existing AAL technologies from former research projects.

2.2 Focus Group and Sampling

Care professionals and residents of the nursing home were contacted to ask for the participation in the study. The aim was to enroll large number of people who can best discuss and share their experiences. The residents for the study were chosen by the nursing home's caring staff since they were knowing their exact medical conditions. For example, this activity should not to disturb any of the medical routines of residents, therefore, the residents suffering from severe chronic diseases (such as cancer and last stage heart diseases), and those with pacemakers and other required monitoring electronic devices such as ECG, pulse oximetry etc. have been excluded from the study. The focus group of the study was composed of residents and professionals as specified below:

1. Residents from the independent living apartments
2. Residents from elderly care apartments
3. Professionals include Physiotherapist, Occupational therapist and Dietician
4. Caring staff of the nursing home
5. Innovation and development experts

The total sample size was consisting of seven residents ($n = 7$) and six professionals ($n = 6$). The characteristics of all the 13 participants interviewed are shown in Tables 1 and 2. We have assigned code to each participant such that *"R"* and *"P"* denote resident and professional respectively. The average age of the residents was 80 years and the average experience of the professional in dealing with problems of older people was 17 years.

Table 1. Demographics of the residents

ID	Age (years)	Gender	Work Status
R1	90	Male	All of them were retired
R2	82	Male	
R3	80	Female	
R4	74	Female	
R5	65	Male	
R6	86	Female	
R7	81	Female	

Table 2. Details of professionals

ID	Age (years)	Gender	Specialization	Experience with elderly(in years)
P1	53	Female	Dietician	20
P2	39	Male	Physiotherapist	16
P3	33	Female	Occupational therapist	11
P4	37	Male	Caregiver	21
P5	39	Male	Innovation expert	17
P6	40	Male	Innovation expert	18

2.3 Data Collection

The interview questions for the residents were finalized after having discussions with care home professionals ranging from nurses (who are in direct contact with the residents) to the innovation experts and vice-versa. The inputs from them were well addressed in the final questionnaires in both the cases. As already stated, the final aim of the interviews was to answer the following questions:

- What are the problems faced by the elderly in performing the daily living activities?
- How smart care homes can contribute to the needs and requirements of the elderly and caregivers?
- An opinion about existing AAL technologies and how it can be improved?

In the beginning of the interview, the objective of the study were explained to the participants and they were told that they can leave or interrupt the interview whenever they feel any discomfort and/or unwillingness to answer. Since the interview were conducted in an informal way of talking, the timing was varied from 25 to 60 min. Consents were taken for audio recording the sessions so as not to miss any important points and actual responses can be quoted. All the participants agreed to audio record the interviews and the data collected were transcribed verbatim for the accuracy immediately after the interview.

The questions were framed in the simplest way for sake of clarity; they were also allowed to seek any sort of clarification needed. In order to make the residents comfortable, the interviewer started with an informal conversation, e.g.
Q: What are your hobbies?
Q: What do you like to do in your free time inside the home?

And the interview progressed in a manner of fluent conversation. However, the framed questions were asked clearly in the running conversation to be consistent with the responses so that it may not lead to any ambiguity in comparison.

The interviews with the residents were performed in two groups, one group was the residents from independent living apartments *(R1–R4)* and other group was of residents from the elderly care apartments *(R5–R7)*. Interviews with professionals and caregivers were conducted on one-to-one basis.

2.4 Ethics

The research was conducted with the prior permission and after the approval from the hospital and care organization authorities. To ensure confidentiality, an agreement was signed by the researcher with the organization. Full anonymity was promised to all the participants.

2.5 Data Analysis

The transcribed data was matched with the audio taped version to remove any discrepancies. Data analysis was performed using qualitative content analysis approach [26,27]. Analysis process includes various steps:

- Organizing and collection of data;
- repeated reading of the data;
- look for meaningful data and labeling it into codes;
- grouping the codes in subcategories;
- grouping subcategories into categories to generate themes.

To ensure the rigor in this study following criteria have been fulfilled: (i) getting familiar with the residents to form trusting relationships, verifying responses with the participants (credibility); (ii) selection of the participants (dependability); (iii) using extracts from the interviews to support findings (transferability); (iv) and establishing an audit trail (confirmability) [28].

3 Findings and Discussions

With consideration to the care needs, we decided to perform a thematic analysis focusing on different important aspects of healthy living; three themes were identified: *daily living, social engagement and technology*. The themes have different categories and subcategories, as shown in Table 3. We have also included some quotes from residents and professionals in the discussion section to highlight the exact needs without generalization.

Table 3. Care needs of the elderly

Themes	Categories	Subcategories
Daily Living	Problems and needs	Eating
		Cooking
		Grooming
		Housekeeping
		Bathroom usage
	Psychological care	Privacy
		Safety
Social Engagement	Participation in social activities	Social events in elderly home
		Special events
		Visit with family members
	Physical activities	Exercise
		Indoor activities
		Outdoor activities
Technology	Adaptability with technology	Devices
		Robots
		Security cameras

3.1 Daily Living

In our study we observed the daily living lifestyle of the residents in the nursing home. All the residents have their own private apartments in assistive environment with 24×7 availability of the caring staff.

The residents have different problems and needs which are mainly defined by their health conditions. The residents suffering from some physical disabilities, mild chronic diseases and fractures, find difficulty in performing most of their daily activities (such as going to toilet, showering, walking or moving inside their apartment, eating), therefore, they seek more assistance from caregivers. The caregivers help the residents in order to follow their daily routine.

P3: "I watch them and analyze that they find difficulties in moving or walking, I look for possibilities so that person can move in their room and actively participate in the environment either with care or some aids."

P3: "The first thing which the patients want to do without any assistance is going to the toilet; as dependence on care givers brings the sense of embarrassment among the patients."

The older people sometimes do not express their urges out of embarrassment which causes poor nutrition and imbalance diet.

R6: "I cannot pick something from refrigerator, or from table because I cannot move. Every time I have to ring and call the nurse"

P1: "I met a patient who told me that I eat less, because I have call somebody to get me something to eat and drink"

In some cases, they need assistance even for small things such as picking up bottle for drinking, cloths from the wardrobe etc.

P2: "Residents also need assistance in household activities like closing the door, curtains, turning off the TV and all these activities takes lot of time of the nurses"

From the interview we found out the major factors which affect the patients physiologically are: feeling of embarrassment; helplessness and loneliness; depression and lack of motivation.

R5: "My hobbies are painting but I do not have motivation of doing it"

One of main challenges for the caregivers is monitoring the resident who do not interact with anyone and keep themselves inside the apartment. In that case, the caregivers cannot track the patients daily activities and prescribed dietary plan. In addition, they sometimes lie and even refuse to take any help by the caregivers.

P1: "A patient answered when offered help: Oh No, I dont need anything. Dont do that, dont make anything for me"

3.2 Social Engagement and Physical Activities

Engagement of the residents in social and physical activities depends on the personal choices, interests and their health conditions. From the interviews, we analyzed that most of the residents participate in the social activities organized within the nursing home but only very few of them go out for some physical activity, walking or swimming. They like to go out only with family member and on some special occasions. This unwillingness could be due the fear of fall, fear of exhaustion and thus always need a companion with them.

There are few challenges with the residents suffering from severe chronic disease, disabilities; such residents do not like to participate mainly due to the feeling of helplessness and embarrassment. The other major factor which cause less physical activity and less exercise is due to the pain.
P2: "Pain has a strong influence on daily living and most important domain where we can act as physical therapist. Would be good if I know in which activities and movement patients suffers pain"

The residents participate in the exercise sessions and trainer motivates them to continue with their exercise. But sometimes exercise during the sessions is not just enough and they are advised to follow a routine. However, it is very difficult to know from the older people that whether they are exercising in their homes or not?
P2: "I do not know whether the patient exercise in their own home or not"
The less physical activity also causes poor nutrition; as the patients do not consume food and/or liquid if they remain inactive for a longer duration.

From the perspective of professionals, quality of life can be improved by motivating and engaging older people in different social activities of their interest which keep them physically active, thus decrease depression and feeling of loneliness.

3.3 Technology

The adaptability with technology is one the most important point to be considered while designing a smart home. It does not solve the purpose if a highly sophisticated device is provided but they do not feel comfortable in using it. We found out from the responses of the older people that they find the technology beneficial and useful especially from the safety point of view; but they have their preferences and restrictions.
P2: "When residents see everybody is using it then they use it"

Table 4. Technology

ID	Smart phone	Tablet	Laptop/Computer
R1	No	Yes	No
R2	No	Yes	Yes
R3	No	Yes	No
R4	Yes	Yes	No
R5	Yes	Yes	Yes
R6	No	Yes	No
R7	No	No	Yes

Table 4 highlights the comfort levels of the older people in using various devices among smart phone, tablet and laptop/computer. As it is clear from the table, among all the devices, majority of the residents find tablet more comfortable in using than smart phones and computer.

In the study, we also inquired from the residents whether they are comfortable with robots and they like to see robots assisting them in daily activities in their home. Most of them disagreed with having robots inside their apartments. Residents pointed out that they are more comfortable in using tablets and agreed in using wearable devices but do not want big robots around them.
R4: "No! I do not want robot inside my apartment"
R1:"I cannot cuddle it and they do not give hugs"

In concern with the home security, we asked whether they like to put cameras outside their apartment door to see who visited them, in their absence. The residents strictly denied that they do not want to put camera inside and outside their apartment.
R2: "If someone wants to meet me, he/she can come again"
From the perspective of the older people and professionals, a personalized intelligent assistive technologies will be desirable which can provide independence, ensure safety, and should be adaptable according to needs of the residents.

3.4 Suggestions/Recommendations

As one of the main motives of the study was to seek suggestions from the residents and professionals for making the nursing home smart.
P1: "when I know how active the patient is, it would be nice for me to know how much energy he/she spent per day and will help me in making diet plan and improve accordingly. Now it is always guess and I have to ask patient every time"
P1: "Would be good if there is some system which help in cooking and preparing the meals. In kitchen, if stove get automatically turned off when not in use"
In general, from the caregivers perspective, some family members of the residents want to keep a track of their health status and daily updates from the nursing home. A system or application which sends out daily notifications to family members and notify them in case any emergency, would be really desirable.
P4: "Would be good to have system for caregiver and resident in the apartment which controls the lighting, door open/close, curtains open/close and inside temperature level"
P3: "If I could know information about small activities from smart home like whether patient is going to bathroom by their own, getting meal or drinks from the kitchen, would be very helpful"
R3: "It would be good if I know my sleeping patterns and activity level inside my home, so I can show that data to my doctor"

From the perspective of the residents safety is the major concern. In the nursing home, residents have a personal alarm system, which they find very useful and beneficial. They use it for calling the caregiver in case of any assistance required and emergency inside the home. However, the residents recommended that such kind of system would be really useful, if they could use it when they are out. In that case, they would feel safer in going out alone, which enhances physical activeness and social involvement.

4 Conclusion

The findings of our study provide insights into the problems and needs of the residents of the nursing home. It also highlights the challenges faced by the caregivers during monitoring of the older people. Suggestions and recommendations have been pointed out from both the sides aiming at an independent and confident life of the residents. We summarize them as follows:

Although, the needs and problems are varied according to the individuals needs. The residents from the independent living apartments *(R1–R4)* do not really want home automation as they can perform basic daily activities. They think it can make them less physically active. On the other hand residents from the elderly care apartments *(R5–R7)* get regular assistance by the caregivers and there is evident need of a smart care home to reduce less personal assistance in basic household activities such as lock/unlock doors, opening/closing window shields, lights controlling etc. However, all the residents agreed to have a system which could monitor their physical activities. Such system would be really useful for the caregivers in order to know the activity level of the residents and deviations from the normal behavior. From patients and caregivers point of view, information about the following would be really useful: walking patterns, sleep analysis, real time location, fall detection and physical activity level.

All the participants agreed that assistive technologies and AAL solutions can have beneficial effects on quality of life and health. Majority of the residents feel comfortable in using tablet over smart phone and laptop. Big robots and cameras are not preferred by the residents; but they are open for wearable devices. On contrary, the professionals find robots a valuable contributions in smart care homes.

Based on the findings of this study it is recommended that a personalized smart home solution which can monitor daily living activities would be really useful for such nursing homes and even for private homes where old people live alone. Such system should be capable of detecting the users activity level and thus sends out recommendations to perform some physical activity when the user is inactive for a while. It can also notify the care staff and/or family members about the abnormalities.

Acknowledgement. This work has been funded by the European Union Horizon2020 MSCA ITN ACROSSING project (GA no. 616757). The authors would like to thank all the participants for their contributions; Esther Veraa, Cindy Wings, Maarten Coolen from Zuyderland, Sittard-Geleen, Netherlands for their valuable inputs and immense support.

References

1. DESA: United nations department of economic and social affairs/population division: world population ageing 2015 (2015)
2. Kleinberger, T., Becker, M., Ras, E., Holzinger, A., Müller, P.: Ambient intelligence in assisted living: enable elderly people to handle future interfaces. In: Stephanidis, C. (ed.) UAHCI 2007. LNCS, vol. 4555, pp. 103–112. Springer, Heidelberg (2007). doi:10.1007/978-3-540-73281-5_11
3. Molinuevo, D.: Services for older people in Europe, October 2008
4. Bettio, F., Verashchagina, A.: Long-term care for the elderly. Provisions and providers in 33 European countries, November 2010
5. Ribbe, M.W., Ljunggren, G., Steel, K., Topinkova, E., Hawes, C., Ikegami, N., Henrard, J.C.: Nursing homes in 10 nations: a comparison between countries and settings. Age Ageing **26**(suppl 2), 3–12 (1997)
6. Blackman, S., Matlo, C., Bobrovitskiy, C., Waldoch, A., Fang, M.L., Jackson, P., Mihailidis, A., Nygård, L., Astell, A., Sixsmith, A.: Ambient assisted living technologies for aging well: a scoping review. J. Intell. Syst. **25**(1), 55–69 (2016)
7. Rashidi, P., Mihailidis, A.: A survey on ambient-assisted living tools for older adults. IEEE J. Biomed. Health Inform. **17**(3), 579–590 (2013)
8. Chan, M., Campo, E., Estève, D., Fourniols, J.Y.: Smart homes current features and future perspectives. Maturitas **64**(2), 90–97 (2009)
9. Das, S.K., Cook, D.J., Battacharya, A., Heierman, E.O., Lin, T.Y.: The role of prediction algorithms in the mavhome smart home architecture. IEEE Wirel. Commun. **9**(6), 77–84 (2002)
10. Helal, S., Mann, W., El-Zabadani, H., King, J., Kaddoura, Y., Jansen, E.: The gator tech smart house: A programmable pervasive space. Computer **38**(3), 50–60 (2005)
11. Rashidi, P., Cook, D.J.: Keeping the intelligent environment resident in the loop (2008)
12. Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., Shafer, S.: Multi-camera multi-person tracking for easyliving. In: Third IEEE International Workshop on Visual Surveillance, Proceedings, pp. 3–10. IEEE (2000)
13. Skubic, M., Alexander, G., Popescu, M., Rantz, M., Keller, J.: A smart home application to eldercare: current status and lessons learned. Technol. Health Care **17**(3), 183–201 (2009)
14. Ni, Q., García Hernando, A.B., de la Cruz, I.P.: The elderlys independent living in smart homes: a characterization of activities and sensing infrastructure survey to facilitate services development. Sensors **15**(5), 11312–11362 (2015)
15. Holzinger, A., Errath, M., Searle, G., Thurnher, B., Slany, W.: From extreme programming and usability engineering to extreme usability in software engineering education (xp+ue/spl rarr/xu). In: 29th Annual International Computer Software and Applications Conference, COMPSAC 2005, vol. 2, pp. 169–172. IEEE (2005)

16. Tapia, E.M., Intille, S.S., Larson, K.: Activity recognition in the home using simple and ubiquitous sensors. In: Ferscha, A., Mattern, F. (eds.) Pervasive 2004. LNCS, vol. 3001, pp. 158–175. Springer, Heidelberg (2004). doi:10.1007/978-3-540-24646-6_10
17. Nguyen, N.T., Phung, D.Q., Venkatesh, S., Bui, H.: Learning and detecting activities from movement trajectories using the hierarchical hidden Markov model. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 2, pp. 955–960. IEEE (2005)
18. Nazerfard, E., Das, B., Holder, L.B., Cook, D.J.: Conditional random fields for activity recognition in smart environments. In: Proceedings of the 1st ACM International Health Informatics Symposium, pp. 282–286. ACM (2010)
19. Sun, H., De Florio, V., Gui, N., Blondia, C.: Promises and challenges of ambient assisted living systems. In: Sixth International Conference on Information Technology: New Generations, ITNG 2009, pp. 1201–1207. IEEE (2009)
20. Dongen, J.J.J., Habets, I.G.J., Beurskens, A., Bokhoven, M.A.: Successful participation of patients in interprofessional team meetings: a qualitative study. Health Expectations (2016)
21. Demiris, G., Rantz, M.J., Aud, M.A., Marek, K.D., Tyrer, H.W., Skubic, M., Hussam, A.A.: Older adults' attitudes towards and perceptions of smart home-technologies: a pilot study. Med. Inform. Internet Med. **29**(2), 87–94 (2004)
22. Alam, M.R., Reaz, M.B.I., Ali, M.A.M.: A review of smart homes past, present, and future. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **42**(6), 1190–1203 (2012)
23. Netuveli, G., Blane, D.: Quality of life in older ages. Br. Med. Bull. **85**(1), 113–126 (2008)
24. Siegel, C., Hochgatterer, A., Dorner, T.E.: Contributions of ambient assisted living for health and quality of life in the elderly and care services-a qualitative analysis from the experts perspective of care service professionals. BMC Geriatrics **14**(1), 112 (2014)
25. Holzinger, A., Schaupp, K., Eder-Halbedl, W.: An investigation on acceptance of ubiquitous devices for the elderly in a Geriatric Hospital environment: using the example of person tracking. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 22–29. Springer, Heidelberg (2008). doi:10.1007/978-3-540-70540-6_3
26. Elo, S., Kyngäs, H.: The qualitative content analysis process. J. Adv. Nurs. **62**(1), 107–115 (2008)
27. Miles, M.B., Huberman, A.M.: Qualitative Data Analysis: An Expanded Sourcebook. Sage, Thousand Oaks (1994)
28. Krefting, L.: Rigor in qualitative research: the assessment of trustworthiness. Am. J. Occup. Ther. **45**(3), 214–222 (1991)

4.3 Publication III: Users' perceptions and attitudes towards smart home technologies

Users' Perceptions and Attitudes Towards Smart Home Technologies

Deepika Singh[1,3](✉), Ismini Psychoula[2], Johannes Kropf[1], Sten Hanke[1], and Andreas Holzinger[3]

[1] AIT Austrian Institute of Technology, Wiener Neustadt, Austria
deepika.singh@ait.ac.at
[2] School of Computer Science and Informatics, De Montfort University, Leicester, UK
ismini.psychoula@dmu.ac.uk
[3] Holzinger Group, HCI-KDD, Institute for Medical Informatics/Statistics, Medical University Graz, Graz, Austria

Abstract. The concept of smart home is a promising and efficient way of maintaining good health, providing comfort and safety thus helps in enhancing the quality of life. Acceptability of smart homes relies on the users' perceptions towards its benefits and their concerns related to monitoring and data sharing. Within this study, an online survey with 234 participants has been conducted to understand the attitudes and perceptions of future smart home users, followed by detailed analysis of their responses. In general, the users agree that the smart home technology would improve the quality of life to a greater extent and enhance the safety and security of residents. On the contrary, they raise several concerns such as the increased dependence on technology and the monitoring of private activities, which may be seen as perceived drawbacks. The obtained results show that the older adults (ages from 36 to 70 years) are more open to monitoring and sharing data especially if it useful for their doctors and caregivers while the young adults (ages up to 35 years) are somewhat reluctant to share information.

Keywords: Smart home · Users' perspective · Acceptability
Data sharing

1 Introduction

In recent years, the concept of assistive technology has been developed tremendously to facilitate self-care, enhance independence, provide comfort and improve the quality of life of the individuals. With the rapid increase in aging population around the world, smart home technology has gained a lot of attention due to its versatile applications in the area of Internet of Things (IoT). It is defined as a living environment where all the devices in the home have the capability to interact with each other and also with the occupants living inside [1]. These devices include smart appliances (refrigerators, washers, TVs etc.), security and

M. Mokhtari et al. (Eds.): ICOST 2018, LNCS 10898, pp. 203–214, 2018.
https://doi.org/10.1007/978-3-319-94523-1_18

safety systems (sensors, monitors, cameras and alarms) and smart energy equipment (thermostats and smart lighting) which are interconnected using standardized communication protocols [2]. There has been major growth in the market of smart home devices owing to the growing aging population, rising demand for home health care, assisted living and energy consumption [3]. According to the report in [4], the global market of smart homes is expected to reach USD 53.45 billion by 2022 and industry analysis shows compound annual growth rate (CAGR) of 14.5% between 2017 and 2022.

Despite many applications and the variety of features, smart homes are still not capable of fully incorporating the demands of the users. High costs and long device replacing cycles are the potential market barriers in the complete adoption of smart homes. Furthermore, privacy and trust related issues with the data collected by the smart home devices are the major challenges [5]. Along with the privacy protection, the actual needs and concerns of future smart home inhabitants need to be considered. Various studies have been conducted with older adults to know their requirements, concerns and perceptions for smart homes [6]. The results of the surveys showed the interest of the participants in the assistive technologies and the necessity of smart home technology for independent living, safety and better quality of life [7,8]. However, in another study [9] participants expressed a variety of concerns including usability, reliability, accessibility and absence of public policy at the state or federal level promoting smart home technology adoption for aging population. According to the national survey of UK homeowners [10], policymakers can play an important role in mitigating perceived risks of smart home technologies by supporting design and operating standards, guidelines on data security and privacy, quality control and energy management in future smart homes. Users' attitudes have also been analyzed towards different type of assisted living services [11]; the study reports that people with many social contacts and high interest in technology showed acceptance for electronic services at home. The results for the different applications were insensitive to gender and age. The wide acceptance of smart homes depends on the way it serves the needs and demands of the user to the best possible extent. In order to do so, opinions of the end users irrespective of their demographics are of great importance. Within the current study, an online survey with a detailed questionnaire was conducted. The questionnaire focuses on identification of needs in performing daily living activities; users' attitudes towards monitoring inside and outside the home and their views regarding robots or personal assistant. There were also questions related to data sharing e.g. which data they would like to share with their doctor. The main aim of the survey was to find out the users' attitudes and their concerns in adopting a smart home.

The paper is structured into four sections. First section introduces the topic and the aim of the study and second section presents the methodology and performed analysis of the responses obtained from the participants. Third section reports the results and discussions and the last section concludes the study and highlights the major findings.

2 Methodology

In this section, the questionnaire design, employed statistical procedures, and demographics of the sample are explained. An online survey was used for this study and data was collected during the period of January - February 2018. The aim was to understand the individuals' attitude towards smart home technologies (their benefits and drawbacks) as well as opinions towards data sharing. The study was approved by the Ethics Committee of De Montfort University in accordance with the Code of the British Computer Science Society and the Chartered Institute for IT. Individuals were asked to participate on a voluntarily basis and the average time taken to complete the survey was approximately 15 min. Full anonymity of the participants was maintained during the study.

In the survey, the questionnaire was divided into six different sections. The first section contained socio-demographic questions (participants' age, gender, location, education and computer or IoT knowledge). The second section of the survey consisted of questions related to the users' ease of performance in daily living activities and their thoughts regarding smart home features. The third section was related to social engagement and devices for outdoor monitoring. The fourth section comprised of questions related to a personal assistant (such as an avatar) and views regarding user interfaces. Benefits, concerns and drawbacks of smart home technologies were asked in the fifth section. The last section of the survey included questions related to users' perception towards monitoring daily living activities and health related data.

2.1 Data Collection

The survey was distributed through e-mails, social media and university groups with an aim of reaching as many respondents as possible (convenience sampling). A sample size of N = 234 was chosen for the analysis; which includes individuals of various backgrounds spread across the globe as shown in Table 1.

The sociodemographic variables used are described as follows:

- Gender: We included gender as a demographic variable to see if there are differences in the perception and acceptance of smart home technologies and data sharing between male and female. The gender breakdown achieved in the study is 58.1% male and 41.9% female. However, the ratio may differ from the real world demographics but the aim was to have significant number of participants from each gender.
- Age: Age is usually negatively correlated with acceptance of technologies thus it was included as a factor to indicate if there is an interest on the adoption of smart home technologies from the upcoming generations. During the survey, we have classified participants into five age groups (as shown in Table 1), which we have classified further on a broader level for the analysis as young adults (ages up to 35 years) and older adults (ages from 36 to 70 years).
- Location: Location was included at the level of continent to show if there are differences in the perception of people from different parts of the world.

- Education: In the previous years, there has been an assumption that educated people who are exposed to advanced technology on a regular basis might accept technology more eagerly than less educated people. However, recent trends indicate that this assumption might no longer be valid due to ease and increased user-friendliness of new technology.
- Computer and IoT Devices: Comfortability and acceptance of technology in daily life varies from person to person, therefore, the participants were asked to classify themselves on their level of computer knowledge and list the IoT devices they already own, if any.
- Familiarity with smart homes: The familiarity with smart homes factor was added to show if the participants are already aware of this technology and if they are familiar with its operation and functions in real life.

Table 1. Demographic breakdown of the participants (N = 234)

		Count	Table N%
Gender	Male	136	58.1%
	Female	98	41.9%
Age	Under 18	5	2.1%
	18–24	58	24.8%
	25–35	100	42.7%
	36–55	48	20.5%
	56–70	23	9.8%
Which of these best describes your location?	Asia	100	42.7%
	Africa	2	0.9%
	North America	24	10.3%
	South America	3	1.3%
	Europe	101	43.2%
	Australia/Oceania	4	1.7%
What is the level of your highest education?	Not completed school	7	3.0%
	Completed school	35	15.0%
	University degree	99	42.3%
	Postgraduate degree	93	39.7%
How would you classify yourself as a computer user?	Beginner	1	0.4%
	Basic Knowledge	22	9.4%
	Moderate	88	37.6%
	Expert	123	52.6%

2.2 Statistical Analysis

For the analysis, Microsoft Excel and IBM SPSS Statistics were used to generate the descriptive statistics of the data and the item-level results of each question of the survey. In addition, Cronbach's alpha value is calculated to test the reliability of the survey with $\alpha = 0.740$. A significance level of 5% is used for all the analyses.

3 Results and Discussion

In the beginning of the survey, participants were asked about familiarity with the concept of smart homes and IoT devices used by them. The data obtained showed that the participants from Europe, Americas (South America and North America) and Australia are more familiar with smart home technologies (Europe = 63%, Americas = 66.6% and Australia = 75%) as compared to participants from Asia (43%). The familiarity with smart homes could also be related with the use of various IoT devices by the participants. As can be seen from Table 2, participants from Europe, Americas and Australia use presence sensors, voice assistant and thermostats comparatively more than that of the participants from Asia and Africa; whereas Asians prefer Fitbit and smart blood pressure cuffs. Other popular IoT devices owned by all the participants include smart phones (94.4%), tablets (59.8%) and smart TVs (57.7%).

Table 2. Smart home familiarity and IoT devices owned by the participants based on location

	Location					
	Asia (N = 100)	Europe (N = 101)	America (N = 27)	Australia (N = 4)	Africa (N = 2)	Total (N = 234)
Smart Home familiarity						
Yes	43%	63.3%	66.6%	75%	50%	55.1%
No	20%	11%	7.4%	25%	None	29.9%
Kind of	37%	24.7%	25.9%	None	50%	15%
IoT Devices owned						
Smart Phone	95%	94%	92.5%	100%	100%	94.4%
Smart Watch	25%	27.7%	40.7%	50%	None	28.2%
Smart TV	55%	60.3%	59.2%	40%	50%	57.7%
FitBit	32%	26.7%	33.3%	50%	None	29.9%
Tablet	47%	66.3%	81.4%	50%	50%	59.8%
Presence sensors	8%	15.8%	25.9%	25%	None	13.7%
Sleep monitors	16%	10.8%	18.5%	50%	None	14.1%
Monitoring cameras	20%	23.7%	33.3%	25%	None	23.1%
Smart Blood pressure cuffs	17%	9.9%	3.7%	50%	None	12.8%
Voice Assistant	16%	20.7%	55.5%	50%	None	23.1%
Thermostat	18%	45.5%	48.1%	50%	50%	34.2%

In order to know the participants' ability in performing routine activities and their preferences in monitoring camera installation, a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) was used. Majority of the participants (upto 75%) agreed or strongly agreed that they "feel safe even when they are alone", "can perform their daily activities" and "are physically active inside the home"; which can be attributed to the age factor. Safety and security are clearly perceived as important aspects in smart homes as 60% of all participants either agreed or strongly agreed on the point of home security system and cameras outside the home, as depicted in Fig. 1. A two-way analysis of variance (ANOVA) test is performed to analyze preferences about cameras outside the home with respect to the different locations and age of the participants.

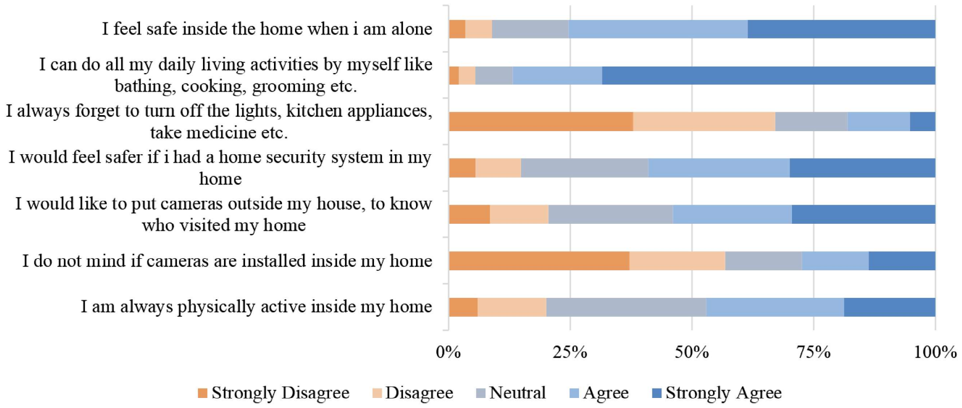


Fig. 1. Attitudes towards home activities

The test result showed that age is not a significant factor while location exhibits a high significance ($p = 0.000$); Asian and American participants agreed or strongly agreed with camera installation outside the home whereas Europeans and Australians disagreed. A similar trend is noticeable with cameras installed inside the home, where only North Americans agreed and the rest disagreed.

In concern to social engagement and outside home activities of the participants, we tried to understand their attitudes towards monitoring of outdoors activities (Fig. 2). Over 50% of all the participants agreed or strongly agreed on the point of "going outside daily for exercise". Most of the participants (70%) disagreed on "not feeling comfortable going outside alone", this is an expected result given the average age of the participants. With respect to the statement "I would not mind if I was monitored when outside the home", 59.4% of all the participants strongly disagreed or disagreed whereas 22.7% agreed or strongly agreed. It was expected that the participants who agreed or strongly agreed should be from the group of older adults but there was no significant contrast observed among the age groups.

Next, the opinions of the participants about personal Artificial Intelligence (AI) assistants are analyzed. As can be seen from Fig. 3, the opinions are quite

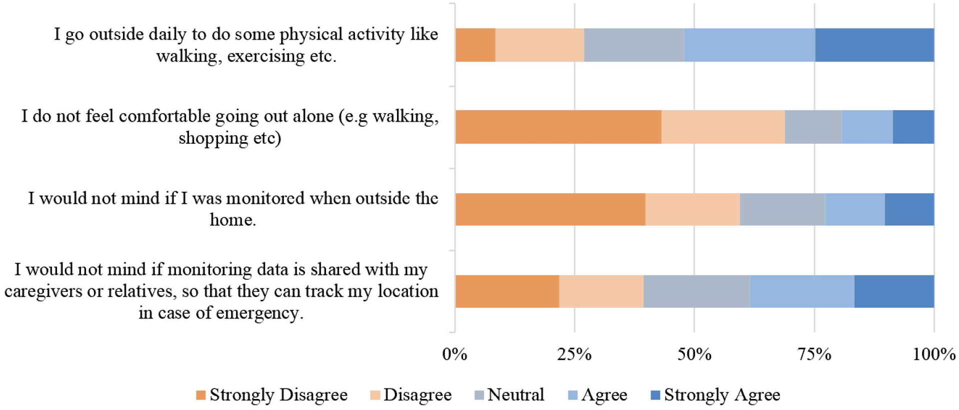


Fig. 2. Attitudes towards outdoors activities

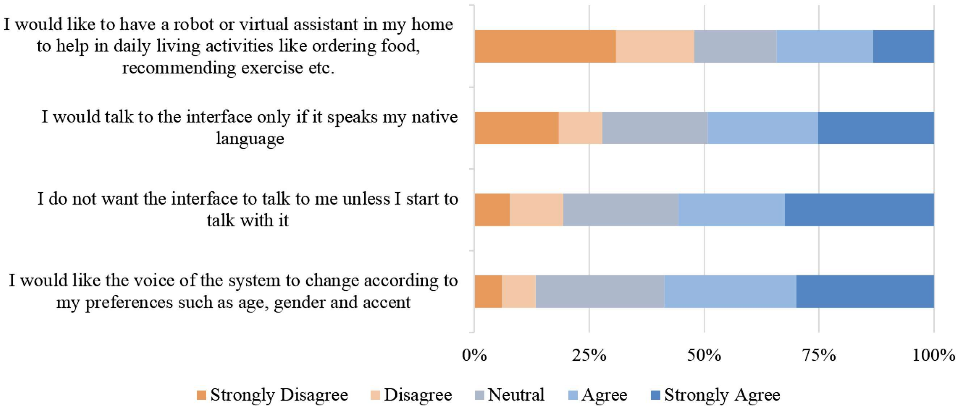
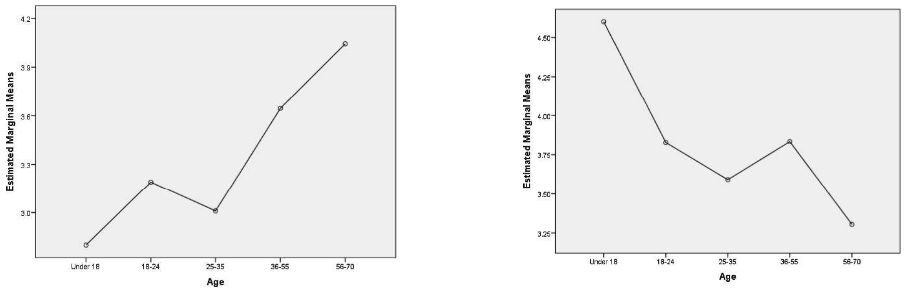


Fig. 3. Attitudes towards personal Artificial Intelligence assistants

balanced in general e.g. almost equal number of participants were either in agreement or disagreement on having a robot or virtual assistant. In the following question with regards to the behavior and abilities of the personal AI assistant, most of participants agreed or strongly agreed that "they would talk to the interface only if it speaks their native language". Also, the majority does not want the interface to start talking with them on its own, they prefer to initiate the conversation and finally, they would like to be able to adjust the speech settings such as gender, age and accent.

A more in-depth analysis was performed to examine if the demographic variables affect the preferences of the participants in context to the capabilities and settings of the personal AI assistant. In Fig. 4, it can be noticed that the effect of gender is not important while there is a trend for age mainly between the age groups of 25–35 years and 56–70 years. For the older adults, it is important ($p = 0.005$) that the system speaks their native language, as seen in Fig. 4a. In contrast, it is quite the opposite from the perspective of young adults (Fig. 4b). Furthermore, it is



(a) "I would talk to the interface only if it speaks my native language"

(b) "I would like the voice of the system to change according to my preferences (such as age, gender and accent)"

Fig. 4. Effect of age on personal assistant settings

important ($p = 0.026$) for the young adults to be able to adjust the speaking/voice settings but for the older adults this feature is of least importance.

In the next part of the study, we investigated about the facilities which the participants would want in their smart homes. The most widely accepted facilities according to the participants are automatic lighting and heating control (79.1%) followed by electrical appliances controlling (68.4%) and emergency alarm system (67.9%). The responses showed (Fig. 5) that security and safety with cameras to monitor visitors, monitoring health status inside home and medicine reminder system are given preference by Asians in comparison to others, while other features were given almost equal preferences. Subsequently, we performed analysis to explore the effect of age on their preferences based on the broad age groups i.e. young adults and older adults. Safety and security with cameras to monitor visitors (74.6%), monitoring health status inside home (50.7%) and medicine reminder system (52%) are preferred by the older adults. More young adults are interested in automation in lighting and heating control (82.2%) as compared to older adults (73.2%).

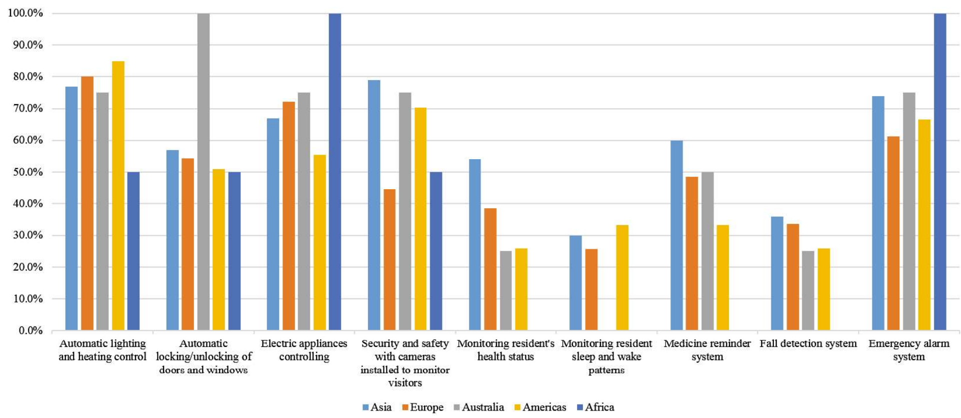


Fig. 5. Facilities the participants would want their smart home to have

The majority of participants perceive the potential benefits as improved quality of life (73.1%), safety (72.6%) and comfort (70.1%), which can be seen in Fig. 6. On the other hand, the drawbacks of smart homes (Fig. 7) are perceived as the increased dependence on technology (76.1 %), the monitoring of private activities (64.5%) and the increased physical idleness (45.3%). This is a general observation of the participants about the perceived benefits and concerns, however, some of the participants expressed their concerns in form of free text as well, e.g. *"Security measures on appliances is lacking and their connectivity to the internet is increasing. This immensely increases the risk of abuse of these appliances. For example a voice activated appliance serving as microphone for spying. Even for transmitting medical data their should be strict guidelines".*

"In my point of view living in a smart home would decrease person's efficiency of doing physical work and increase dependency on machines and technologies".

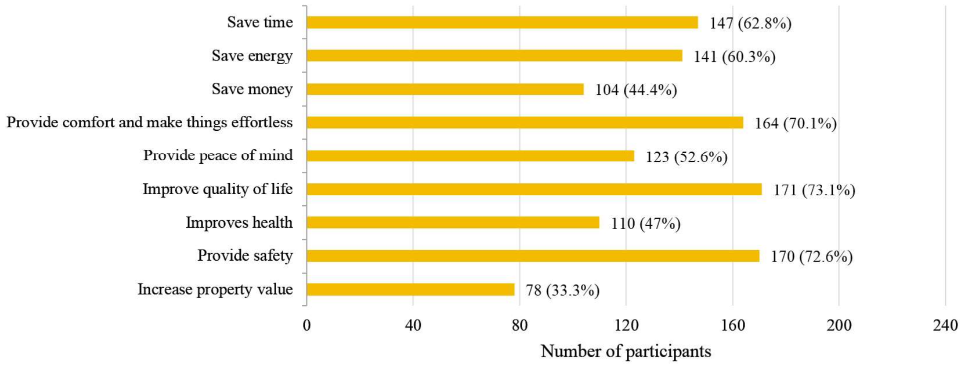


Fig. 6. Perceived smart home benefits

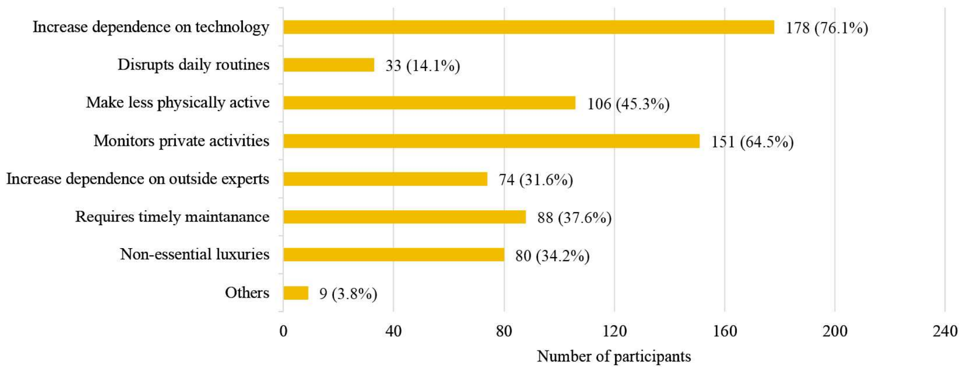


Fig. 7. Perceived smart home drawbacks

From the analyses, it has been observed that Europeans are more concerned towards privacy (74%) in comparison with Asians (54%), whereas older adults group of Asians prefers safety and security (77%) over privacy.

Further analysis of the individual questions in the section about attitudes towards smart home monitoring (Fig. 8) shows that age plays an important role

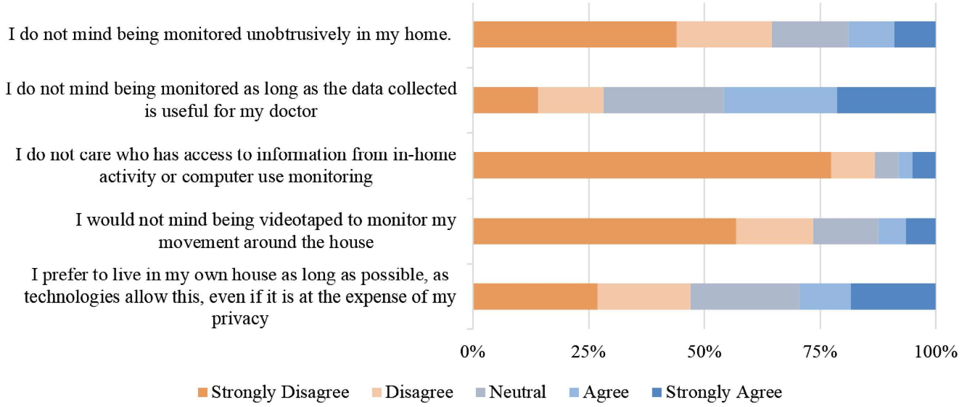
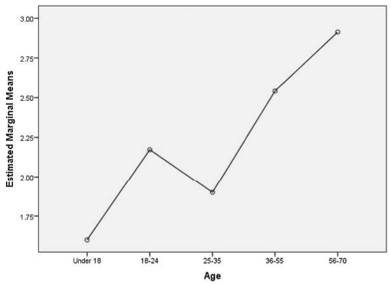
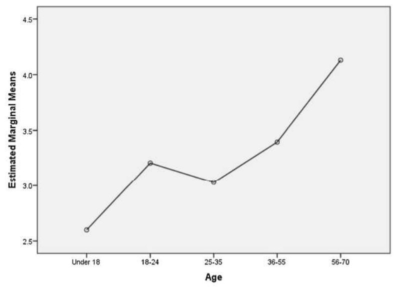


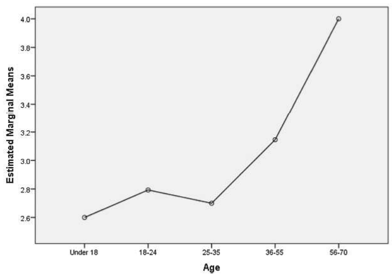
Fig. 8. Attitudes towards smart home monitoring



(a) I do not mind being unobtrusively monitored in my home



(b) I do not mind being monitored as long as the data collected is useful for my doctor



(c) I would not mind if monitoring data is shared with my caregivers or relatives so they can track my location in case of emergency

Fig. 9. Effect of age on monitoring attitudes

in the attitude of the participants. More specifically in the context to the statement "I do not mind being monitored unobtrusively in my home", a significant difference (p = 0.003) can be observed (Fig. 9a) in the opinions, especially between the age groups 23–35 years and 56–70 years. There is also an effect of age on the statement "I do not mind being monitored as long as the data collected is useful for my doctor" (Fig. 9b). The young adults disagree with the statement while the older adults agree ($p = 0.004$). In context to the statement "I would not mind if monitoring data is shared with my caregivers or relatives so they can track my location in case of emergency" the responses were quite equally distributed on the Likert scale. It was further investigated using a two-way ANOVA test to see the effect of age on the responses. In Fig. 9c, it is visible that age has a significant effect ($p = 0.004$) on the willingness of people to share their location data with their relatives or caregivers in case of emergency.

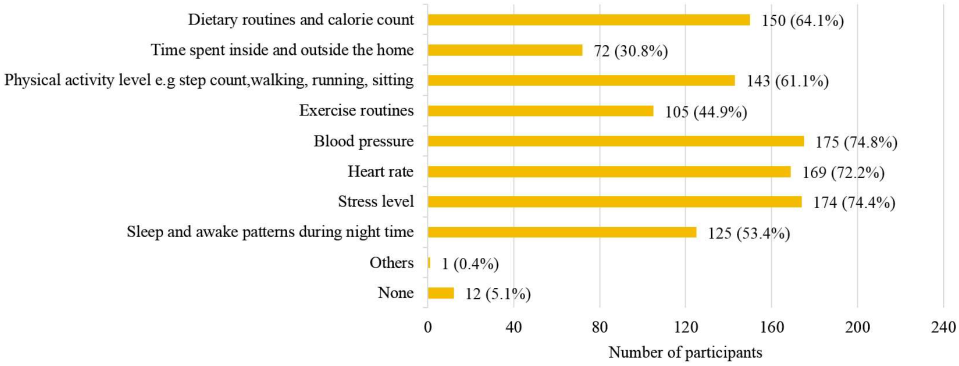


Fig. 10. Smart home data the participants would want to share with their doctor

The last section of this study focuses on data sharing concerns and the kind of data which participants would like to share with their doctors (Fig. 10). Most of the participants agreed with sharing their health related data such as blood pressure (74.8%), stress level (74.4%) and heart rate (72.2%). Also, they do not mind sharing dietary regimes and calorie count (64%) and physical activity level (61%). It is also important to note that 5.1% of the total participants mentioned that they do not want to share any data with their doctor; interestingly all these participants are from Europe and North America.

4 Conclusion

The vision of smart homes can only be realized when the people get interested in adopting these technologies in their daily lives; and the interest can be raised when the developers incorporate their specific needs and concerns into it. This study gives an insight about those needs and concerns for the upcoming smart home users. We investigated the various aspects of usual indoor and outdoor activities to know their ease and activity level. Most of participants irrespective of their location are physically active which could be related to the age (approx. 90% are below 55 years). On having a personal AI assistant to help in daily living activities, around 30% of the participants agreed, majority of them are over 35 years. The most widely accepted facilities according to the participants are automatic lighting and heating control (79.1%) followed by electrical appliances controlling

(68.4%) and emergency alarm system (67.9%). The prospective users of smart home technology perceive the benefits as comfort, safety and improved quality of life but show concerns over the increased dependence on technology and the monitoring of private activities, which may be seen as perceived drawbacks. The older adults are more open to monitoring and data sharing compared to the young adults, especially in cases where the data are beneficial for their doctors and caregivers. The reluctance of the young adults towards smart home technologies and monitoring needs to be further assessed with dedicated studies to understand their concerns and needs. It will help in making the smart home technologies more acceptable by the future generations with minimal or no concern.

Acknowledgement. This work has been funded by the European Union Horizon2020 MSCA ITN ACROSSING project (GA no. 616757). The authors would like to thank the participants and members of the project's consortium for their valuable inputs.

References

1. Cheek, P., Nikpour, L., Nowlin, H.D.: Aging well with smart technology. Nurs. Adm. Q. **29**(4), 329–338 (2005)
2. Cook, D.J.: How smart is your home? Science **335**(6076), 1579–1581 (2012)
3. Chan, M., Campo, E., Estéeve, D., Fourniols, J.Y.: Smart homescurrent features and future perspectives. Maturitas **64**(2), 90–97 (2009)
4. Zion Market Research: Smart Home Market 2016. www.globenewswire.com/news-release/2018/01/03/1281338/0/en/Global-Smart-Home-Market-to-Exceed-53-45-Billion-by-2022-Zion-Market-Research.html/. Accessed 23 Feb 2018
5. AlAbdulkarim, L., Lukszo, Z.: Impact of privacy concerns on consumers' acceptance of smart metering in the Netherlands. In: 2011 IEEE International Conference on Networking, Sensing and Control (ICNSC), pp. 287–292. IEEE (2011)
6. Demiris, G., Oliver, D.P., Dickey, G., Skubic, M., Rantz, M.: Findings from a participatory evaluation of a smart home application for older adults. Technol. Health Care **16**(2), 111–118 (2008)
7. Visutsak, P., Daoudi, M.: The smart home for the elderly: perceptions, technologies and psychological accessibilities: the requirements analysis for the elderly in Thailand. In: 2017 XXVI International Conference on Information, Communication and Automation Technologies (ICAT), pp. 1–6. IEEE (2017)
8. Singh, D., Kropf, J., Hanke, S., Holzinger, A.: Ambient assisted living technologies from the perspectives of older people and professionals. In: Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-MAKE 2017. LNCS, vol. 10410, pp. 255–266. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66808-6_17
9. Coughlin, J.F., D'Ambrosio, L.A., Reimer, B., Pratt, M.R.: Older adult perceptions of smart home technologies: implications for research, policy & market innovations in healthcare. In: 29th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, EMBS 2007, pp. 1810–1815. IEEE (2007)
10. Wilson, C., Hargreaves, T., Hauxwell-Baldwin, R.: Benefits and risks of smart home technologies. Energy Policy **103**, 72–83 (2017)
11. Ziefle, M., Rocker, C., Holzinger, A.: Perceived usefulness of assistive technologies and electronic services for ambient assisted living. In: 2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), pp. 585–592. IEEE (2011)

4.4 Publication IV: Users' Privacy Concerns in IoT based Applications

Users' Privacy Concerns in IoT based Applications

Ismini Psychoula*, Deepika Singh[†] [‡], Liming Chen*, Feng Chen*,
Andreas Holzinger[‡], Huansheng Ning[§]
*School of Computer Science and Informatics, De Montfort University, Leicester, UK
Email: ismini.psychoula@dmu.ac.uk
[†]Center for Health & Bioresources, AIT Austrian Institute of Technology, Wiener Neustadt, Austria
[‡]Holzinger Group, Institute for Medical Informatics/Statistics, Medical University Graz, A-8036 Graz, Austria
[§]School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China

***Abstract*—In recent years user privacy has become an important aspect in the development of the Internet of Things (IoT) services due to their privacy invasive nature. However, there has been comparatively little research so far that aims to understanding users' notion of privacy in connection with IoT. In this work, we aim to understand how and if contextual factors affect users' privacy perceptions of IoT environments. To ascertain privacy perceptions, we deployed a public online survey (N=236) and contacted interviews (N=41) to explore factors that could have an influence. Although a lot of the participants identified privacy risks in IoT and rated the collected information items with high privacy ratings, we find that quite a large number of participants would still decide to have the offered IoT service if they find it useful and practical for their daily lives despite the infringement on their privacy. We conclude by highlighting and analyzing the qualitative comments of the participants and suggest possible solutions for the identified issues.**

I. Introduction

With the rapid deployment of the Internet of Things (IoT) and the advances in artificial intelligence technologies, technology companies are developing new products aiming to make the users' lives more convenient. By some estimates, the installed base of Internet of Things devices is forecast to grow to almost 75 billion worldwide by 2025 [1]. The rise of consumer smart home platforms as well as connected devices, means end users are empowered to set up their own automated IoT environments. These smart devices support desirable features, such as voice-controlled appliances and remote-controlled thermostats. However, these devices also collect and use personal data in order to tailor the offered services to each individual user. These data collection practices will become even more powerful in future IoT environments, given that nearly all of the IoT devices are connected to the Internet and can collectively monitor and gather personal information of users. For instance, the IoT devices may collect users personal information without asking for their permission, or may not give notice to them when collecting potentially sensitive information which raises security and privacy issues.

There is a need for new tools to provide transparency, user control, and ensure that individual privacy requirements are met. To develop these tools, it is important to better understand how people feel about the privacy implications of IoT and the situations they prefer to have control of their privacy. Hence, providing services with minimized privacy risks is very important for both protecting users privacy expectations and making intelligent IoT services more acceptable to the public. In order to achieve these objectives, researchers and developers should understand how different factors influence peoples privacy perceptions in an IoT environment. This understanding will enable them to better design privacy-preserving IoT systems and services. To address this gap, we conduct a survey and semi-structured interviews to learn about what are the features they want from IoT, and to understand their security and privacy related attitudes, expectations, and actions.

The paper is organized as follows. First we discuss related work. Then we describe the design of our study and discuss the qualitative and quantitative analysis of the collected data. Next we present and discuss our results along with highlighting and categorizing the qualitative comments of the participants. Finally we discuss the limitations of the study and ways to address the issues identified.

II. Related Work

In most IoT environments users are surrounded by a variety of sensors and devices that monitor their activities. Prior research on privacy has produced contradictory results. On one hand it shows that privacy is a primary concern for users in the IoT era [2]. On the other hand experimental studies show that individuals reveal personal information for relatively small rewards [3]. Several studies have shown a difference between privacy concerns and attitudes and actual privacy behaviour, while other studies indicate that individuals privacy behaviour is in line with their concerns and attitudes. Although information privacy in general and data sharing and control are universal issues, a lot of researchers argue that the precise concerns and responses to data sharing requests depend on the users characteristics, including their culture [4], [5] and age [6]. Studies have investigated various factors that can impact privacy [7], but the new methods of data collection in the IoT environments have led to new privacy challenges and introduced new factors. Some of these challenges include obtaining consent for data collection, allowing users to control and choose the data they share, while at the same time ensuring the use of collected data is limited to the stated purpose [7]. These challenges become even more difficult by the increased potential for misuse of personal information in the IoT environment. This arises from the pervasive tracking of habits, behaviors, and locations over a long period of time.

There are new risks to personal safety as a result of the use of IoT systems [8], [9]. Several studies were aimed at understanding individuals IoT-related privacy concerns, and proposing potential solutions for them [8], [10]. Prior work has shown that user adoption of innovative and recent technologies is significantly related to their individual characteristics [11] and the trustworthiness of IoT services is affected by the implemented privacy and security practices [12].

The work presented in [13] discusses the use of smart home devices in the living environment of residents aged 65 or over in residential care. Interestingly, a key point of the study's findings is that the privacy concerns of the users reduces when the smart devices are being used for a medical intentions or in the case of a potential emergency. While, it appears that it is the thought of being watched and tracked that causes privacy concerns.

Other studies more closely related to this work have evaluated several factors that may impact privacy concerns related to IoT data collection. In the research presented in [14] the authors studied the relative importance of two factors with the first being the entity collecting data, and the second being the situation in which the data are being collected, in order to determine users privacy preferences in ubiquitous computing settings. Their results indicate that individuals base their privacy decisions on who is collecting their data, rather than the context in which it is being collected. In the studies discussed in [15] and [16] the authors tested five factors related to the context of data collection in two separate studies and found that individuals generally thought that monitoring in personal spaces was unacceptable, along with monitoring by an unknown entity or the government. Their results also indicate that photo and video monitoring may cause some privacy concern regardless of context. Additional smaller, qualitative studies have focused on individuals privacy preferences related to wearable sensors. These studies showed that users want to have ownership of the data they produce, and that privacy concerns vary depending on factors including retention time and the perceived value of the data collected [17], [18].

According to the findings in [8] a limitation of prior work is that the studies usually focus on a single scenario in which sensing is occurring. Thus, many of the proposed solutions do not generalize to other contexts. This work will attempt to address this gap by identifying privacy concerns in heterogeneous scenarios which use different types of data collection and data formats. This way, we can determine which factors have the greatest impact on measures of individuals privacy concerns in regards to IoT and data sharing. The results will enable the design of privacy-enabling solutions appropriate to a variety of contexts applicable in IoT environments. Moreover, it will address the gap of identifying privacy concerns individuals have in data collection scenarios which are not obviously aligned with specific privacy risks. While the impact of the location of the data collection, type of data being collected, and purpose for collection have already been studied in prior work considering IoT contexts [15], [16]. This study aims capturing more contextual nuances that are specific to IoT scenarios. These findings will help understand the relative importance of different privacy concerns to individuals and act as a basis for the development of privacy-preserving mechanisms.

III. Methodology

In this section we discuss the collection and analysis of the data. In the first part of the research we conducted an interview study with 41 older adults (31 female and 10 male) affected by chronic diseases with the goal of qualitatively assessing their perceptions regarding different IoT scenarios and their privacy concerns. Also an online survey was used where 236 people participated from around the world. The goal was to understand the individuals' attitude towards IoT devices and technologies as well as the participants' opinions in regards to monitoring, data sharing and privacy. The questionnaire was advertised via social media and email. No screening criteria were applied, other than the participants being able to respond in English. The study was approved by the Ethics Committee of De Montfort University. The participation in both the interviews and the survey was voluntary and the duration was approximately 15 minutes. The study provided full anonymity to the participants.

The questions followed a similar pattern in both the interviews and the online survey. The first part of the questions focused on socio-demographic questions such as age, gender, location, education and familiarity with technology. The second part consisted of questions about the IoT devices the users' already own and what do they perceive as benefits, drawbacks and concerns related to IoT systems. The third part had questions on the participants' attitudes towards monitoring and data sharing while the last section focused on the participants' privacy and security concerns. The study was described as a Survey on IoT Technologies to avoid having the participants believe that their privacy perceptions are being evaluated, because then they might adjust their responses [19]. Thus, our survey and interviews were framed as concerning general opinions. More obvious privacy related questions were placed near the end to ensure earlier responses were not primed. For the analysis, we used two software packages Microsoft Excel for sentiment analysis and IBM SPSS Statistics to generate the descriptive statistics of the data and the item-level results of each question. A significance level of 5% was used for all the reported results. To test the reliability of the measure we used Cronbach's alpha. For this set of questions it was $\alpha = 0.808$ which indicates a high level of internal consistency.

IV. Results & Discussion

A. *Interviews*

Interviews were conducted one-on-one, with informed consent received at the start. All interviews were recorded and then transcribed in order to be analyzed. The majority of the older adults with chronic diseases that participated in the interviews were not aware of smart homes and IoT technology they use. After the researcher explained to them the concept and services that IoT could potentially offer a small part of them (30%)

found the technology useful and something they would like to have if they could afford it but the majority of them (70%) felt that this kind of technology cannot help them at this point in their life and that it would be too difficult to learn to use it. In regards to data sharing most of the participants (80%) of the interview participants said they would not mind sharing medical and daily life data. They thought the sharing of daily behavior data is particularly useful for their doctors and family members.

B. Online Survey

Characteristics of the Sample		Count	Percentage
Gender	Male	137	58.1%
	Female	99	41.9%
Age	Under 18	5	2.1%
	$18-24$	58	24.6%
	$25-35$	100	42.4%
	$36-55$	50	21.2%
	$56-70$	23	9.7%
Location	Asia	102	43.2%
	Africa	2	0.8%
	North America	24	10.2%
	South America	3	1.3%
	Europe	101	42.8%
	Australia/Oceania	4	1.7%
Education	Not completed school	7	3.0%
	Completed school	35	14.8%
	University degree	101	42.8%
	Postgraduate degree	93	39.4%
Familiarity with Technology	Beginner	2	0.8%
	Basic Knowledge	21	8.9%
	Moderate	90	38.1%
	Expert	123	52.1%

TABLE I: Demographic information of the online survey participants (N=236)

Socio-demographic information was gathered on the factors of gender, age, location, familiarity with technology and education level. The reason for recording these factors is that previous research has suggested that women possess larger privacy concerns than men [20]. It has also reported that older people care more about privacy [21], and this could be reflected in the responses. Previous work also found that education correlates with privacy concern [22]. The data sample consists of 236 responses with 58.1% male and 41.9% female. In regards to age 42.4% of the responses came from the 25-35 age group, with 39.4% of the participants having a postgraduate degree, implying a well-educated participant group (Table I).

C. IoT Devices

Participants reported having a large variety of internet-connected devices. We summarize these devices in Table II. Most common are wearable devices, thermostats, cameras, and smart personal assistants. Then we classified the devices on 5 categories Wearable Devices, Ubiquitous Sensors, Cameras, Personal Gadgets and Virtual Assistants (Figure 1) in order to examine if there are differences in the participant's perceived drawbacks based on the type of devices owned by them. The most prominent drawbacks as perceived by the users are

Type of IoT Device	**Count**
Smartphone	224
Tablet	143
Smart Watch	68
Smart TV	134
Fitness Bracelet	73
Presence Sensors	32
Sleep Monitors	34
Cameras	55
Smart Blood Pressure Cuff	31
Smart Personal Assistant	55
Smart Thermostat	82
Smart Plugs & Lights	4

TABLE II: IoT Devices owned by participants

the increased dependence on technology (more than 65% for all devices) and the monitoring of private activities which is high (around 65%) for owners of ubiquitous sensors, cameras, personal gadgets and virtual assistants and comparatively less (around 55%) for owners of wearable devices. There is no denying that technology has made things easier and helps people in day to day lives but users' concerns also need to be addressed while developing such intelligent systems. Another drawback users perceived is the increased physical inactivity which can be correlated to the dependence on technology. Furthermore, regular charging, periodical maintenance, system updates and connectivity issues (internet, mobile services) are also drawbacks raised by users.

D. Data Sharing

The survey investigated the data sharing attitudes of the participants. To measure the participants' view on privacy, the participants were asked by means of multiple questions to what extent they favor convenience and safety over privacy. The questions were related to the infringement on privacy by surveillance cameras, the sharing of health data with family members and doctors, and the manner in which government agencies handle users' personal data in general. Every question was written as a statement, for example: "I am willing to have information from activity monitoring shared with my doctor". The participant could indicate how much he or she agrees or disagrees with the statement by selecting a value from 1 to 5 on a 5-point Likert scale, where 1 means Strongly Disagree and 5 means Strongly Agree.

In regards to information sharing (Figure 2) most of the participants strongly agree (40%) or agree (28%) on monitored activity data being send to their doctor if it suggests they might suffer from a chronic disease. But these percentages change significantly if the data are to be shared with family members with only 24% strongly agreeing and 20% agreeing. Also the percentages that would want information send to them are much less than being send to the doctor with 29% of the participants strongly agreeing and 28% agreeing. With respect to the statement "It bothers me that my data might be visible and accessible by others", 59% of all the participants strongly agreed and 25% agreed. We notice similar percentages to the statement "I am concerned about privacy in regards to in-home monitoring". While for the statement

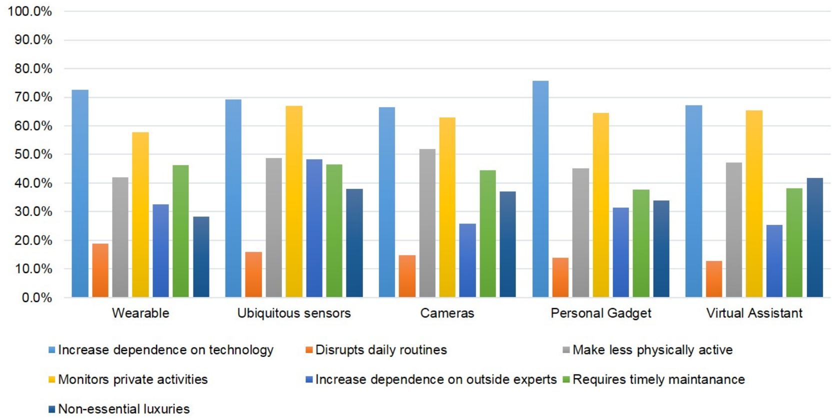


Fig. 1: Perceived Drawbacks Based on IoT Device Type

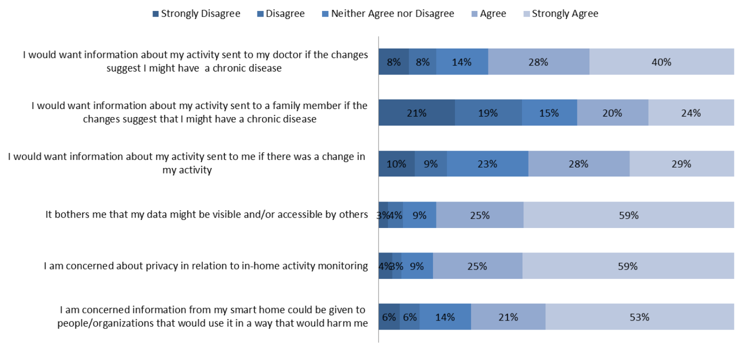


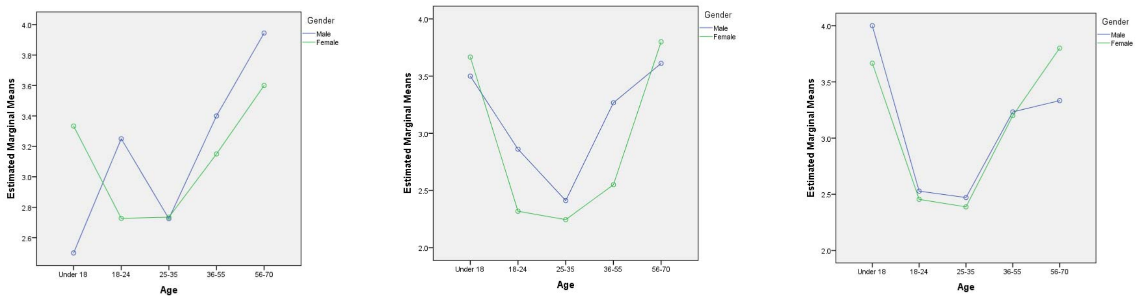
Fig. 2: Participants' Data Sharing Attitudes

"I am concerned information from my smart home could be given to people/organization that would use it in a way that would harm me" the percentages are a bit different with 53% strongly agreeing and 21% agreeing.

Further analysis of the individual questions in the section about attitudes towards data sharing (Figure 3) shows that age plays an important role in the attitude of the participants. More specifically in the statement "I would want information about my activity monitoring send to a family member if the changes suggest I might have a chronic disease", a difference (p=0.05) can be seen (Figure 3a) in the opinions depending on the Age of the participants, especially between the age groups 25-35 and 56-70. There is also an effect of Age on the statement "I am willing to have information from activity monitoring (e.g sleeping, eating, exercising) shared with my family" (Figure 3b). The age group 25-35 disagree with the statement while the older generations agree (p=0.001). Regarding these questions, we can conclude that people with an age ranging from 25 and 35 years are, compared to people of a much older age between 56 and 70, significantly less willing to share their personal data that is being collected by sensor technologies. A two way analysis of variance (ANOVA) was performed to examine if Gender and Age have an effect on the participants' preferences in regards to continuing to stay at home with the help of technology even if that might be at the expense of their privacy. In Figure 3c, it can be seen that the effect of gender is not important while there is a trend for age especially between the age groups 18-35 and 35 and older. For the older generations, it is important (p=0.001) to be able to live in their own house as long as possible, even if it comes at the expense of privacy, in contrast with the younger generation that disagreed with the statement.

E. Information Items & Privacy Ratings

In Figure 4 we show the participants' rating of information items between 0 - "Non private" and 5 - "Extremely Private". The majority of participants (76%) considers baking details "Very Private" followed by data collected from monitoring cameras (54%), medical information (42%) and home security information (41%). In regards to shopping habits, energy



(a) "I would want information about my activity monitoring send to a family member, if the changes suggest a chronic disease."

(b) "I am willing to have information from activity monitoring shared with my family"

(c) "I prefer to live in my own house as long as technology allows this even if it is at the expense of my privacy"

Fig. 3: Effect of age and gender on data sharing preferences

consumption, movement from room to room, exercise and location data the participants' opinions are divided on the level of privacy with some considering them "Neutral" and others "Somewhat Private" or "Private".

To analyze the results in more detail we examined the effect of Culture (Figure 5) and Age (Figure 6). Here we present the the information items that are most affected by these factors. Banking Details are considered extremely private for people from Africa, South America and Australia in comparison to people from Asia . The same can be observed for the Age group 25-35 . With regards to monitoring cameras Europeans and South Americans consider them more private in comparison to people in Africa and Asia , also there is a clear difference in the opinions between the 25-35 age group and the people over 35 . When it comes to location data the it is considered more private for participants in the age groups of Under 18 and 25-35 while it is not ranked as "Somewhat private" for people from Asia and South America . For medical information we observe a similar pattern where the age group 25-35 considers it much more private than the age group 56-70 also participants from Asia consider it less private compared to the rest . Energy consumption and Exercise Data are considered as Neutral by all cultures and age groups. We also performed the same analysis for the factors Familiarity with technology and Education. In regards to familiarity with technology the participants with higher expertise gave higher privacy ratings to all information items compared to beginners or moderate users. The same cannot be said for education as the privacy ratings were quite varied depending on the information item. In some cases the privacy rating of people that have not completed school is higher than that of people with postgraduate degrees. Furthermore, on the gender factor there appears to be a trend with females have giving higher privacy ratings than males.

Figure 7 shows the participants responses when asked what would concern them for each of the information items in the following scenario "Suppose there was a leak of information that gave access to all the information in your smart home. Specifically someone else would have access to: medical information, banking details, location, movement from room to room, energy consumption, monitoring cameras, home security system, exercise and shopping habits." (The participants had the option to select multiple items). A lot of the participants where concerned about potentially being embarrassed about the content and the financial risk of the content, while the ease of access by third parties is not a very prominent concern with the exception of Location data.

The participants where also asked if they would have any concerns about living in a smart home in an open ended type of question. (This question was asked at the beginning of the questionnaire in order to gather the participants unbiased answers before we presented them with other risky scenarios and questions). From the 236 participants 108 (45.7%) replied that they would have no concerns about living in a smart home environment. Participants mentioned a diverse set of potential security and privacy issues, but few concrete concerns were articulated by a majority of participants. Moreover, participants were sometimes aware of potential issues but were not concerned about them. After analyzing the survey answers of the participants that mentioned they have concerns, we identified several concern categories such as privacy and security , data sharing, cost, dependence on technology and trust on technology (Table III). It is interesting to note that from the people that mentioned concerns most of them belong in the 25-35 age group.

F. Internet of Things Scenarios

The participants of the online survey where also presented with IoT scenarios, two of those along with example responses are presented below.

Scenario I: *"Imagine there is a new inexpensive thermostat sensor for your house that can learn about your temperature preferences and movements around the house and potentially save money on your energy bill. It is programmable remotely in return for sharing data about some of the basic activities*

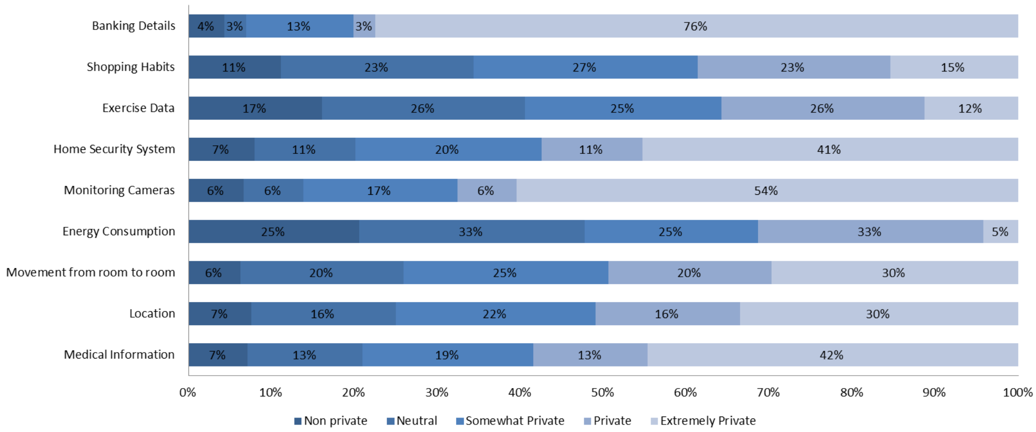


Fig. 4: Participants' Privacy Rating of Collected Data

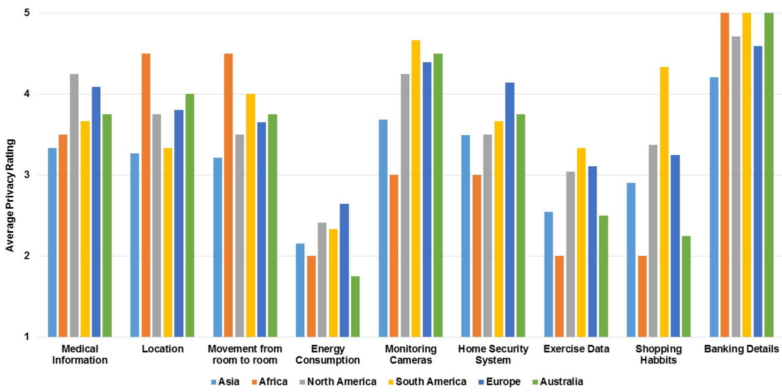


Fig. 5: Effect of Culture on Participants' Privacy Rating of Collected Data

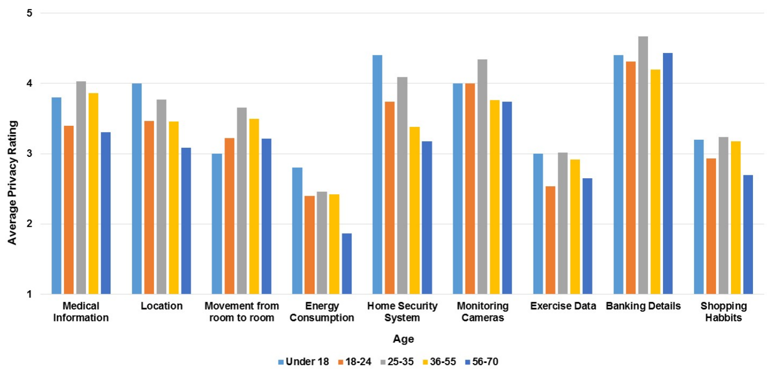


Fig. 6: Effect of Age on Participants' Privacy Rating of Collected Data

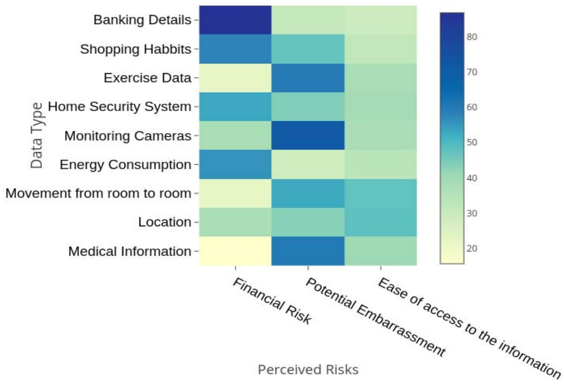


Fig. 7: Summary statistics showing the relationship between various factors and participants concerns for each one. (E.g. 71% of the participants were worried about potential embarrassment upon potential exposure of monitoring camera data.)

Categories	Representative Quotes
Cost	[P8]: "If a whole automated system already existed and it was affordable it would be really helpful in my condition (paraplegia). Although I am aware of the potential threats in my condition the benefits outweigh the risks."
Dependence on Technology	[P37]: "In my point of view living in a smart home would decrease persons efficiency of doing physical work and increase dependency on machines and technologies."
Trust on Technology	[P196]: "The worry of it backfiring, alarms going off incorrectly, doors unlocking by mistake. AI is improving at an alarming rate...
Information Control	[P224]: "The major thing that concerns me is the gathering of information. Who will be able to access in database? Would it be easy for me to shut down all of these systems ,at once, in emergency cases?"
Privacy & Security	[P154]: "I would be concerned with the possibility that they could easily be hacked", [P161] "I might feel uncomfortable because of lack of privacy "

TABLE III: Participants' Concerns

that take place in your house like when people are there and when they move from room to room."

After performing sentiment analysis on the comments from the participants we found that 104 found the service acceptable, 108 found it unacceptable and 23 where neutral depending on the benefits they might receive or what are the terms of service that where offered. Table IV shows the identified categories resulting from the sentiment analysis and showcases some of the participants' responses.

Scenario II: *" Imagine that in the future a smart fridge would be able to know when you run out of food and order for you all the groceries you need. The smart fridge will keep track of your shopping habits and might give them to third parties."*
In the second scenario 74 participants found it acceptable and another 30 participants found it partly acceptable depending on various factors (see Table V). Most of the participants that found this scenario unacceptable mentioned privacy concerns, with a small percentage of them reporting practical issues (how is the selection of products done, how to order things that are not store in the fridge etc.) as the reasons they found the scenario unacceptable.

Categories	Representative Quotes
Type of Information	[P97]: "Yes, as I consider such information to not be too intrusive. However, it strictly depends on the type of information collected and whether it is specifically stated it collects such information or not."
Data Sharing	[P3]: " I wouldn't want to share information but if there isn't any other way for the thermostat to work, then I am ok with sharing"
Terms of Service	[P115]: "If there is a clear ToS where the user is protected and can decide then I see no problem. Informed consent is a must if others see the data."
Privacy & Security	P[185]: "It may be acceptable, but there is always a concern about privacy and big brother watching you all the time. If it was anonymous ie not linked specifically to individuals but monitoring movement of people in the house I would not mind."

TABLE IV: Participants' Responses to Scenario I

Categories	Representative Quotes
Benefits	P[49]: " It is acceptable to lose some privacy for the benefits this smart fridge provides to me.", [P217]: " Not 100% sure. It's all about privacy vs easiness!"
Exploitation	[P224]: No I would not like that. It is private data! I would not like someone to profit with my daily collected data or even just know them and create a pattern for my daily food habits. Only if it would be my doctor I would accept it and under specific conditions.
Data Sharing	P[105]: "Partially acceptable. Will only be completely acceptable if there is a clear option to enable and disable data sharing with 3rd parties and functions related to it at any time without any cost to my person."
Privacy & Security	[P209]: " Security concerns. I don't like third parties knowing when I'm about to go out of my home and be at a specific place.

TABLE V: Participants' Responses to Scenario II

In all the open ended questions the identified concern themes are recurring with the participants mentioning that their acceptance of the technology depends on the terms on service, the informed consent, the security measures that are implemented and their want for data sharing control. Another concern for the users was the availability and reliability of the service. Also it is noteworthy that some participants identified availability of device functionality as an asset that might be attacked. Although several participants voiced concerns about reliability and trust on technology none connected this concern to security risks. Interestingly when the participants speculated about potential attacks on their IoT environment, they did not articulate specific adversaries, they often referred to potential adversaries as someone.

Many participants identified a trade off between security and privacy with functionality and convenience, in some cases sounding resigned to it. There is a growing need to present a better trade-off to the end users. For example, certain technical design choices can reduce risks without significantly impacting functionality, like not always requiring the cloud but also running automation services locally. By minimizing these trade offs where possible, we can ease the decision-making burden from users and enable adoption of IoT technologies by people who are not willing to make the trade offs required today.

V. Conclusion

Our results provide contributions to privacy literature that comes from a qualitative and quantitative angle to provide insights into privacy issues from the perspective, of participants around the world. The findings in this study offer a more profound consideration of the various issues that worry IoT users and provide a detailed assessment of the similarities and differences in peoples attitudes and behaviours, depending on socio-demographic background. Moreover our findings analyze and identify themes for the participants' privacy and security concerns (or lack thereof). We notice that With the IoT environments, people have a sense that there's some privacy issues, but they dont always know or understand what data is being collected, or how or why. Out results also indicate that elderly people are more open to data sharing and are less concerned about privacy issues compared to younger generations. This result can be explained by the tendency of younger people to be more familiar with technology and more aware about the risks IoT technologies pose. Our study focused mainly on peoples perceptions, attitudes, and opinions but it also tests causal links among the interesting concepts and factors. Moreover we find that to the people that have privacy concerns or are on the fence about IoT environments control has a significant influence along with trust. Regulations and security mechanisms are important factors for the acceptance of the technology. One of the limitations of our study is the small participant number which makes the results not easily statistically generalizable to all people and settings. Our future work will investigate which of factors are direct or indirect determinants of privacy, whether it is trust, risk, perception, knowledge, awareness or all of the above that shape a user's privacy attitude with a larger number of participants and test whether these processes are similar or different between different groups.

Acknowledgment

This work has been funded by the European Union Horizon 2020 MSCA ITN ACROSSING project (GA no. 616757) and partly supported by the National Natural Science Foundation of China under Grant 61811530335. The authors would also like to thank the participants and AKTIOS care units for facilitating the interviews with people with chronic diseases.

References

[1] Internet of Things (IoT) connected devices installed worldwide. https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/.

[2] C. L. Miltgen and D. Peyrat-Guillard, "Cultural and generational influences on privacy concerns: a qualitative study in seven european countries," *European Journal of Information Systems*, vol. 23, no. 2, pp. 103–125, 2014.

[3] J. P. Carrascal, C. Riederer, V. Erramilli, M. Cherubini, and R. de Oliveira, "Your browsing behavior for a big mac: Economics of personal information online," in *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 189–200.

[4] R. Thomson, M. Yuki, and N. Ito, "A socio-ecological approach to national differences in online privacy concern: The role of relational mobility and trust," *Computers in Human Behavior*, vol. 51, pp. 285–292, 2015.

[5] S. Trepte and P. Masur, "Cultural differences in social media use, privacy, and self-disclosure," Research Report on a multicultural survey study. University of Hohenheim. Retrieved from http://opus. uni-hohenheim. de/volltexte/2016/1218/pdf/Trepte_Masur_ResearchReport. pdf, Tech. Rep., 2016.

[6] D. Singh, I. Psychoula, J. Kropf, S. Hanke, and A. Holzinger, "Users perceptions and attitudes towards smart home technologies," in *International Conference on Smart Homes and Health Telematics*. Springer, 2018, pp. 203–214.

[7] C. Perera, R. Ranjan, L. Wang, S. U. Khan, and A. Y. Zomaya, "Big data privacy in the internet of things era," *IT Professional*, vol. 17, no. 3, pp. 32–39, 2015.

[8] P. Bhaskar and S. I. Ahamed, "Privacy in pervasive computing and open issues," in *Availability, Reliability and Security, 2007. ARES 2007. The Second International Conference on*. IEEE, 2007, pp. 147–154.

[9] F. T. Commission *et al.*, "Internet of things: Privacy & security in a connected world," *Washington, DC: Federal Trade Commission*, 2015.

[10] R. Chow, S. Egelman, R. Kannavara, H. Lee, S. Misra, and E. Wang, "Hci in business: A collaboration with academia in iot privacy," in *International Conference on HCI in Business*. Springer, 2015, pp. 679–687.

[11] A. Burton-Jones and G. S. Hubona, "Individual differences and usage behavior: revisiting a technology acceptance model assumption," *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, vol. 36, no. 2, pp. 58–77, 2005.

[12] I. D. Addo, S. I. Ahamed, S. S. Yau, and A. Buduru, "A reference architecture for improving security and privacy in internet of things applications," in *Mobile Services (MS), 2014 IEEE International Conference on*. IEEE, 2014, pp. 108–115.

[13] K. L. Courtney, "Privacy and senior willingness to adopt smart home information technology in residential care facilities," 2008.

[14] S. Lederer, J. Mankoff, and A. K. Dey, "Who wants to know what when? privacy preference determinants in ubiquitous computing," in *CHI'03 extended abstracts on Human factors in computing systems*. ACM, 2003, pp. 724–725.

[15] H. Lee and A. Kobsa, "Understanding user privacy in internet of things environments," in *Internet of Things (WF-IoT), 2016 IEEE 3rd World Forum on*. IEEE, 2016, pp. 407–412.

[16] ——, "Privacy preference modeling and prediction in a simulated campuswide iot environment," in *Pervasive Computing and Communications (PerCom), 2017 IEEE International Conference on*. IEEE, 2017, pp. 276–285.

[17] D. Barua, J. Kay, and C. Paris, "Viewing and controlling personal sensor data: what do users want?" in *International Conference on Persuasive Technology*. Springer, 2013, pp. 15–26.

[18] P. Klasnja, S. Consolvo, T. Choudhury, R. Beckwith, and J. Hightower, "Exploring privacy concerns about personal sensing," in *International Conference on Pervasive Computing*. Springer, 2009, pp. 176–183.

[19] P. Rajivan and L. J. Camp, "Influence of privacy attitude and privacy cue framing on android app choices." in *WPI@ SOUPS*, 2016.

[20] K. B. Sheehan, "An investigation of gender differences in on-line privacy concerns and resultant behaviors," *Journal of Interactive Marketing*, vol. 13, no. 4, pp. 24–38, 1999.

[21] P. Han and A. Maclaurin, "Do consumers really care about online privacy?" *Marketing Management*, vol. 11, no. 1, p. 35, 2002.

[22] D. ONeil, "Analysis of internet users level of online privacy concerns," *Social Science Computer Review*, vol. 19, no. 1, pp. 17–31, 2001.

CHAPTER 5

Human Activity Recognition in Smart Home

5.1 Introduction

Numerous previous works have used statistical machine learning methods for human activity recognition. Such conventional algorithms such as support vector machine and random forest, require manual extraction of some representative features from large and noisy sensor datasets. The manual feature engineering requires export knowledge and can inevitably miss implicit features. Until recently, there were no universal or systematical feature extraction techniques to effectively capture distinguishable features for human activities. Deep learning has gained popularity and tremendous success in image recognition and natural language processing tasks due to its ability to automatically learn representative features from a large amount of complex data. This learning ability also enables the activity recognition system to analyze multimodal sensory data for accurate recognition deeply. Diverse structures of deep neural networks encode features from multiple perspectives. For example, the Recurrent neural networks (RNNs) algorithm can extract the temporal dependencies and incrementally learn information through time intervals and Convolutional neural networks (CNNs) can capture local dependence and a signal's scale invariance. Since sensor-based activity data has high dimension characteristics and data features, therefore extracting data features using traditional statistical methods is difficult. CNNs are competent in extracting complex activity recognition data features, local connections of multimodal sensor data, and the translational invariance introduced by locality that leads to accurate recognition [94] [95]. Therefore, we were first to developed models based on Long short term memory (LSTM) and CNN for sensor-based activity recognition in the single-occupancy dataset. The first publication of the chapter introduces the LSTM model for activity recognition and compared performance with traditional models such as Naive Bayes, Hidden Markov Model (HMM), Hidden Semi-Markov Model (HSMM), and Conditional Random Fields (CRF). The second publication presented the 1D-CNN model and evaluated it on the same activity recognition dataset to compare performance with LSTM and other statistical models.

Also, work has been performed for multiple resident activity recognition since, in real-life scenarios, a home is mostly occupied by more than one resident. Therefore, our third publication focuses on handling class imbalance problem in multiple resident settings using deep learning techniques.

The content of this chapter is based on the publications:

Singh, D., Merdivan, E., Psychoula, I., Kropf, J., Hanke, S., Geist, M. and Holzinger, A., 2017, August. Human activity recognition using recurrent neural networks. In International Cross-Domain Conference for Machine Learning and Knowledge Extraction (pp. 267-274). Springer, Cham.

Contribution: Deepika Singh proposed the idea and both Deepika Singh and Erinc Merdivan performed the experiments and evaluation on sensor dataset. The manuscript was written and presented at CD-MAKE 2017 in Regio di Calabria, Italy by Deepika Singh. The other authors contributed to the scientific discussion and revision of the manuscript.

Singh, D., Merdivan, E., Hanke, S., Kropf, J., Geist, M. and Holzinger, A., 2017. Convolutional and recurrent neural networks for activity recognition in smart environment. In Towards integrative machine learning and knowledge extraction (pp. 194-205). Springer, Cham.

Contribution: Deepika Singh initiated the work and both Deepika Singh and Erinc Merdivan performed the experiments and evaluation on sensor dataset. The manuscript was written by Deepika Singh. The other authors contributed to the scientific discussion and revision of the manuscript.

Singh, D, Merdivan, E., Kropf, J. and Holzinger, A. Handling Imbalanced Data in Deep Learning for Multiple Resident Activity Recognition. Submitted to IEEE Transactions on Neural Networks and Learning Systems (2021).

Contribution: Deepika Singh designed and performed all the experiments. The manuscript was written by Deepika Singh and the other authors contributed to the scientific discussion and revision of the manuscript.

5.2 Publication V: Human Activity Recognition using Recurrent Neural Networks

Human Activity Recognition Using Recurrent Neural Networks

Deepika Singh[1(✉)], Erinc Merdivan[1(✉)], Ismini Psychoula[2], Johannes Kropf[1], Sten Hanke[1], Matthieu Geist[3], and Andreas Holzinger[4]

[1] AIT Austrian Institute of Technology, Wiener Neustadt, Austria
{deepika.singh,erinc.merdivan}@ait.ac.at
[2] School of Computer Science and Informatics, De Montfort University, Leicester, UK
[3] CentraleSupelec, Châtenay-Malabry, France
[4] Holzinger Group, HCI-KDD, Institute for Medical Informatics/Statistics, Medical University Graz, Graz, Austria

Abstract. Human activity recognition using smart home sensors is one of the bases of ubiquitous computing in smart environments and a topic undergoing intense research in the field of ambient assisted living. The increasingly large amount of data sets calls for machine learning methods. In this paper, we introduce a deep learning model that learns to classify human activities without using any prior knowledge. For this purpose, a Long Short Term Memory (LSTM) Recurrent Neural Network was applied to three real world smart home datasets. The results of these experiments show that the proposed approach outperforms the existing ones in terms of accuracy and performance.

Keywords: Machine learning · Deep learning · Human activity recognition · Sensors · Ambient assisted living · LSTM

1 Introduction

Human Activity recognition has been an active research area in the last decades due to its applicability in different domains and the increasing need for home automation and convenience services for the elderly [1]. Among them, activity recognition in Smart Homes with the use of simple and ubiquitous sensors, has gained a lot of attention in the field of ambient intelligence and assisted living technologies for enhancing the quality of life of the residents within the home environment [2].

The goal of activity recognition is to identify and detect simple and complex activities in real world settings using sensor data. It is a challenging task, as the data generated from the sensors are sometimes ambiguous with respect to the activity taking place. This causes ambiguity in the interpretation of activities. Sometimes the data obtained can be noisy as well. Noise in the data can be caused by humans or due to error in the network system which fails to give

A. Holzinger et al. (Eds.): CD-MAKE 2017, LNCS 10410, pp. 267–274, 2017.
DOI: 10.1007/978-3-319-66808-6_18

correct sensor readings. Such real-world settings are full of uncertainties and calls for methods to learn from data, to extract knowledge and helps in making decisions. Moreover, the inverse probability allows to infer unknowns and to make predictions [3].

Consequently, many different probabilistic, but also non-probabilistic models, have been proposed for human activity recognition. Patterns corresponding to the activities are detected using sensors such as accelerometers, gyroscopes or passive infrared sensors, *etc.*, either using feature extraction on sliding window followed by classification [4] or with Hidden Markov Modeling (HMM) [5].

In recent years, there has been a growing interest in deep learning techniques. Deep learning is a general term for neural network methods which are based on learning representations from raw data and contain more than one hidden layer. The network learns many layers of non-linear information processing for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. Deep learning techniques have already outperformed other machine learning algorithms in applications such as computer vision [6], audio [7] and speech recognition [8].

In this paper, we introduce a recurrent neural network model for human activity recognition. The classification of the human activities such as cooking, bathing, and sleeping is performed using the Long Short-Term Memory classifier (LSTM) on publicly available Benchmark datasets [9]. An evaluation of the results has been performed by comparing with the standardized machine learning algorithms such as Naive Bayes, HMM, Hidden Semi-Markov Model (HSMM) and Conditional Random Fields (CRF).

The paper is organized as follows. Section 2 presents an overview of activity recognition models and related work in machine learning techniques. Section 3 introduces Long Short-Term Memory (LSTM) recurrent neural networks. Section 4 describes the datasets that were used and explains the results in comparison to different well-known algorithms. Finally, Sect. 5 discusses the outcomes of the experiments and suggestions for future work.

2 Related Work

In previous research, activity recognition models have been classified into data-driven and knowledge-driven approaches. The data-driven approaches are capable of handling uncertainties and temporal information [10] but require large datasets for training and learning. Unfortunately, the availability of large real world datasets is a major challenge in the field of ambient assisted living. The knowledge-driven techniques are used in predictions and follow a description-based approach to model the relationships between sensor data and activities. These approaches are easy to understand and use but they cannot handle uncertainty and temporal information [11].

Various approaches have been explored for activity recognition, among them the majority of the techniques focuses on classification algorithms such as Naive Bayes (NB) [12], Decision Trees [13], HMM [5], CRF [14], Nearest Neighbor (NN) [15], Support Vector Machines (SVM) [16] and different boosting techniques.

A simple probabilistic classifier in machine learning is the Naive Bayes classifier which yields good accuracy with large amounts of sample data but does not model any temporal information. The HMM, HSMM, and CRF are the most popular approaches for including such temporal information. However, these approaches sometimes discard pattern sequences that convey information through the length of intervals between events. This motivates the study of recurrent neural networks (RNN) which promises the recognition of patterns that are defined by temporal distance [17].

LSTM is a recurrent neural network architecture that is designed to model temporal sequences and learn long-term dependency problems. The network is well suited for language modeling tasks; it has been shown that the network in combination with clustering techniques increases the training and testing time of the model [18] and outperforms the large scale acoustic model in speech recognition systems [19].

3 LSTM Model

LSTM is a recurrent neural network architecture that was proposed in [20]. Another version without a forget gate was later proposed in [21] and extended in [22]. LSTM has been developed in order to deal with gradient decay or gradient blow-up problems and can be seen as a deep neural network architecture when unrolled in time. The LSTM layer's main component is a unit called memory block. An LSTM block has three gates which are input, output and forget gates. These gates can be seen as write, read and reset operations for the cells. An LSTM cell state is the key component which carries the information between each LSTM block. Modifications to the cell state are controlled with the three gates described above. An LSTM single cell, as well as how each gate is connected to each other and the cell state itself, can be seen in Fig. 1.

Each gate and cell state are governed by multiplicative equations that are given by:

$$\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o), \\
c_t &= f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \\
h_t &= o_t \tanh c_t,
\end{aligned}$$

with W being the weight matrix and x is the input, σ being the sigmoid and tanh is the hyperbolic tangent activation function. The terms i, f and o are named after their corresponding gates and c represents the memory cell [23].

By unrolling LSTM single cells in time we construct an LSTM layer where h_t is the hidden state and y_t is the output at time t as shown in Fig. 2.

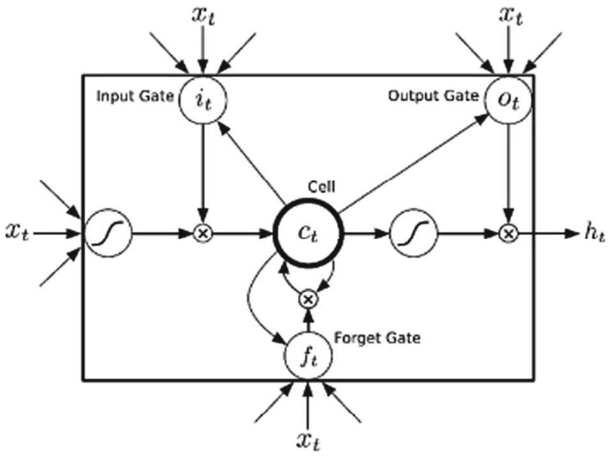


Fig. 1. LSTM single cell image [23].

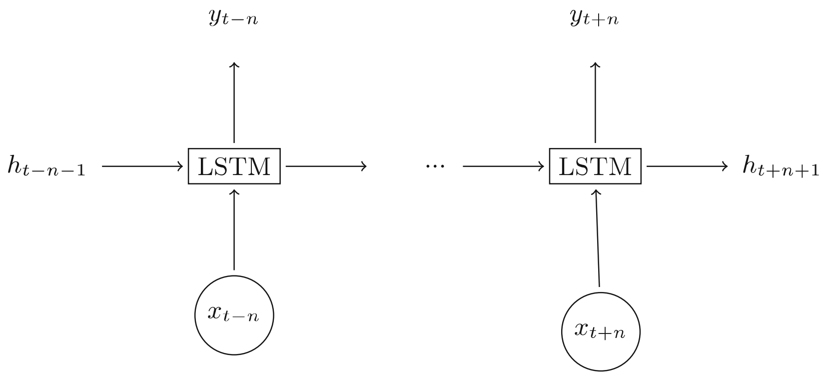


Fig. 2. Illustrations of an LSTM network with x being the binary vector for sensor input and y being the activity label prediction of the LSTM network.

4 Experiments

4.1 Dataset

Publicly available and annotated sensor datasets have been used to evaluate the performance of the proposed approach [9]. In this dataset, there are three houses with different settings to collect sensory data. The three different houses were all occupied by a single user named A, B, and C respectively. Each user recorded and annotated their daily activities. Different number of binary sensors were deployed in each house such as passive infrared (PIR) motion detectors to detect motion in a specific area, pressure sensors on couches and beds to identify the user's presence, reed switches on cupboards and doors to measure open or close status, and float sensors in the bathroom to measure toilet being flushed or not. The data were annotated using two approaches: (1) keeping a diary in

Table 1. Details of the datasets.

	House A	House B	House C
Age	26	28	57
Gender	Male	Male	Male
Setting	Apartment	Apartment	House
Rooms	3	2	6
Duration	25days	14days	19days
Sensors	14	23	21
Activities	10	13	16
Annotation	Bluetooth	Diary	Bluetooth

which the activities were logged by hand and (2) with the use of a blue tooth headset along with a speech recognition software. A total of three datasets were collected from the three different houses. Details about the datasets are shown in Table 1 where each column shows the details of the house with the information of the user living in it, the sensors placed in the house and the number of activity labels that were used.

The data used in the experiments have different representation forms. The first form is raw sensor data, which are the data received directly from the sensor. The second form is last-fired sensor data which are the data received from the sensor that was fired last. The last firing sensor gives continuously 1 and changes to 0 when another sensor changes its state. For each house, we left one day out of the data to be used later for the testing phase and used the rest of the data for training. We repeated this for every day and for each house. Separate models are trained for each house since the number of sensors varies, and a different user resides in each house. Sensors are recorded at one-minute intervals for 24 h, which totals in 1440 length input for each day.

4.2 Results

The results presented in Table 2 show the performance of the LSTM model on raw sensor data in comparison with the results of NB, HMM, HSMM and CRF [9]. Table 3 shows the results of the LSTM model on last-fired sensor data again in comparison with the results of NB, HMM, HSMM and CRF. For the LSTM model, a time slice of (70) with hidden state size (300) are used. For the optimization of the network, Adam is used with a learning rate of 0.0004 [24] and Tensorflow was used to implement the LSTM network. The training took place on a Titan X GPU and the time required to train one day for one house is approximately 30 min, but training times differ amongst the houses. Since different houses have different days we calculated the average accuracy amongst all days. The training is performed using a single GPU but the trained models can be used for inference without losing performance when there is no GPU.

Table 2. Results of raw sensor data

Model	House A	House B	House C
Naive Bayes	77.1 ± 20.8	80.4 ± 18.0	46.5 ± 22.6
HMM	59.1 ± 28.7	63.2 ± 24.7	26.5 ± 22.7
HSMM	59.5 ± 29.0	63.8 ± 24.2	31.2 ± 24.6
CRF	89.8 ± 8.5	78.0 ± 25.9	46.3 ± 25.5
LSTM(Ours)	$\mathbf{89.8 \pm 8.2}$	$\mathbf{85.7 \pm 14.3}$	$\mathbf{64.22 \pm 21.9}$

Table 2 shows the results of different models on raw data from three different houses. The LSTM model has the best performance for all three data sets. In House B and House C, LSTM improves the best result significantly especially on House C where the improvement is approximately 40%.

Table 3. Results of last-fired sensor data

Model	House A	House B	House C
Naive Bayes	95.3 ± 2.8	86.2 ± 13.8	87.0 ± 12.2
HMM	89.5 ± 8.4	48.4 ± 26.0	83.9 ± 13.9
HSMM	91.0 ± 7.2	67.1 ± 24.8	84.5 ± 13.2
CRF	$\mathbf{96.4 \pm 2.4}$	$\mathbf{89.2 \pm 13.9}$	$\mathbf{89.7 \pm 8.4}$
LSTM	95.3 ± 2.0	88.5 ± 12.6	85.9 ± 10.6

Table 3 shows the results on last fired data from three different houses using the same models as in Table 2. The LSTM model did not improve the results in this section but it matched the best performance for two data sets with a slight drop in House C.

5 Discussion

The results presented in this paper show that the deep learning based approaches for activity recognition from raw sensory inputs can lead to significant improvement in performance, increased accuracy, and better results. As shown in Sect. 4.2 our LSTM based activity predictor matched or outperformed existing probabilistic models such as Naive Bayes, HMM, HSMM and CRF on raw input and in one case improved the best result by 40%. Predicting on raw input also reduces the human efforts required on data preprocessing and handcrafting features which can be very time consuming when it comes to an AAL (Ambient Assisted Living) environment.

6 Future Work

Our future work will focus on reducing the variance on our predictions and early stopping criteria while training on different days. The LSTM model has different hyperparameters which affect the performance of the model significantly. Different optimization and hyperparameter search techniques could be investigated in the future. Since the LSTM model has proven to be superior on raw data it would be interesting to also apply other deep learning models. One problem is that deep learning badly captures model uncertainty. Bayesian models offer a framework to reason about model uncertainty. Recently, Yarin and Ghahramani (2016) [25] developed a theoretical framework casting dropout training in deep neural networks as approximate Bayesian inference in deep Gaussian processes. This mitigates the problem of representing uncertainty in deep learning without sacrificing either computational complexity or test accuracy.

Acknowledgement. This work has been funded by the European Union Horizon2020 MSCA ITN ACROSSING project (GA no. 616757). The authors would like to thank the members of the project's consortium for their valuable inputs.

References

1. Roecker, C., Ziefle, M., Holzinger, A.: Social inclusion in ambient assisted living environments: home automation and convenience services for elderly users. In: Proceedings of the International Conference on Artificial Intelligence (ICAI 2011), pp. 55–59. CSERA Press, New York (2011)
2. Chen, L., Hoey, J., Nugent, C.D., Cook, D.J., Yu, Z.: Sensor-based activity recognition. IEEE Trans. Syst. Man Cybern., Part C (Applications and Reviews) **42**, 790–808 (2012)
3. Holzinger, A.: Introduction to machine learning and knowledge extraction (MAKE). Mach. Learn. Knowl. Extr. **1**, 1–20 (2017)
4. Roggen, D., Cuspinera, L.P., Pombo, G., Ali, F., Nguyen-Dinh, L.-V.: Limited-Memory Warping LCSS for real-time low-power pattern recognition in wireless nodes. In: Abdelzaher, T., Pereira, N., Tovar, E. (eds.) EWSN 2015. LNCS, vol. 8965, pp. 151–167. Springer, Cham (2015). doi:10.1007/978-3-319-15582-1_10
5. Duong, T.V., Bui, H.H., Phung, D.Q., Venkatesh, S.: Activity recognition and abnormality detection with the switching hidden semi-markov model. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 838–845. IEEE (2005)
6. Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 609–616. ACM (2009)
7. Lee, H., Pham, P., Largman, Y., Ng, A.Y.: Unsupervised feature learning for audio classification using convolutional deep belief networks. In: Advances in Neural Information Processing Systems, pp. 1096–1104 (2009)
8. Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., et al.: Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Sig. Process. Mag. **29**, 82–97 (2012)

9. Kasteren, T.L., Englebienne, G., Kröse, B.J.: Human activity recognition from wireless sensor network data: Benchmark and software. In: Chen, L. (ed.) Activity Recognition in Pervasive Intelligent Environments, pp. 165–186. Atlantis Press, Amsterdam (2011)
10. Yuen, J., Torralba, A.: A data-driven approach for event prediction. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6312, pp. 707–720. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15552-9_51
11. Ye, J., Stevenson, G., Dobson, S.: Kcar: a knowledge-driven approach for concurrent activity recognition. Pervasive Mob. Comput. **19**, 47–70 (2015)
12. Tapia, E.M., Intille, S.S., Larson, K.: Activity recognition in the home using simple and ubiquitous sensors. In: Ferscha, A., Mattern, F. (eds.) Pervasive 2004. LNCS, vol. 3001, pp. 158–175. Springer, Heidelberg (2004). doi:10.1007/978-3-540-24646-6_10
13. Bao, L., Intille, S.S.: Activity recognition from user-annotated acceleration data. In: Ferscha, A., Mattern, F. (eds.) Pervasive 2004. LNCS, vol. 3001, pp. 1–17. Springer, Heidelberg (2004). doi:10.1007/978-3-540-24646-6_1
14. Van Kasteren, T., Noulas, A., Englebienne, G., Kröse, B.: Accurate activity recognition in a home setting. In: Proceedings of the 10th International Conference on Ubiquitous Computing, pp. 1–9. ACM (2008)
15. Wu, W., Dasgupta, S., Ramirez, E.E., Peterson, C., Norman, G.J.: Classification accuracies of physical activities using smartphone motion sensors. J. Med. Internet Res. **14**, e130 (2012)
16. Zhu, Y., Nayak, N.M., Roy-Chowdhury, A.K.: Context-aware activity recognition and anomaly detection in video. IEEE J. Sel. Top. Sig. Proces. **7**, 91–101 (2013)
17. Ribbe, M.W., Ljunggren, G., Steel, K., Topinkova, E., Hawes, C., Ikegami, N., Henrard, J.C., JÓNnson, P.V.: Nursing homes in 10 nations: a comparison between countries and settings. Age Ageing **26**, 3–12 (1997)
18. Sundermeyer, M., Schlüter, R., Ney, H.: Lstm neural networks for language modeling. In: Interspeech, pp. 194–197 (2012)
19. Sak, H., Senior, A., Beaufays, F.: Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: Fifteenth Annual Conference of the International Speech Communication Association (2014)
20. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**, 1735–1780 (1997)
21. Gers, F.A., Schmidhuber, J., Cummins, F.: Learning to forget: continual prediction with lstm. Neural Comput. **12**, 2451–2471 (2000)
22. Gers, F.A., Schraudolph, N.N., Schmidhuber, J.: Learning precise timing with lstm recurrent networks. J. Mach. Learn. Res. **3**, 115–143 (2002)
23. Zhang, S., Zheng, D., Hu, X., Yang, M.: Bidirectional long short-term memory networks for relation classification. In: PACLIC (2015)
24. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
25. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: representing model uncertainty in deep learning. In: Balcan, M.F., Weinberger, K.Q. (eds.) Proceedings of The 33rd International Conference on Machine Learning (ICML), vol. 48, pp. 1050–1059. PMLR (2016)

5.3 Publication VI: Convolutional and RNN for Activity Recognition in Smart Environment

Convolutional and Recurrent Neural Networks for Activity Recognition in Smart Environment

Deepika Singh[1(✉)], Erinc Merdivan[1,4], Sten Hanke[1], Johannes Kropf[1], Matthieu Geist[2,3,4], and Andreas Holzinger[5]

[1] AIT Austrian Institute of Technology, Wiener Neustadt, Austria
{deepika.singh,erinc.merdivan}@ait.ac.at
[2] Université de Lorraine, LORIA, UMR 7503, 54506 Vandoeuvre-lès-Nancy, France
[3] CNRS, LORIA, UMR 7503, 54506 Vandoeuvre-lès-Nancy, France
[4] LORIA, CentraleSupélec, Université Paris-Saclay, 57070 Metz, France
[5] Holzinger Group, HCI-KDD, Institute for Medical Informatics/Statistics, Medical University Graz, Graz, Austria

Abstract. Convolutional Neural Networks (CNN) are very useful for fully automatic extraction of discriminative features from raw sensor data. This is an important problem in activity recognition, which is of enormous interest in ambient sensor environments due to its universality on various applications. Activity recognition in smart homes uses large amounts of time-series sensor data to infer daily living activities and to extract effective features from those activities, which is a challenging task. In this paper we demonstrate the use of the CNN and a comparison of results, which has been performed with Long Short Term Memory (LSTM), recurrent neural networks and other machine learning algorithms, including Naive Bayes, Hidden Markov Models, Hidden Semi-Markov Models and Conditional Random Fields. The experimental results on publicly available smart home datasets demonstrate that the performance of 1D-CNN is similar to LSTM and better than the other probabilistic models.

Keywords: Deep learning · Convolutional neural networks · 1D-CNN · LSTM · Activity recognition · Smart homes

1 Introduction

The advancement in sensing, networking and ambient intelligence technologies has resulted in emergence of smart environments and different services for a better quality of life and well being of the aging population. The aim are services providing comfort and security in their private space. Among them, the research in Smart Home (SH) has gained a lot of interest in the field of Ambient Assisted Living (AAL) technologies. The motivation behind the smart home research is the rapid increase in the world's aging population. According to the World Health Organization (WHO), the number of older people (aged 60 years or above) has

A. Holzinger et al. (Eds.): Integrative Machine Learning, LNAI 10344, pp. 194–205, 2017.
https://doi.org/10.1007/978-3-319-69775-8_12

increased substantially in the past decade and expected to reach about 2 billion by 2050 [1].

The concept of smart homes gained popularity in early 2000s. Lutolf [2] defined smart home concept as the integration of different services within a home environment by using a common communication system. According to Satpathy [3] a smart home provides independence and comfort to the residents by using all mechanical and digital devices interconnected in a network and able to communicate with the user to create an interactive space.

Smart home equipped with simple, easy to install and low cost interconnected sensors are providing variety of services such as health care, well being, energy conservation by ensuring safety and security to the residents. Activity recognition in the home environment facilitates the remote monitoring for the purpose of detecting so called Activities of Daily Living (ADL), residents' behavior and their interaction with the smart environment. The large amount of data collected from the installed sensors is analyzed by employing machine learning models to detect meaningful features and abnormal behavioral patterns in ADLs. Several models have been proposed to recognize the activities inside smart homes using intrusive and non-intrusive approaches. Activity recognition by intrusive approaches is opposed to ethical aspects, e.g. devices such as video cameras, microphones in private environment raise privacy concern and therefore, unlikely to be accepted by the residents. On the other hand, non-intrusive approaches are preferable as they include simple and ubiquitous sensors to measure activities of the residents and the surroundings without hindering their privacy.

In the recent years, there has been extensive interest in deep learning in the field of image analysis [4], speech recognition [5] and sensor informatics [6]. Activity recognition using deep learning has several advantages in terms of system performance and flexibility. It provides an effective tool for extracting high-level feature hierarchies from high-dimensional sensory data which is useful for classification and regression tasks [7]. Deep learning models are based on learning representations from raw data and contain more than one hidden layer. The network learns many layers of non-linear information processing for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. The well known deep learning models include Long Short Term Memory (LSTM) [8], Convolutional Neural Network (CNN) [9], Deep Belief Network (DBN) [10] and autoencoders [11].

In this work we exploit activity recognition using convolutional neural network model on publicly available smart homes dataset [12] which is an extension of our previous work of activity recognition using LSTM model [13]. The classification of the daily human activities such as cooking, bathing and sleeping is performed using temporal 1D-CNN model and evaluation of results has been carried out in comparison with LSTM and other machine learning algorithms such as Naive Bayes, Hidden Markov Model (HMM), Hidden Semi-Markov Model (HSMM) and Conditional Random Fields (CRF).

The paper is structured in different sections; the introduction is followed by Sect. 2 which presents an overview of existing work in activity recognition using various machine learning techniques in the field of AAL. Section 3 introduces Long Short-Term Memory and Convolutional Neural Network model. Section 4 describes the datasets that were used and explains the results. Finally, Sect. 5 discusses the outcomes of the experiments and suggestions for future work.

2 Related Work

The section is divided into three parts. The first part gives an overview of the existing smart home projects in the field of AAL. The second lists the available smart home datasets and the last part presents the existing work in activity recognition using machine learning techniques.

2.1 Existing Smart Homes

Several smart home projects have been implemented in the past decade, which use sensors for activity recognition inside the home environment. The Gator-tech smart house built by University of Florida contained smart appliances equipped with sensors such as smart blinds, smart refrigerator, smart stove which monitor user activities and provide services to the residents [14]. The Aware Home developed by Georgia Institute of Technology uses radio frequency identification (RFID) tags for the localization of the resident [15]. For the purpose of activity recognition, House_n project has been developed by Massachusetts Institute of Technology [16]. Various sensors have been installed to detect the routined activities such as toileting, bathing and grooming using supervised learning algorithms. The Center for Advanced Studies in Adaptive Systems (CASAS) introduced smart home in a box technology which is easy to install and provides various services with no customization and training [17]. Several other smart environment efforts have been demonstrated such as Easy Living project of Microsoft implements an intelligent environment to track and identify multiple residents through an active badge system [18]. In all of the smart home projects [19], activity recognition plays an important role.

2.2 Publicly Available Dataset

There has been several efforts to collect datasets from the sensors installed in the smart homes for human activity recognition. As these datasets are important for the research community since collecting real house annotated datasets is costly, time consuming and difficult to obtain.

The publicly available datasets are useful as they provide the baseline for the comparison of different machine learning algorithms. Eventually, it helps in collecting real house dataset using the baseline by identifying loopholes (if any) and corresponding improvement. The Table 1 summarizes the widely used publicly available smart home datasets.

Table 1. Publicly available smart home datasets

Dataset	Number of houses	Residents	Number of sensors	Number of activities
CASAS [17]	7	Multi	20 – 86	11
Kasteren [12]	3	Single	14 – 21	10 – 16
Ordonez [20]	2	Single	12	10 – 11
House_n [16]	2	Single	77 – 84	9 – 13
ARAS [21]	2	Multi	20	12 – 14
HIS [22]	1	Multi	20 – 30	7
OPPURTUNITY [23]	1	Multi	72	15 – 20

2.3 Activity Recognition in AAL

Activity recognition in smart home has been viewed as a promising approach to improve healthcare services and providing independent life to the older people. Activity recognition in SH has been broadly classified into two approaches: data driven and knowledge driven approaches. Data driven approaches use probabilistic and statistical models to learn from the datasets. Various supervised techniques have been widely used for classification of activity such as Naive Bayes [16], HMM [24], CRF and Support Vector Machine (SVM) [22]. The data driven approaches support modeling of uncertainty and temporal information but require large dataset to learn the model. Knowledge driven approaches do not require large datasets. Instead, these approaches use domain knowledge and prior heuristics to generate activity models, which make the model static and limited to specific representation of ADLs [25]; additionally, they cannot handle uncertainty and temporal parameters. Knowledge driven approaches for activity models and recognition are categorized into three types: Mining-based approach, Logic-based approach and Ontology-based approach.

The recent research has been focusing on activity recognition using unsupervised and semi-supervised machine learning algorithms to minimize the need of user annotation of activity datasets which requires a considerable amount of time and effort. The K-means clustering [26] is the most common as it performs best in temporal complexity and cluster set flexibility for large sensor datasets. The other clustering algorithms include hierarchical clustering, Density based clustering (DBSCAN) and Self-organizing map (SOM) which provides higher accuracy with random datasets [27]. However, the above clustering algorithms have some ambiguity in processing noise in the dataset. In addition, active learning techniques [28] have been used in the training process to improve accuracy in recognizing and forecasting activities in smart home.

3 Deep Learning

Nowadays, activity recognition using deep learning has become one of the most preferred techniques owing to its ability to learn data representations and classifiers. Their performance on different activity recognition tasks has been explored by researchers [29,30]. Deep architectures with multiple layers of Restricted Boltzmann Machines (RBM) handle binary sensory data and use DBN-ANN and DBN-R algorithms for human behavior prediction [31]. Convolutional neural networks [32] are type of deep neural network (DNN) which use convolutions over the input layer to compute the output. Each layer applies different filters to extract hierarchical features, dependency, translation equi-variance of data and automates feature learning, which make it suitable to use for time series raw sensor data. The CNN model has performed well in extracting features and recognizing activities from the raw sensor data [33] and video frames in comparison to the other machine learning approaches on publicly available datasets [34]. In this paper, we performed 1D-CNN on publicly available smart home datasets.

3.1 LSTM Model

LSTM, proposed by [35], is a recurrent neural network architecture which is capable of learning long term dependencies. LSTM has been developed in order to deal with gradient decay or gradient blow-up problems and can be seen as a deep neural network architecture when unrolled in time. The LSTM layer's main component unit is called memory cell. A memory cell is composed of four main elements: an input gate, a neuron with self-recurrent connection, a forget gate and an output gate [13]. The input provided to the LSTM controls the operations to be performed by the gates in the memory cell: write (input gate), read (output gate) and reset (forget gate). Following equations explain the way a layer of memory cells is updated at each timestep t.

$$
\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i),\\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f),\\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o),\\
c_t &= f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c),\\
h_t &= o_t \tanh c_t,
\end{aligned}
$$

where W_i, W_f, W_o are the weight matrix and x_t is the input to the memory cell layer at time t, σ being the sigmoid and tanh is the hyperbolic tangent activation function. The terms i, f and o are the input gate, forget gate and output gate. The term c represents the memory cell and b_i, b_f, b_c and b_o are bias vectors.

Figure 1 illustrates an LSTM single cell layer at time t where x_t,h_t and y_t are the input, hidden and output state.

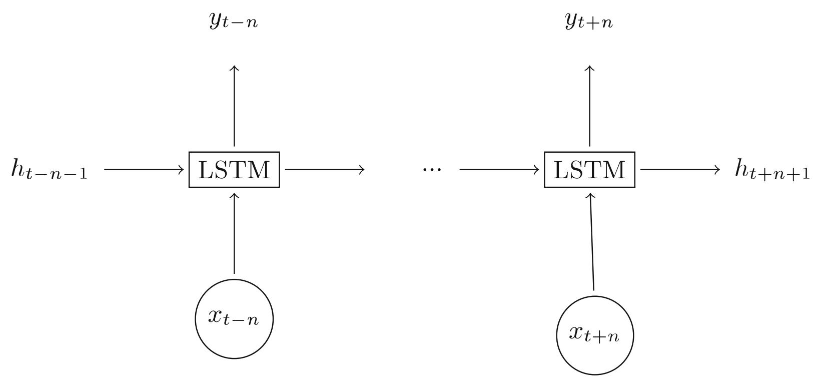


Fig. 1. Illustrations of an LSTM network with x being the binary vector for sensor input and y being the activity label prediction of the LSTM network.

3.2 CNN Model

Convolutional neural network is a type of deep neural network, consists of multiple hidden layers which can be either convolutional, pooling or fully connected. A single convolutional layer of CNN extracts features from the input signal through convolution operation of the signal with a kernel. The activation of a unit in a CNN represents the output of the convolution of the kernel with the input signal. CNNs are able to learn hierarchical data representations for fast feature extraction and classification. The CNN model has advantages when used for activity recognition task [36]. It can capture local dependencies of the activity signal and preserves feature scale invariant, thus able to capture variations in the similar activity efficiently through feature extraction. Figure 2 shows the structure of CNN for Activity Recognition.

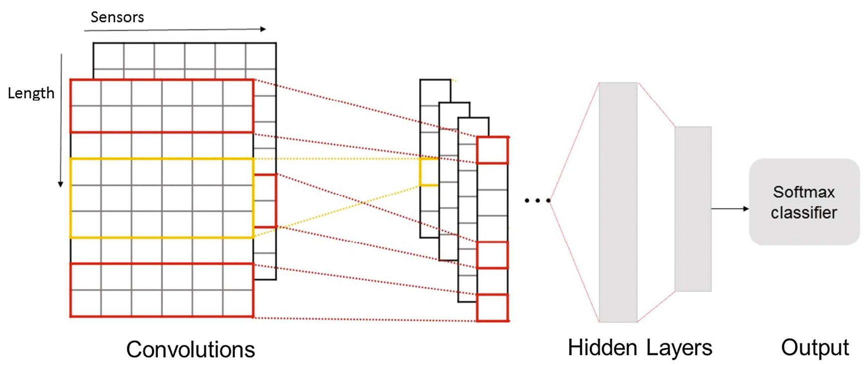


Fig. 2. CNN for activity recognition

1D temporal convolutional model used in this work, has four layers: (1) an input layer, (2) convolution layer with multiple feature widths and feature map, (3) fully connected layer and (4) the output layer.

4 Experiments

4.1 Dataset

Kasteren publicly available annotated sensor datasets of three houses, as listed in Table 1 have been used to evaluate the performance of the proposed approach. The binary sensors installed in each house to monitor activities are passive infrared (PIR) motion detectors to detect motion in a specific area, pressure sensors on couches and beds to identify the user's presence, reed switches on cupboards and doors to measure open or close status and float sensors in the bathroom to measure toilet being flushed or not. The details of the three houses with the information of the resident, the sensors and the number of activity labels are provided in Table 2.

Table 2. Details of the datasets.

	House A	House B	House C
Age	26	28	57
Gender	Male	Male	Male
Setting	Apartment	Apartment	House
Rooms	3	2	6
Duration	25 days	14 days	19 days
Sensors	14	23	21
Activities	10	13	16
Annotation	Bluetooth	Diary	Bluetooth

The data in the experiments are represented in two different forms. The first is raw sensor data, which are the data received directly from the sensor. The second form is last-fired sensor data. The last firing sensor gives continuously 1 and changes to 0 when another sensor changes its state. For each house, we performed leave-one-out cross validation and repeated this for every day and for each house. Separate models are trained for each house since the number of sensors varies and a different user resides in each house. Sensors are recorded at one-minute interval for 24 h, which totals in 1440 length input for each day.

4.2 Results

The results presented in Table 3 show the performance (accuracy) of the 1D-CNN model together with LSTM on raw sensor data and Table 4 shows the results of the last-fired sensor data in comparison with the results of Naive Bayes, HMM, HSMM and CRF [12]. We calculated accuracy of the model, which represents the correctly classified activities in each time. For the LSTM model, a time slice of (70) with hidden state size (300) are used. We implemented 1D(temporal) convolution with a time slice of (15). 128 filters are used for each layer and 1D kernel sizes were 5, 5, 3, 3, 3, 3 with a fully connected layer of 128 in the end. Dropout of 0.5 is used in order to reduce the overfitting in the data. We also tested longer timeslices but they tend to overfit considerably. Adam method [37] is used with a learning rate of 0.0004 for optimization of the networks and Tensorflow library of Python has been used to implement the CNN and LSTM network. The training took place on a Titan X GPU and the time required to train one day for one house is 4 min for CNN and approximately 30 min for LSTM, but training time differs amongst the houses. Since different houses have different number of days of data, we calculated the average accuracy amongst all days. The training is performed using a single GPU but the trained models can be used for inference without losing performance when there is no GPU.

Table 3. Results of raw sensor data

Model	House A	House B	House C
1D-CNN*	88.2 ± 8.6	79.4 ± 20.1	49.2 ± 25.6
LSTM**	$\mathbf{89.8 \pm 8.2}$	$\mathbf{85.7 \pm 14.3}$	$\mathbf{64.22 \pm 21.9}$
Naive Bayes	77.1 ± 20.8	80.4 ± 18.0	46.5 ± 22.6
HMM	59.1 ± 28.7	63.2 ± 24.7	26.5 ± 22.7
HSMM	59.5 ± 29.0	63.8 ± 24.2	31.2 ± 24.6
CRF	89.8 ± 8.5	78.0 ± 25.9	46.3 ± 25.5

*Current work, **Previous work [13]

Each model for each house is trained with leave-one-day out strategy. If house has k days of data k-1 days are used to train and 1 day is used to test and this processed is repeated for each day. In order to compare models average accuracy with variance are calculated. Table 3 shows the average accuracy with the variance of accuracies of different models on raw data from three different houses. Among all the models, the LSTM performs the best for all three datasets and 1D-CNN performs second best. In House B and House C, LSTM improves the best result significantly especially on House C where the improvement is approximately 40% from CRF and 30% from CNN.

Table 4. Results of last-fired sensor data

Model	House A	House B	House C
1D-CNN*	95.3 ± 2.8	86.8 ± 12.7	86.23 ± 12.4
LSTM**	95.3 ± 2.0	88.5 ± 12.6	85.9 ± 10.6
Naive Bayes	95.3 ± 2.8	86.2 ± 13.8	87.0 ± 12.2
HMM	89.5 ± 8.4	48.4 ± 26.0	83.9 ± 13.9
HSMM	91.0 ± 7.2	67.1 ± 24.8	84.5 ± 13.2
CRF	$\mathbf{96.4 \pm 2.4}$	$\mathbf{89.2 \pm 13.9}$	$\mathbf{89.7 \pm 8.4}$

*Current work, **Previous work [13]

Table 4 shows the accuracy on last fired data from three different houses. The 1D-CNN matches the best performance achieved by CRF in case of House A but drops slightly in case of House B and C. In comparison to LSTM, 1D-CNN performs similar except a slight decrease in case of House B. It is also important to notice the high variance in all models. Variance is halved for the last-fired sensor data compared to raw sensor data.

5 Conclusions and Future Work

In this work, we used deep learning techniques for activity recognition from raw sensory inputs in smart home environment. As data preprocessing and feature engineering is expensive for real world applications especially in AAL environment, the prediction from raw input data can eliminate most of the feature engineering efforts performed by humans. Deep learning models (1D-CNN and LSTM) lead to significant improvement in performance, especially on raw data in comparison to existing probabilistic models such as Naive Bayes, HMM, HSMM and CRF. In case of last fired data, both deep learning models (1D-CNN and LSTM) match the best performance of existing models. Although, LSTM gives better performance than CNN in general cases, but when it comes to training times CNN is much faster than LSTM based models. The selection of CNN over LSTM can help in reducing time to design a prototype and see if there is a temporal dependence between input and output. In addition, CNN also helps in evaluating performance of different architectures to achieve best results which could be very time consuming with LSTM.

In general, there are many future research directions of deep learning approaches in medical applications and specifically, in ambient assisted living scenarios. One problem in the medical domain is that the deep learning approaches are so-called black-box approaches, thus are lacking transparency. However, in the medical domain trust and acceptance among end-users is of eminent importance. Consequently, a big research challenge will emerge through rising legal and privacy aspects, e.g. with the new European General Data Protection Regulations [38], it will become a necessity to explain why a decision has been made [39]. An interesting emerging field in the ever-increasing complexity and large number of heterogeneous sensors (in sensor networks of ambient assistant living) is to combine deep learning approaches with graphical and topological approaches [40], which leads to geometric deep learning [41]. This is just to outline the enormous potential of future research in the area of deep learning applied to ambient assisted living - as part of health systems.

Nevertheless, the immediate future work will be focusing on combining CNN with LSTM to utilize the fast training time with high accuracy. More detailed architectures search for only CNN based models is also planned to be performed. In order to avoid overfitting, different methods will be investigated for a better generalization. It could also be interesting to investigate the high variance experienced in some cases.

Acknowledgement. This work has been funded by the European Union Horizon2020 MSCA ITN ACROSSING project (GA no. 616757). The authors would like to thank the members of the project's consortium for their valuable inputs.

References

1. United Nations, Department of Economic and Social Affairs: United nations department of economic and social affairs, population division: World population ageing 2015 (2015)
2. Lutolf, R.: Smart home concept and the integration of energy meters into a home based system. In: Seventh International Conference on Metering Apparatus and Tariffs for Electricity Supply, 1992, IET, pp. 277–278 (1992)
3. Satpathy, L.: Smart housing: technology to aid aging in place: new opportunities and challenges. Ph.D. thesis, Mississippi State University (2006)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
5. Deng, L., Li, J., Huang, J.T., Yao, K., Yu, D., Seide, F., Seltzer, M., Zweig, G., He, X., Williams, J., et al.: Recent advances in deep learning for speech research at microsoft. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8604–8608. IEEE (2013)
6. Längkvist, M., Karlsson, L., Loutfi, A.: A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recogn. Lett. **42**, 11–24 (2014)
7. Salakhutdinov, R.: Learning deep generative models. Ann. Rev. Stat. Appl. **2**, 361–385 (2015)
8. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems, pp. 3104–3112 (2014)
9. Matsugu, M., Mori, K., Mitari, Y., Kaneda, Y.: Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Netw. **16**(5), 555–559 (2003)
10. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)
11. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science **313**(5786), 504–507 (2006)
12. Kasteren, T.L., Englebienne, G., Kröse, B.J.: Human activity recognition from wireless sensor network data: benchmark and software. In: Chen, L., Nugent, C., Biswas, J., Hoey, J. (eds.) Activity Recognition in Pervasive Intelligent Environments, pp. 165–186. Atlantis Press, Amsterdam (2011)
13. Singh, D., Merdivan, E., Psychoula, I., Kropf, J., Hanke, S., Geist, M., Holzinger, A.: Human Activity recognition using recurrent neural networks. In: Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-MAKE 2017. LNCS, vol. 10410, pp. 267–274. Springer, Cham (2017). doi:10.1007/978-3-319-66808-6_18
14. Helal, S., Mann, W., El-Zabadani, H., King, J., Kaddoura, Y., Jansen, E.: The gator tech smart house: a programmable pervasive space. Computer **38**(3), 50–60 (2005)
15. Kidd, C.D., et al.: The aware home: a living laboratory for ubiquitous computing research. In: Streitz, N.A., Siegel, J., Hartkopf, V., Konomi, S. (eds.) CoBuild 1999. LNCS, vol. 1670, pp. 191–198. Springer, Heidelberg (1999). doi:10.1007/10705432_17

16. Tapia, E.M., Intille, S.S., Larson, K.: Activity recognition in the home using simple and ubiquitous sensors. In: Ferscha, A., Mattern, F. (eds.) Pervasive 2004. LNCS, vol. 3001, pp. 158–175. Springer, Heidelberg (2004). doi:10.1007/978-3-540-24646-6_10
17. Cook, D.J., Crandall, A.S., Thomas, B.L., Krishnan, N.C.: Casas: a smart home in a box. Computer **46**(7), 62–69 (2013)
18. Brumitt, B., Meyers, B., Krumm, J., Kern, A., Shafer, S.: EasyLiving: technologies for intelligent environments. In: Thomas, P., Gellersen, H.-W. (eds.) HUC 2000. LNCS, vol. 1927, pp. 12–29. Springer, Heidelberg (2000). doi:10.1007/3-540-39959-3_2
19. Alam, M.R., Reaz, M.B.I., Ali, M.A.M.: A review of smart homespast, present, and future. IEEE Trans. Syst. Man Cybern. Part C (Applications and Reviews) **42**(6), 1190–1203 (2012)
20. Ordónez, F.J., de Toledo, P., Sanchis, A.: Activity recognition using hybrid generative/discriminative models on home environments using binary sensors. Sensors **13**(5), 5460–5477 (2013)
21. Alemdar, H., Ertan, H., Incel, O.D., Ersoy, C.: Aras human activity datasets in multiple homes with multiple residents. In: Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), pp. 232–235 (2013)
22. Fleury, A., Noury, N., Vacher, M.: Supervised classification of activities of daily living in health smart homes using svm. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2009, pp. 6099–6102. IEEE (2009)
23. Roggen, D., Calatroni, A., Rossi, M., Holleczek, T., Förster, K., Tröster, G., Lukowicz, P., Bannach, D., Pirkl, G., Ferscha, A., et al.: Collecting complex activity datasets in highly rich networked sensor environments. In: 2010 Seventh International Conference on Networked Sensing Systems (INSS), pp. 233–240. IEEE (2010)
24. Monekosso, D.N., Remagnino, P.: Anomalous behavior detection: supporting independent living. In: Monekosso, D., Remagnino, P., Kuno, Y. (eds.) Intelligent Environments, pp. 33–48. Springer, London (2009)
25. Chen, L., Hoey, J., Nugent, C.D., Cook, D.J., Yu, Z.: Sensor-based activity recognition. IEEE Trans. Syst. Man Cybern. Part C (Applications and Reviews) **42**(6), 790–808 (2012)
26. Lapalu, J., Bouchard, K., Bouzouane, A., Bouchard, B., Giroux, S.: Unsupervised mining of activities for smart home prediction. Procedia Comput. Sci. **19**, 503–510 (2013)
27. Li, C., Biswas, G.: Unsupervised learning with mixed numeric and nominal data. IEEE Trans. Knowl. Data Eng. **14**(4), 673–690 (2002)
28. Longstaff, B., Reddy, S., Estrin, D.: Improving activity classification for health applications on mobile devices using active and semi-supervised learning. In: 2010 4th International Conference on- Pervasive Computing Technologies for Healthcare (PervasiveHealth), pp. 1–7. IEEE (2010)
29. Hammerla, N.Y., Halloran, S., Ploetz, T.: Deep, convolutional, and recurrent models for human activity recognition using wearables. arXiv preprint arXiv:1604.08880 (2016)
30. Yang, J., Nguyen, M.N., San, P.P., Li, X., Krishnaswamy, S.: Deep convolutional neural networks on multichannel time series for human activity recognition. IJCA **I**, 3995–4001 (2015)

31. Choi, S., Kim, E., Oh, S.: Human behavior prediction for smart homes using deep learning. In: RO-MAN, 2013 IEEE, pp. 173–179. IEEE (2013)
32. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016). http://www.deeplearningbook.org
33. Chen, Y., Xue, Y.: A deep learning approach to human activity recognition based on single accelerometer. In: 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1488–1492. IEEE (2015)
34. Geng, C., Song, J.: Human action recognition based on convolutional neural networks with a convolutional auto-encoder. In: 5th International Conference on Computer Sciences and Automation Engineering (ICCSAE 2015) (2015)
35. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
36. Zeng, M., Nguyen, L.T., Yu, B., Mengshoel, O.J., Zhu, J., Wu, P., Zhang, J.: Convolutional neural networks for human activity recognition using mobile sensors. In: 2014 6th International Conference on Mobile Computing, Applications and Services (MobiCASE), pp. 197–205. IEEE (2014)
37. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
38. Barnard-Wills, D.: The technology foresight activities of european union data protection authorities. Technol. Forecast. Soc. Change **116**, 142–150 (2017)
39. Holzinger, A., Plass, M., Holzinger, K., Crisan, G.C., Pintea, C.M., Palade, V.: A glass-box interactive machine learning approach for solving np-hard problems with the human-in-the-loop. arXiv:1708.01104 (2017)
40. Holzinger, A.: On topological data mining. In: Holzinger, A., Jurisica, I. (eds.) Interactive Knowledge Discovery and Data Mining in Biomedical Informatics. LNCS, vol. 8401, pp. 331–356. Springer, Heidelberg (2014). doi:10.1007/978-3-662-43968-5_19
41. Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P.: Geometric deep learning: going beyond euclidean data. IEEE Sig. Process. Mag. **34**(4), 18–42 (2017)

5.4 Publication VII: Handling Imbalanced Data in Deep Learning for Multiple Resident Activity Recognition

Handling Imbalanced Data in Deep Learning for Multiple Resident Activity Recognition

Deepika Singh, Erinc Merdivan, Johannes Kropf and Andreas Holzinger

Abstract—Multiple residents' activity recognition is an important area of research in the field of active and assisted living as in real-life scenarios a home is mostly occupied by more than one resident. Moreover, the distribution of activities of residents in a home environment is not balanced as some activities are performed frequently, while some activities are performed occasionally by the residents. Therefore, handling class imbalance is very important for developing robust activity recognition systems from raw sensor data in a smart home environment. Deep learning methods have shown their effectiveness in many areas, including single resident activity recognition systems on balanced datasets. In this paper, we have utilized Long Short Term Memory (LSTM) and Bidirectional Long Short Term Memory (BiLSTM), deep learning model, to evaluate the performance of multiple resident activity recognition using separate activity labels as well as combined activity labels of the residents. We have conducted extensive experiments on both data level and algorithm level approaches on both the deep learning models to address class imbalance problems and compared their performances on three highly imbalanced smart home datasets.

Index Terms—Human activity recognition, multiple residents; smart homes; class imbalance; deep learning

1 Introduction

Human activity recognition in an intelligent environment is a highly dynamic research area which has gained a lot of attention due to its varied applications. The applications of activity recognition systems are categorized as: active and assisted living systems for smart homes (SH), monitoring and surveillance systems for indoor and outdoor, health care monitoring and tele-immersion applications [1], [2]. Among these, SH plays an important role, especially in user behavior analyses, health monitoring, and assistance. Most of the research on activity recognition in SH have investigated on single resident activity monitoring [3] [4] [5] [6]. However, in real-life scenarios, a home is not always occupied by a single resident but often occupied by more than one resident. Therefore, developing an SH solution from the perspective of multiple residents is extremely crucial.

In recent years, there has been an increase in multiple occupancy-based research related to activity modeling and data association. However, there are still various challenges to be addressed in multiple occupancies, such as finding the suitable models for data association i.e. identification of the residents by whom each sensor is triggered and capturing interactions between the occupants [7]. Another major challenge that occurs while developing real-life applications is the class imbalance problem. Activity recognition is mainly considered as a classification problem where the performance of the system depends on the model selection, features involved, number of classes, and the size of the datasets available for training the system. In most of the SH datasets, there is a lack of uniformity in different daily living activities of residents; which is obvious as in real-life situations, some activities are performed more often than the others. Hence, handling class imbalance is extremely important for developing an effective activity recognition model. Different techniques have been designed in previous works to address the imbalance problems. These techniques are divided into resampling methods, cost-sensitive learning, kernel-based methods, and ensemble methods. Resampling methods are mainly divided into oversampling [8] and undersampling [9] methods. Oversampling methods duplicate samples of the minority class and then augment them into the original dataset, whereas undersampling randomly removes a certain number of instances from the majority class in order to achieve a balanced dataset. However, sampling techniques sometimes cause problems: random oversampling may lead to overfitting while random undersampling may lose some important information from the dataset. Cost-sensitive learning [10] and ensemble learning [11] are algorithmic level methods that are achieved by improving existing algorithms. The cost-sensitive learning method assigns misclassification costs to different classes by modifying the loss function such that the misclassification cost of the minority class is higher than the cost of the majority class. In Kernal based methods, there are many works which use sampling and ensemble techniques to the support vector machine (SVM) algorithm in which different error costs were suggested to different classes to bias the SVM to shift the decision boundary away from positive samples and make them more densely distributed [12]. The ensemble learning method trains multiple classifiers to improve the reliability of a single classifier and predict the class through voting or averaging the results of all classifiers. The standard techniques for constructing ensemble classifiers are overall accuracy oriented and presented to be more effective in enhancing classification performance, moreover, they still

- *D. Singh and A. Holzinger are from Human-Centered AI Lab (Holzinger Group), Institute for Medical Informatics/Statistics, Medical University Graz, Austria. E. Merdivan and J. Kropf are from the AIT Austrian Institute of Technology GmbH, Wiener Neustadt, Austria (Email: deepika.singh@medunigraz.at, andreas.holzinger@medunigraz.at , merdivane@gmail.com and Johannes.Kropf@ait.ac.at).*

have difficulty in recognizing minority class in imbalanced dataset [13]. Therefore, a combination of ensemble learning with data level approaches has been designed, specifically to handle the class imbalance problem [14]. The most widely used ensemble learning methods are AdaBoost [15] and Bagging [16] algorithms whose applications in several classification tasks have led to significant improvements.

Although several studies have been conducted for class imbalance, there remains a lack of empirical work on addressing the class imbalance in multiple residents activity recognition. In this work, we reported an empirical study of both data-driven and algorithm-driven techniques for handling class imbalance. Data-driven approaches modify the original dataset by oversampling the minority samples and can provide a balanced distribution without losing information on the majority class. Undersampling techniques alter the dataset by removing samples from the majority class. The main advantage of undersampling lies in the reduction of the training time, which is significant in the case of highly imbalanced large datasets [17]. In algorithm level techniques, we employed cost-sensitive learning to deep learning models, which has performed well as reported in previous works in class imbalance problem. However, the majority of work uses statistical methods such as SVM and Naive Bayes as a base classifier in cost sensitive learning approach [18]. In other works, machine learning methods have been used in activity recognition which relies on feature extraction techniques including time-frequency transformation and statistical approaches. In such methods, the extracted features are carefully engineered and heuristic. There is no universal feature extraction method that can effectively capture distinguishable features of human activities. Consequently, we chose the Long short term memory (LSTM) network as it allows extracting highly discriminative non-linear feature representations while modeling temporal sequences by learning long-term dependency. In addition, LSTM and 1D-convolutional neural network outperformed other statistical machine learning models on single resident activity recognition [19].

To summarize, the main contributions of this paper are:

i employing LSTM and BiLSTM networks for multiple resident activity recognition;
ii evaluating model performance by taking each resident separately and also with combined activity labels of the residents;
iii conducting extensive experiments using both data level and algorithm level class imbalance techniques; and,
iv investigating model performance at different sample ratios and cost coefficients on three benchmark datasets.

The paper is further structured as follows: section 2 reports the related works. Section 3 introduces the SH datasets, LSTM networks and different data imbalance methods that are used in the paper and section 4 provides the description of experiments performed. Results of the paper are demonstrated and discussed in the next section, which is followed by a concluding section highlighting the major findings.

2 Related Works

In this section, we review related works on multiple resident activity recognition and imbalanced data classification approaches and discuss them in detail which eventually laid the foundation of the current work.

2.1 Multiple resident activity recognition

Activity recognition have been categorized mainly into two approaches: Vision based [20] [21] [22] and pervasive sensing based [23] [24] [25]. Vision-based activity recognition can provide good results but have raised various privacy concerns among the residents due to required camera installations in their private spaces [26] [27] whereas pervasive sensing-based activity recognition approaches use data from wearable sensors and non-intrusive environment sensors. A significant amount of work has been performed on activity recognition using wearable sensors. A new technology called Body Sensor Network (BSN) has emerged which consists of different wearable sensors that capture and process physiological signals on the human body. BSNs then collect data from wearable sensors and process them to extract useful information [28] [29]. A major issue with wearable sensors is that wearing or carrying a tag is often not feasible especially with the old people, who often forget to wear, or not willing to wear tags at all. Pervasive sensing using environment sensors offers the advantage of being non-intrusive to the inhabitants and do not require to carry any tag or device. In pervasive sensing, the sensors are deployed in the environment and capture activities of the residents, which then can be used for activity recognition. But there are some challenges in this approach as well. Recognizing human activities using environment sensors is challenging because sometimes the data captured by the sensors can be disturbed from the surroundings which can make data noisy and human activities are complex. In such a case, sensor deployment, its configuration, and selection of the classification model play an important role in the identification of activities of residents and the residents themselves.

In previous works, diverse computational models have been applied in the context of single resident activity recognition which includes standard data mining approaches, probabilistic models, and machine learning models such as neural networks, support vector machines, decision trees, and ontologies. However, for multi-resident activity recognition, such a diversity of models hasn't been used yet. The problem in multiple resident activity recognition using non-intrusive sensors is the association of sensor data when such sensors cannot directly identify residents and interactions between them, whereas, in a single resident setting, sensors' states reflect directly the activity of the sole resident. Multiple residents' activities can have different scenarios: the same activity can be performed by two or more residents (e.g. eating a meal or watching TV together) or multiple residents perform different activities independently (e.g. one resident is watching TV and the other preparing meal). Evidently, there is a need for a model that is capable of capturing the complex nature of both joint and independent activities. Previous works have addressed multiple resident activity recognition using wearable sensors such as RFID

[25], accelerometer [30] and videos [31]. Machine learning approaches used previously for multi-resident activity recognition are naive Bayes, Markov model classifier [32] and conditional random field (CRF) [33] on CASAS [34] dataset in which data association problem was investigated. In [35], the authors proposed a two-stage activity recognition method in order to exploit more knowledge in multi-resident activities. The two phases in the model include the building phase and activity recognition phase and it converts multi-label problems into a single-label problem by treating activities of residents as combined label state using HMM (Hidden Markov model) and CRF (Conditional random field) classifiers. In recent works, deep learning models have shown impressive performances in various fields. LSTM network which is variants of recurrent neural network (RNN) is good at solving time series problems as its design enables gradients to flow through time readily [36]. Deep Residual Bidirectional LSTM network has been used for activity recognition using wearable sensors on UCI dataset (which uses data from a smartphone) and Opportunity dataset (data from wearable, object, and ambient sensor) [37]. In [38], CNN (Convolutional Neural Network) and LSTM have been used for extracting Spatio-temporal features from multisensory and multimodal data which includes RFID, audio data, and videos for concurrent activity recognition. In [39], a joint diverse temporal learning framework using LSTM and 1D-CNN models has been proposed to improve human activity recognition.

However, the existing state of the art approaches in multi-resident activity recognition focuses on the improvement of recognition algorithms using accuracy as performance metrics rather than handling imbalanced dataset. Furthermore, there is a lack of comprehensive studies on how different class imbalance approaches perform in the multiple residents' activity recognition domain.

2.2 Imbalanced data classification

Existing research uses various machine learning and deep learning models for activity classification but lacks in analyzing the class imbalance problems in the dataset, through which a model achieves very high accuracy but did not reveal the actual performance of the model due to class imbalance. Likewise, in some machine learning problems, not every mistake is treated equally. This is very true in the SH setting; for example, if the system makes a mistake in detecting a resident fall, it is much more harmful than making a mistake in detecting if a user is brushing his teeth. Training with equal importance for each activity in the home environment is not suitable to provide high user experience and satisfaction. In the multi-resident setting, common methods to help with the imbalance dataset needs to be altered since instead of one classification there are multiple classifications (for each resident of the house).

Three major approaches have been defined to learn imbalanced data [40]: data-level methods, algorithm level methods, and hybrid methods. Data level methods concentrate on modifying training sets to balance the data distributions by adding or removing samples. Such methods use oversampling (addition of new sampling to minority class) and undersampling (removing samples from majority class) approaches for balancing the data distributions. This way, data level approaches to avoid the modification of the learning algorithm by decreasing the effect caused by an imbalance with a preprocessing step. Synthetic Minority Oversample Technique (SMOTE) is the popular oversampling method [41] with an idea to create new minority samples by interpolating several minority class instances that lie together. The strategy used in SMOTE is problematic as it blindly generalizes the minority class without regard to the majority class, particularly in the case of highly skewed class distribution where the minority class is very sparse in comparison to the majority class, thus resulting in a high chance of class mixture [42]. In undersampling techniques, four K-Nearest Neighbour (KNN) methods [43], namely, NearMiss-1, NearMiss-2, NearMiss-3, and the "most distant" method were proposed, in which instead of using the entire set of majority samples, a small subset of these samples was selected such that resulting training data is less skewed. Results of the experiment suggest that the NearMiss-2 method provides competitive results in comparison to SMOTE and random undersampling. Algorithm level methods modify existing learning algorithms to alleviate the bias towards majority classes. These methods require special knowledge of both the learning algorithm and the application domain, comprehending the reason behind the failure of the classifier when the class distribution is uneven. The most popular approach in such methods is cost-sensitive approaches [44]. In such approaches, the given learning algorithm is modified by incorporating varying costs for each of the considered groups of samples. Another algorithm level approach is one-class learning which focuses only on the target group and thus helps in eliminating the bias towards any group. Hybrid methods combine data-level methods and algorithm level methods by extracting strong features from both the approaches, merging data level solutions with classifier ensembles is one of the widely used hybrid approaches [45]. There exist some works that propose hybridization of cost-sensitive learning and sampling methods [46].

Numerous other works have been performed for handling class imbalance in traditional classification problems using data preprocessing and algorithm level techniques [47] [48] [49]. These studies have shown that for the various base classifier, a balanced dataset provides improved overall performance compared to an imbalanced dataset. Traditional machine learning algorithms such as SVM, Decision tree, Naive Bayes, Random forest, Hidden Markov Models, and their ensembles were used to balance between minimizing the total recognition error and maximizing the accuracy of classification on minority class [50]. The major drawback of these methods is that they rely on handcrafted and classical heuristic feature extraction techniques. Recently, deep learning methods have shown promising results in various applications such as natural language processing, image classification, speech recognition, and also in human activity recognition systems by outperforming on raw sensor datasets [3].

2.2.1 Handling class imbalance in deep learning

Some of the works with deep learning methods for handling class imbalance use CNN for representation learning [51] and proposed quintuplet sampling with a triple-header loss for imbalanced learning. Another work proposed Deep Over-sampling (DOS) with CNN architecture [52] to address the effect of class imbalance on both classifier and representation learning. Empirical results of the proposed DOS framework showed improvement in addressing the class imbalance problem. A new loss function in a deep neural network is proposed in [53] which captures classification errors from both majority and minority classes. Another method was presented in [54] to optimize the network parameters and class sensitive costs. Deep reinforcement learning has shown promising results in various applications, therefore recent work also explores the performance of deep reinforcement learning model for imbalanced classification and evaluated their approach on text and image data [55] where classification problem was formulated as a sequential decision-making process. The environment returns a high reward to minority class samples but a low reward to the majority class sample and the process was terminated when the agent misclassifies the minority class sample. Deep Q-learning was used to find the optimal classification policy for the Imbalanced Classification Markov Decision Process (ICMDP). Experiments showed better classification performance on imbalanced text datasets. The survey on class imbalance for deep learning presents classical methods such as random oversampling and cost-sensitive target function, which show promising results when applied in deep learning situations [56].

2.2.2 Handling class imbalance in activity recognition

Few studies have been performed in single resident activity recognition using improved SMOTE algorithm to address issues concerning imbalanced activity classes [57]. SMOTE [41] is the widely used algorithm as it creates new non-replicated examples by interpolating neighboring minority class instances but it fails to preserve the class covariance structure and increases overlapping between classes. Another work uses a cost-sensitive SVM approach for the classification of activities and compared the results with HMM, CRF, and traditional SVM models [58]. Most of the works on handling imbalance classes focus on vision and text classification problems but very less work has been performed in handling class imbalance in multiple resident activity recognition. In addition, existing works lack comparative studies of different class imbalance approaches.

Therefore, this paper presents a comprehensive study of both data level and algorithm level class imbalance approaches in multiple resident activity recognition. Since temporal deep learning methods have shown promising results on raw sensor datasets in single resident activity recognition, therefore we used LSTM and BiLSTM networks as a classifier for addressing the class imbalance in activity recognition.

3 Methodology

3.1 Smart home datasets

In this work we have used publicly available ARAS [59] and CASAS-Kyoto Multiresident ADL Activities dataset (fourth number dataset on Casas dataset list: http://casas.wsu.edu/datasets/) [60] [24]. ARAS is a widely used dataset in activity recognition systems whereas the CASAS-Kyoto Multiresident ADL Activities dataset has not been used much in previous works. As the collection of real SH data is time-consuming, costly, and difficult to annotate, therefore, the publicly available datasets are used to provide a baseline for comparison.

3.1.1 ARAS multi-resident ADL dataset

ARAS datasets use ambient sensors such as contact sensor, temperature sensor, sonar distance sensor, force sensor, photocells, resistors, and infrared receivers in the SH setting. The dataset consists of 20 different types of sensor signals as features together with the activity labels of two residents for two different houses which are termed as House A and House B. Each house has 30 days of a dataset with 30 separate files for a month and every file contains 86400 instances. The dataset consists of 27 different types of activities for each resident. The distribution of activities in House A and House B of the ARAS dataset are shown in Figure 1.

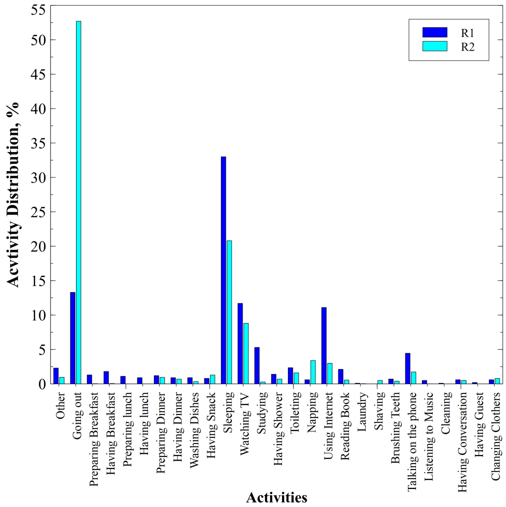
As visible from the Figure 1 that the dataset of both the residents in two houses are highly imbalance where only few activities in the distribution are more than 35% and most of the activities are less than 10% of the whole dataset.

3.1.2 CASAS-Kyoto Multiresident ADL Activities dataset

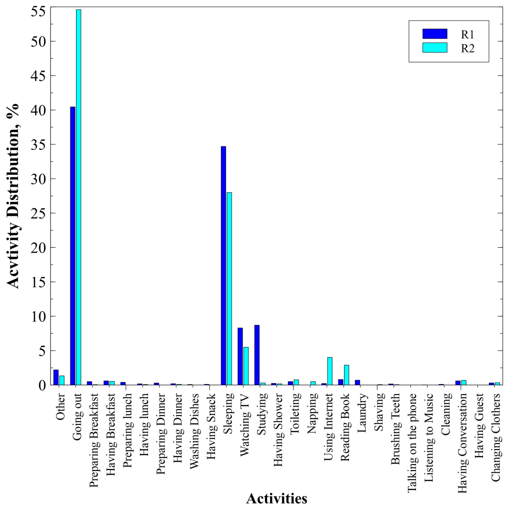
The CASAS-Kyoto Multiresident ADL Activities dataset was collected in a smart apartment testbed located at Washington State University (WSU). The sensors used in the dataset are motion, item, cabinet, water sensors, burner, phone, and temperature sensors. The smart space was occupied by two residents at the same time where they performed daily living tasks concurrently. The collected sensor events were labeled with activity and person identifications. The dataset has 15 different daily living activities that were performed by both the residents, in which few activities (moving furniture, playing checkers, paying bills, and packing picnic supplies) were jointly accomplish by both residents. Since some activities were performed jointly by both the residents and some individually, when an activity is performed by only one resident, there is no label for the activity of the other resident. As both the residents are present in the apartment, therefore in such a scenario, we assigned a label (named as Other) to the activity of the second resident which is not known, which makes our dataset of 16 activity labels for both the residents. However, in many cases, there were sensor readings for both the residents and their activity labels. The frequency distribution of activities in the dataset is shown in Figure 2.

3.2 LSTM models

LSTM networks [61] are a very successful extension of RNNs and designed to avoid the long-term dependency problem associated with RNN. LSTM model introduces a new state called cell state and constant error carousel (CEC) node which allows constant propagation of error signals over time, thus solving the problem of vanishing gradients. In addition, LSTM uses a gating mechanism over an internal memory cell to control access to CEC and to learn a more complex representation of the long-term dependencies. Therefore, LSTM is better at classifying, processing, and predicting time series data with the time lags of unknown



(a) House A



(b) House B

Fig. 1: Activity distribution of both residents (R1 and R2) in ARAS dataset

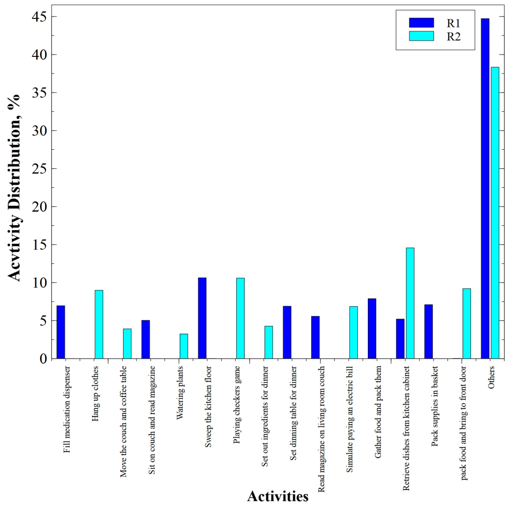


Fig. 2: Frequency count of activities in CASAS-Kyoto Multiresident ADL Activities dataset

sizes. An LSTM block consists of input, output, and forget gates which are responsible for write, read and reset operations respectively for the memory cell. The main component of LSTM is the memory cell which is responsible for remembering states for short or long periods over arbitrary time intervals. Each LSTM cell operates as a memory to write, read, and erase information based on the outcomes rendered by input, output, and forget gates respectively. Forget gate receives new time step X_t and previous output h_{t-1} as input and gives output using sigmoid activation function to decide which information will be kept or deleted. The information will be deleted if the output of the sigmoid activation function is 0, while information will be kept if the output is 1. The forget gate computation is shown in Equation (1). The next step decides what new information will be stored in the cell state. This step has two parts, first, the input gate layer decides which new information from the current input (X_t, h_{t-1}) is updated to the cell state. In the second step, tanh activation function that generates a new candidate value $\tilde{C}_t$, could be appended to the cell state. The multiplication of these two parts will be added to the multiplication of forget gate (f_t) with the previous cell state (C_{t-1}) to generate a new cell state (C_t). The forget gate (f_t) is multiplied with the previous cell state (C_{t-1}), forgetting the information which was specified to be deleted earlier. Then we append $i_t * \tilde{C}_t$, which is the new candidate value, scaled by how much the cell state is updated. The computation of the input gate, new candidate value and cell state is shown in Equation (2) - Equation (4). In the final step, the output gate is computed based on the filtered version. First, the previous hidden state and the current input time step are passed to the sigmoid activation function, and then the new state is put through $\tanh$ function. Then the output of the sigmoid function is multiplied with the output of $\tanh$ function to generate the next hidden state. The update cell state and new hidden state forward the information to the next time step. Equation (5) and (6) shows the computation of output gate and hidden state (h_t).

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{1}$$
$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{2}$$
$$\tilde{C}_t = \tanh(W_C.[h_{t-1}, x_t] + b_c) \tag{3}$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$
$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{5}$$
$$h_t = o_t * \tanh(C_t) \tag{6}$$

where σ is the sigmoid activation function, $\tanh$ is hyperbolic tangent function, x is the input data and W is the weight matrix. The LSTM equations are adapted from [62].

The architecture of LSTM and BiLSTM network is shown in Figure 3. The input layer of the network comprises an

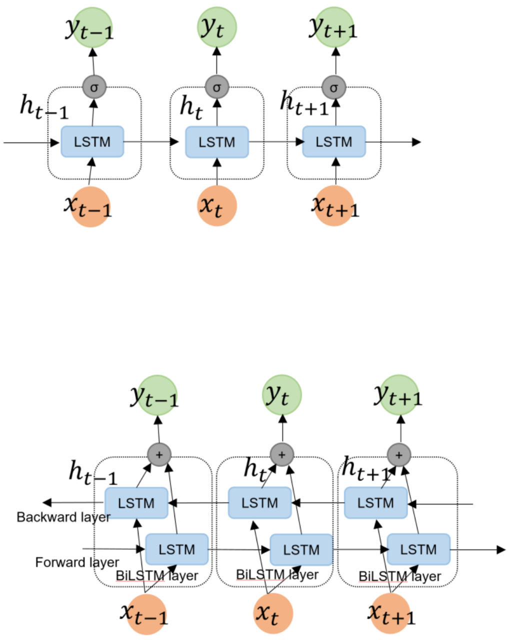


Fig. 3: LSTM and Bidirectional LSTM

embedded vector that contains a sequence of sensor events and then n LSTM cells are fully connected to the inputs and have recurrent connections with the other LSTM cells. Finally, a dense output layer of the network performs the classification task. In the BiLSTM network, two parallel LSTM are used for forward and backward loops, which extracts patterns from the past and future events. The forward layer reads the input from the left to right direction whereas the backward layer reads the input from right to left direction. The output prediction is the weighted sum of the prediction score from both the forward and backward layer. In both networks, the Adam optimizer is used for training the network and minimizing the softmax cross-entropy loss function.

3.3 Handling class imbalance in LSTM and BiLSTM networks

The following three methods are used with LSTM and BiLSTM networks to address the class imbalance problem in multiple resident activity recognition.

3.3.1 Oversampling

Oversampling is the data level approach that aims to balance the class distribution by increasing samples of the minority class. The oversampling is performed by computing the sampling ratio (also known as the class imbalance ratio) between the minority class and majority class. We selected the most frequent activity and reduced the imbalance of less frequent activities in the training set. We oversampled less frequent activities with varying sampling ratios but we never, in any case, oversampled less frequent activity to the amount where it was more frequent than the actual most frequent activity. For example: Suppose Resident 1 has 1000 samples and Resident 2 has 5000 samples and maximum activity is 10000, in this case, we threshold oversample at 2, even though we could apply oversample by 10 (if only Resident 1 was taken into consideration). We used different sampling ratios (from range 1 to 10) and conducted experiments over these ranges. The optimal difference in model performance was observed at sampling ratios 2 and 5.

3.3.2 Undersampling

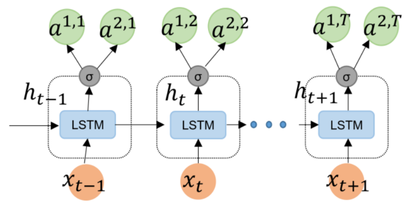
Similar to oversampling, undersampling is also a data level approach and performed by computing sampling ratio where we reduced the samples of most frequent activities of the residents. We limited the undersampling ratio in a way that the most frequent activity will still be the most frequent even after being undersampled. For example: if any of the Resident 1 or Resident 2 activity ratios are lower than average (uniform distribution for all activities in the original count) we keep these instances and do not undersample. We only undersample if both activities are over-represented and again keeping in mind that we threshold undersampling ratio taking into account average. We also tried different sampling ratios from range 0.25 to 1.0 and conducted experiments over these ranges, however, the optimal difference was observed at 0.25 and 0.5 undersample ratio.

Data level approaches are not dependent on the classifier as they avoid the modification of the learning model by reducing the effect caused due to imbalanced data with a preprocessing step. Thus, these approaches are more versatile.

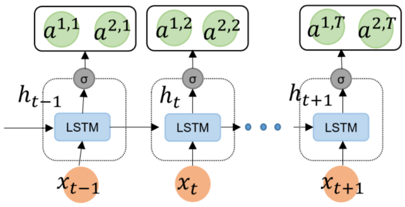
3.3.3 Cost-sensitive learning

Cost-sensitive learning lies between data level and algorithm level approach as it incorporates both data-level processing by adding costs to samples and algorithm level modifications by modifying the learning process [14]. This method evaluates the cost associated with the misclassifying samples. It does not create a balanced data distribution rather assigns the training samples of different classes with different weights, where weights are in proportion to the misclassification costs. In the cost-sensitive version, we scaled the loss according to the cost coefficients in frequent activities and limit the ratio of cost coefficient below the ratio of most frequent/given activity frequency. In this approach, we also conducted experiments with different cost coefficients (from range 1 to 10), and the best model performance was observed at cost coefficients 2 and 5.

Since the dataset is multiple residents, we performed experiments by keeping frequency ratios for residents separately and also with combined activities ratios. For example, in the case of separate activity labels: we chose activity 1 and activity 3 separately for different residents and applied class imbalance methodologies for users separately, where we always kept the most frequent activity samples more than any other activity. In the case of combined activities of both residents, we used a tuple of activities and calculate the frequency of tuple activities, such as activity (1, 3), and applied class imbalance methodologies to these tuple activities. Figure 4 depicts the sequence model for multi-resident activity recognition with each resident separately and with combined activities of residents.



(a) Model for separate activities

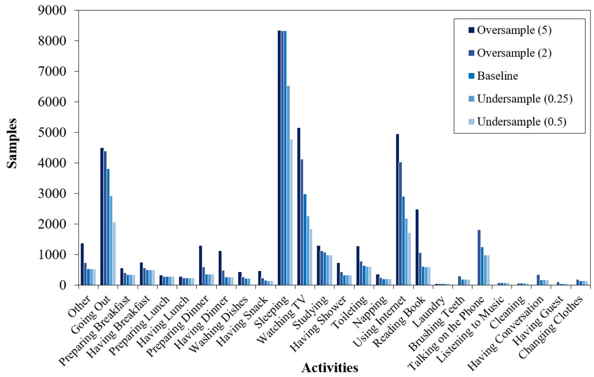


(b) Model for combined activities

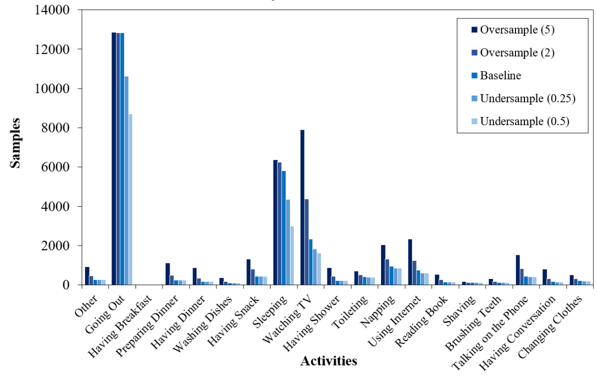
Fig. 4: LSTM model for multi-resident activity recognition

The activity distribution of ARAS and CASAS dataset after performing oversampling and undersampling at different sampling ratio for both the residents are shown in Figure 5, Figure 6 and Figure 7.

As can be seen from Figures 5, 6 and 7 that oversampling with sampling ratio = 2 or 5, does not represent multiplying each resident activity with sample ratio 2 or 5. We took into consideration that increasing one resident activity will also change the distribution of other resident activity as in the dataset we have sensor information for both the residents together. Similarly, while undersampling the dataset with a sampling ratio = 0.25 and 0.5 does not mean, reducing the activity distribution to one-fourth and half. Instead, we performed sampling such that when we undersample with 0.25 sampling ratio, we selected 0.25 probability if a certain data point should be added or not. Therefore, the exact distribution of activities may vary in each case.



(a) Resident 1 activity distribution



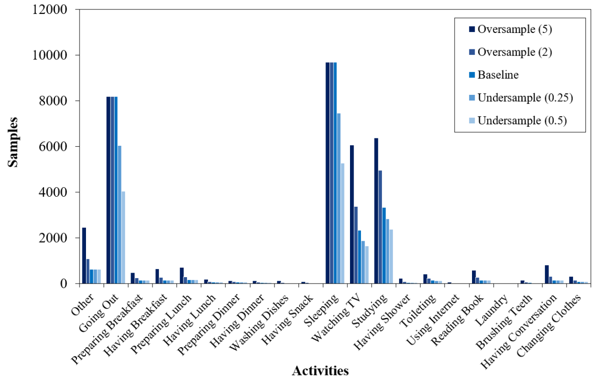
(b) Resident 2 activity distribution

Fig. 5: ARAS House A activity distribution after oversampling and undersampling

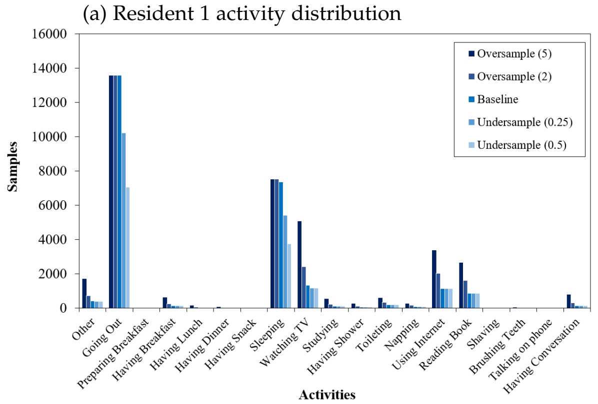
4 Experiments

The experiments are performed on three SH datasets, in which two houses (House A and House B) are from the ARAS dataset and the third house is of the CASAS dataset. Both the dataset has sensor observations of two residents. The experiments are designed such that for all the three houses classification of activities of the residents are performed using different LSTM and BiLSTM network and in each model, we explored oversampling, undersampling, and cost-sensitive learning methods to handle class imbalance problem. Each house of ARAS dataset consists of 30 days of human activities dataset, where each day consists of 86400 data points. The dataset is divided into training, validation and test set such that the first 18 days are used for training, the next six days of the dataset are used for validation and the last six days are used as a test set. In the CASAS-Kyoto Multiresident ADL activities dataset, human activities of two residents were carried out for 26 days and each file has a different number of data points. We followed a similar approach as other datasets, where the first 16 days are used as training (10572 instances), the next five days are used as validation test (3051 instances) and the last 5 days are used as test set (3608 instances of sensor readings). The experiments are computed first with the original dataset (without applying class imbalance methods) and then twelve different experiments are conducted for each model by applying class imbalance techniques to the training data.

The evaluation metrics play an important role in measuring the performance of models in handling class imbalance in multiple resident activity recognition. Hence, we used the Exact Match Ratio (EMR), Balanced accuracy, and micro average of F1-score to evaluate all the models. EMR metrics indicate the percentage of samples that have all their labels classified correctly (shown in Equation 7). Balanced accuracy metric is used in multi-class classification problems to deal with imbalanced datasets and it is based on two most commonly used metrics: sensitivity (also known as



(a) Resident 1 activity distribution

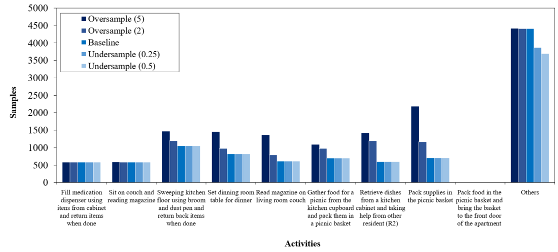


(b) Resident 2 activity distribution

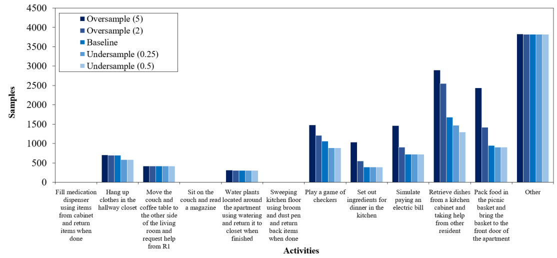
Fig. 6: ARAS House B activity distribution after oversampling and undersampling

true positive rate or recall) and specificity (also known as a false-positive rate), shown in Equation 8. Also, we used a micro average of F1-score as it is a weighted average of recall and precision, shown in Equation 9. The Exact match ratio of both residents, balanced accuracy, and micro average of F1-score of each resident of the test set are computed at best validation accuracy for all the models.

In both LSTM and BiLSTM networks, we used a range of sequence lengths from 10 to 100, a range of batch sizes from 32 to 512, and a range of several epochs from 5 to 100. A series of trial and error experiments were conducted over these ranges. We observed that epochs = 30, batch size = 64, sequence length = 30, and hidden units (n) = 128 are found to be optimal parameters to avoid overfitting and achieved a low generalization error in training both the models. The model parameters are kept the same for all the datasets. The training of the network is performed on a single Quadro RTX 4000 8GB GPU, also trained models can be used for inference without losing much performance when there is no GPU. In addition, we also performed experiments on a single NVIDIA 12GB GeForce GTX 1080Ti GPU, and the same results were observed on both the computer environment.



(a) Resident 1 activity distribution



(b) Resident 2 activity distribution

Fig. 7: CASAS-Kyoto activity distribution after oversampling and undersampling

$$ExactMatchRatio, EMR = \frac{1}{n}\sum_{i=1}^{n} I(Y_i = Z_i) \quad (7)$$

where, I is the indicator function, Y_i is target class and Z_i is predicted class.

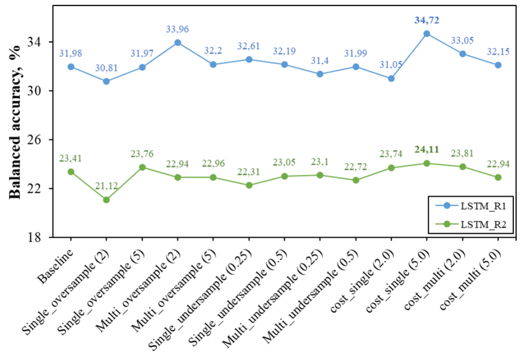
$$BalancedAccuracy = \frac{Sensitivity + Specificity}{2} \quad (8)$$

$$F1 - score = \frac{2 * (precision * recall)}{(precision + recall)} \quad (9)$$

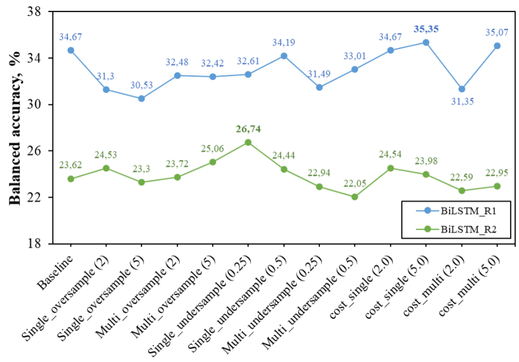
5 Results and Discussions

In this section, the experimental results of both LSTM and BiLSTM network together with different class imbalance approaches in terms of exact match ratio, balanced accuracy, and a micro average of F1-score are presented and discussed. Figure 8, 9 and Figure 10 present the balanced accuracy results of each resident of the dataset. Table 1 and Table 2 reports the results of House A (ARAS) dataset, Table 3 and Table 4 reports the results of House B (ARAS) and Table 5 and Table 6 presents the results of CASAS-Kyoto Multiresident ADL Activities dataset in terms of EMR and micro average F1-score.

As discussed in the previous section, each table shows the experiment results of the baseline model which is without applying class imbalance techniques, and then 12 different experiment results with data level and algorithm level techniques on deep learning models. In all three approaches (oversample, undersample, and cost-sensitive learning), the term "single" represents activity recognition of each resident separately and the term "multi" represents combined activity recognition for both the results. The models are evaluated at different oversample (2 and 5) and undersample (0.25 and 0.5) class ratios, together with different cost coefficient values (2 and 5) to have a detailed study and comparison of different class imbalance approaches in a multi-resident setting.



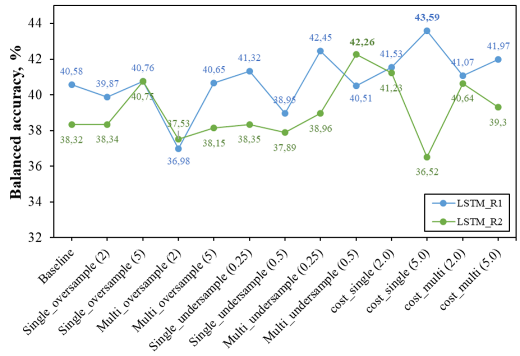
(a) Balanced accuracy results of LSTM model



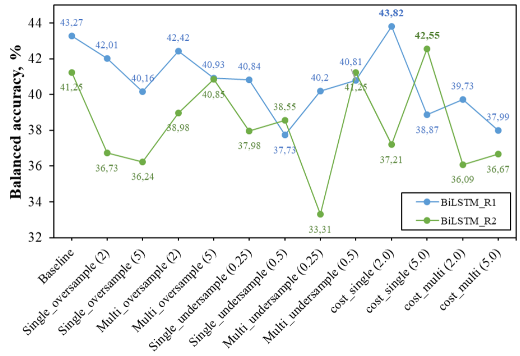
(b) Balanced accuracy results of BiLSTM model

Fig. 8: ARAS House A Balanced accuracy results

Balanced accuracy results show that a single cost-sensitive learning approach outperforms all the other class imbalance approaches in the majority of the cases. In the ARAS house A dataset, the single cost-sensitive learning approach of R1 improves by 3% in LSTM and 1% in BiLSTM in comparison to the baseline model, whereas in R2 cost-sensitive approach increases balanced accuracy by 1% in LSTM network but in BiLSTM model single-undersampling approach improves the balance accuracy by 3%. In House B of the ARAS dataset, the cost-sensitive approach performs better in both LSTM and BiLSTM models, except in the LSTM model of R2, where the undersampling approach is slightly better. In the CASAS dataset, a single cost-sensitive approach clearly outperforms all the other approaches and improves the balance accuracy results of R1 by 9% and 13% in LSTM and BiLSTM, 11% and 14% increase in balance accuracy of R2 LSTM and BiLSTM models in comparison to a baseline model. To summarize, from the following results it has been observed that in almost all the networks cost-



(a) Balanced accuracy results of LSTM model

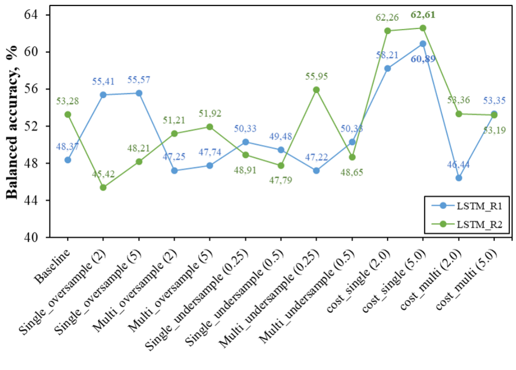


(b) Balanced accuracy results of BiLSTM model

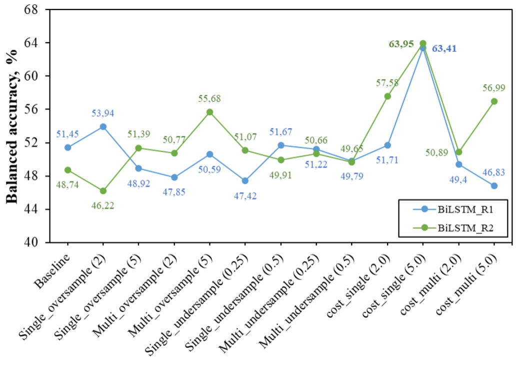
Fig. 9: ARAS House B Balanced accuracy results

sensitive learning performs better in terms of balanced accuracy. In the EMR of both residents, no clear trend has been observed as the results as in House B the difference in EMR results is minimal for both LSTM and BiLSTM networks, in House A, baseline model performed better in comparison with other models, whereas in BiLSTM network the results of EMR in undersampling and cost-sensitive approach are similar. In the CASAS-Kyoto dataset, EMR results are better in undersampling and cost-sensitive approach. The F1-score of R2 is better than R1 in the case of House A, whereas for House B high F1 score is achieved for both the residents in comparison to House A. Furthermore, in CASAS-Kyoto smart home no significant difference can be seen in F1 scores of R1 and R2.

Since each SH dataset has a different configuration, sensor readings, activity labels, and class imbalance, therefore the difference in model performance is observed in all three datasets. The computation time of the CASAS-Kyoto dataset was much faster in comparison to the ARAS dataset due to less number of sensor observations in each day of the dataset. In terms of model computational time, the undersampling method was faster to train in comparison to oversampling and cost-sensitive learning, where multi-oversampling took a quite long time to train which is quite obvious due to the increase in the number of samples to train the models. Among the deep learning models, training with LSTM was faster in comparison to the BiLSTM model.



(a) Balanced accuracy results of LSTM model



(b) Balanced accuracy results of BiLSTM model

Fig. 10: CASAS-Kyoto Balanced accuracy results

Figure 11 shows the computational time of both the models for all the three datasets.

TABLE 1: LSTM-HouseA (ARAS)

LSTM model	Both person (EMR)	F1-score (R1)	F1-score (R2)
Baseline	54.03%	0.65	0.73
Single_oversample (2)	52.46%	0.64	0.72
Single_oversample (5)	50.52%	0.60	0.73
Multi_oversample (2)	53.25%	0.64	0.74
Multi_oversample (5)	51.50%	0.63	0.72
Single_undersample (0.25)	52.87%	0.65	0.73
Single_undersample (0.5)	53.15%	0.64	0.73
Multi_undersample (0.25)	53.39%	0.64	0.73
Multi_undersample (0.5)	52.26%	0.64	0.73
Cost_single (2)	53.03%	0.66	0.71
Cost_single (5)	51.14%	0.63	0.72
Cost_multi (2)	52.80%	0.64	0.73
Cost_multi (5)	52.85%	0.64	0.73

5.1 Results on frequent activities

In order to have a further comprehensive analysis of different class, imbalance approaches on multi-resident activity recognition datasets, we extended our experiments by selecting the first top-five activities of the dataset, and performs classifications using the same LSTM and BiLSTM as

TABLE 2: BiLSTM-HouseA (ARAS)

LSTM model	Both person (EMR)	F1-score (R1)	F1-score (R2)
Baseline	53.89%	0.65	0.74
Single_oversample (2)	52.76%	0.64	0.74
Single_oversample (5)	50.89%	0.62	0.72
Multi_oversample (2)	53.60%	0.65	0.74
Multi_oversample (5)	51.59%	0.63	0.72
Single_undersample (0.25)	53.96%	0.65	0.74
Single_undersample (0.5)	53.57%	0.65	0.73
Multi_undersample (0.25)	54.28%	0.65	0.74
Multi_undersample (0.5)	53.41%	0.65	0.72
Cost_single (2)	51.67%	0.63	0.72
Cost_single (5)	50.44%	0.64	0.68
Cost_multi (2)	54.06%	0.65	0.74
Cost_multi (5)	53.80%	0.66	0.74

TABLE 3: LSTM-HouseB (ARAS)

LSTM model	Both person (EMR)	F1-score (R1)	F1-score (R2)
Baseline	90.56%	0.94	0.93
Single_oversample (2)	90.83%	0.94	0.93
Single_oversample (5)	90.25%	0.93	0.93
Multi_oversample (2)	90.44%	0.94	0.93
Multi_oversample (5)	90.24%	0.93	0.93
Single_undersample (0.25)	90.58%	0.94	0.93
Single_undersample (0.5)	90.56%	0.93	0.93
Multi_undersample (0.25)	90.75%	0.94	0.93
Multi_undersample (0.5)	90.90%	0.94	0.93
Cost_single (2)	90.83%	0.93	0.93
Cost_single (5)	90.82%	0.94	0.93
Cost_multi (2)	90.64%	0.93	0.94
Cost_multi (5)	90.93%	0.94	0.93

TABLE 4: BiLSTM-HouseB (ARAS)

LSTM model	Both person (EMR)	F1-score (R1)	F1-score (R2)
Baseline	91.42%	0.94	0.94
Single_oversample (2)	90.82%	0.94	0.93
Single_oversample (5)	90.23%	0.94	0.92
Multi_oversample (2)	91.18%	0.95	0.93
Multi_oversample (5)	90.64%	0.94	0.93
Single_undersample (0.25)	90.73%	0.94	0.93
Single_undersample (0.5)	90.83%	0.94	0.93
Multi_undersample (0.25)	90.82%	0.94	0.93
Multi_undersample (0.5)	91.09%	0.94	0.93
Cost_single (2)	91.29%	0.94	0.93
Cost_single (5)	90.29%	0.94	0.93
Cost_multi (2)	91.15%	0.94	0.94
Cost_multi (5)	90.68%	0.93	0.93

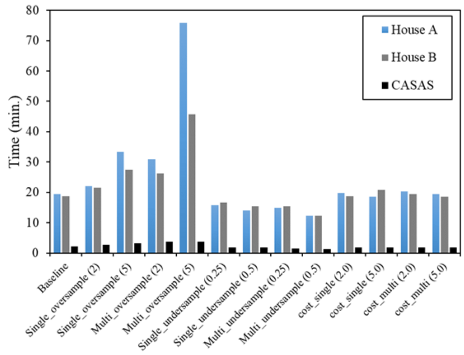
described above. As can be seen from Figure 5, Figure 6 and Figure 7 that even after oversampling and undersampling the data, the distribution is still imbalance, so we selected the top five activities in all the datasets to analyze model performance on frequent activities. The model configurations are exactly the same as previous experiments and the results of the experiments of ARAS are shown in Table 7-10. The model configurations in CASAS frequent activities dataset are also the same as previous CASAS experiments. Table 11-12 presents the results of class imbalance techniques on frequent activities of CASAS-Kyoto dataset.

TABLE 5: LSTM (CASAS-Kyoto)

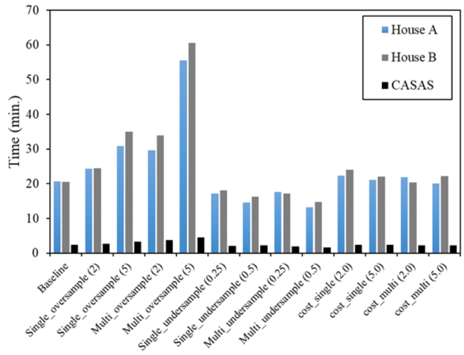
LSTM model	Both person (EMR)	F1-score (R1)	F1-score (R2)
Baseline	37.07%	0.55	0.54
Single_oversample (2)	35.44%	0.57	0.54
Single_oversample (5)	33.91%	0.56	0.52
Multi_oversample (2)	35.88%	0.57	0.54
Multi_oversample (5)	34.20%	0.56	0.52
Single_undersample (0.25)	37.01%	0.57	0.53
Single_undersample (0.5)	37.53%	0.58	0.55
Multi_undersample (0.25)	38.86%	0.58	0.54
Multi_undersample (0.5)	34.20%	0.58	0.51
Cost_single (2)	35.16%	0.54	0.53
Cost_single (5)	26.00%	0.54	0.48
Cost_multi (2)	35.62%	0.58	0.53
Cost_multi (5)	37.04%	0.58	0.54

TABLE 6: BiLSTM (CASAS-Kyoto)

BiLSTM model	Both person (EMR)	F1-score (R1)	F1-score (R2)
Baseline	37.64%	0.57	0.55
Single_oversample (2)	36.52%	0.58	0.54
Single_oversample (5)	34.89%	0.58	0.53
Multi_oversample (2)	36.78%	0.58	0.55
Multi_oversample (5)	34.08%	0.59	0.53
Single_undersample (0.25)	35.82%	0.58	0.53
Single_undersample (0.5)	37.85%	0.57	0.54
Multi_undersample (0.25)	37.04%	0.60	0.54
Multi_undersample (0.5)	35.82%	0.59	0.54
Cost_single (2)	33.10%	0.57	0.52
Cost_single (5)	24.32%	0.53	0.51
Cost_multi (2)	35.21%	0.58	0.53
Cost_multi (5)	32.11%	0.57	0.52



(a) LSTM model execution time



(b) BiLSTM model execution time

Fig. 11: Model execution time

TABLE 7: LSTM-House A (ARAS)

LSTM model	Both person (EMR)	Balanced accuracy (R1)	Balanced accuracy (R2)	F1-score (R1)	F1-score (R2)
Baseline	77.95%	**75.02**%/	60.13%	0.87	0.86
Single_oversample (2)	74.94%	73.63%	58.43%	0.84	0.85
Single_oversample (5)	75.68%	71.58%	61.47%	0.86	0.84
Multi_oversample (2)	78.09%	74.90%	60.27%	0.87	0.86
Multi_oversample (5)	75.29%	74.64%	57.05%	0.85	0.84
Single_undersample (0.25)	78.76%	73.07%	62.29%	0.86	0.86
Single_undersample (0.5)	75.86%	73.66%	60.83%	0.86	0.85
Multi_undersample (0.25)	77.13%	71.96%	61.94%	0.86	0.85
Multi_undersample (0.5)	76.21%	73.15%	61.96%	0.85	0.85
Cost_single (2)	75.58%	73.17%	62.97%	0.85	0.84
Cost_single (5)	73.68%	73.82%	57.86%	0.84	0.82
Cost_multi (2)	78.11%	73.27%	**64.27**%	0.87	0.87
Cost_multi (5)	76.62%	73.42%	62.67%	0.85	0.85

TABLE 9: LSTM-House B (ARAS)

LSTM model	Both person (EMR)	Balanced accuracy (R1)	Balanced accuracy (R2)	F1-score (R1)	F1-score (R2)
Baseline	96.73%	82.14%	88.07%	0.98	0.98
Single_oversample (2)	96.98%	83.48%	**91.80**%	0.98	0.98
Single_oversample (5)	96.46%	81.66%	87.94%	0.98	0.98
Multi_oversample (2)	96.55%	80.42%	88.43%	0.97	0.98
Multi_oversample (5)	96.41%	83.72%	85.03%	0.98	0.97
Single_undersample (0.25)	97.14%	77.90%	91.07%	0.98	0.98
Single_undersample (0.5)	97.07%	81.54%	84.37%	0.98	0.98
Multi_undersample (0.25)	96.75%	**83.66**%	90.37%	0.97	0.99
Multi_undersample (0.5)	96.97%	83.53%	86.51%	0.98	0.99
Cost_single (2)	97.23%	83.07%	91.02%	0.98	0.98
Cost_single (5)	96.51%	82.51%	81.22%	0.97	0.98
Cost_multi (2)	96.38%	80.51%	84.69%	0.98	0.98
Cost_multi (5)	96.49%	76.98%	86.75%	0.97	0.98

TABLE 8: BiLSTM-House A (ARAS)

BiLSTM model	Both person (EMR)	Balanced accuracy (R1)	Balanced accuracy (R2)	F1-score (R1)	F1-score (R2)
Baseline	78.58%	73.02%	64.71%	0.87	0.87
Single_oversample (2)	77.95%	72.51%	65.19%	0.87	0.88
Single_oversample (5)	75.64%	74.61%	59.92%	0.86	0.84
Multi_oversample (2)	79.09%	75.30%	65.31%	0.88	0.88
Multi_oversample (5)	75.49%	72.25%	62.72%	0.84	0.86
Single_undersample (0.25)	79.01%	74.46%	60.81%	0.88	0.87
Single_undersample (0.5)	79.35%	75.22%	64.12%	0.87	0.88
Multi_undersample (0.25)	77.37%	71.13%	65.19%	0.85	0.87
Multi_undersample (0.5)	78.05%	74.18%	64.19%	0.87	0.86
Cost_single (2)	76.48%	**75.88**%	62.08%	0.87	0.84
Cost_single (5)	74.23%	72.71%	62.47%	0.86	0.82
Cost_multi (2)	78.27%	74.02%	63.52%	0.88	0.88
Cost_multi (5)	77.01%	74.73%	**65.96**%	0.85	0.88

TABLE 10: BiLSTM-House B (ARAS)

BiLSTM model	Both person (EMR)	Balanced accuracy (R1)	Balanced accuracy (R2)	F1-score (R1)	F1-score (R2)
Baseline	96.88%	84.98%	**93.88**%	0.98	0.98
Single_oversample (2)	96.79%	84.21%	93.14%	0.98	0.98
Single_oversample (5)	97.06%	83.85%	87.47%	0.98	0.98
Multi_oversample (2)	96.83%	81.08%	90.61%	0.98	0.98
Multi_oversample (5)	96.45%	84.14%	83.19%	0.97	0.99
Single_undersample (0.25)	97.23%	84.56%	83.96%	0.98	0.99
Single_undersample (0.5)	97.31%	83.10%	92.67%	0.98	0.99
Multi_undersample (0.25)	97.85%	85.68%	90.76%	0.98	0.99
Multi_undersample (0.5)	96.95%	80.78%	83.73%	0.98	0.98
Cost_single (2)	97.35%	83.40%	90.82%	0.98	0.98
Cost_single (5)	96.91%	**86.67**%	81.27%	0.98	0.98
Cost_multi (2)	96.91%	80.21%	88.02%	0.97	0.99
Cost_multi (5)	96.89%	82.16%	92.60%	0.98	0.99

The EMR, balanced accuracy, and F1 score of both House A and House B of the ARAS dataset improved a lot in comparison to previous experiments when we took frequent activities of the datasets, which also makes the dataset quite balanced and thus improving the performance of LSTM and BiLSTM models. The results of EMR improved a lot in comparison to previous experiments but are similar in each dataset for all the approaches. In the balanced accuracy

TABLE 11: LSTM-(CASAS-Kyoto)

LSTM model	Both person (EMR)	Balanced accuracy (R1)	Balanced accuracy (R2)	F1-score (R1)	F1-score (R2)
Baseline	22.87%	27.01%	38.19%	0.36	0.41
Single_oversample (2)	19.49%	25.17%	25.17%	0.36	0.39
Single_oversample (5)	24.66%	30.30%	44.51%	0.34	0.41
Multi_oversample (2)	23.97%	27.74%	33.41%	0.42	0.32
Multi_oversample (5)	25.41%	30.04%	37.59%	0.35	0.44
Single_undersample (0.25)	20.45%	24.18%	30.81%	0.25	0.45
Single_undersample (0.5)	24.45%	27.71%	30.85%	0.30	0.40
Multi_undersample (0.25)	27.07%	27.38%	36.48%	0.36	0.46
Multi_undersample (0.5)	18.46%	30.10%	35.47%	0.29	0.44
Cost_single (2)	25.14%	**30.48%**	34.86%	0.36	0.32
Cost_single (5)	22.59%	29.63%	38.53%	0.33	0.41
Cost_multi (2)	22.11%	27.77%	**45.75%**	0.36	0.42
Cost_multi (5)	26.72%	26.71%	44.59%	0.38	0.41

TABLE 12: BiLSTM-(CASAS-Kyoto)

BiLSTM model	Both person (EMR)	Balanced accuracy (R1)	Balanced accuracy (R2)	F1-score (R1)	F1-score (R2)
Baseline	22.52%	25.23%	33.89%	0.32	0.41
Single_oversample (2)	26.65%	24.80%	46.07%	0.35	0.46
Single_oversample (5)	21.42%	20.54%	42.72%	0.27	0.43
Multi_oversample (2)	29.27%	31.47%	40.81%	0.40	0.40
Multi_oversample (5)	21.28%	25.55%	29.76%	0.33	0.37
Single_undersample (0.25)	21.28%	23.94%	36.21%	0.29	0.43
Single_undersample (0.5)	21.14%	20.71%	36.93%	0.34	0.36
Multi_undersample (0.25)	25.00%	**32.58%**	31.00%	0.38	0.39
Multi_undersample (0.5)	26.93%	32.21%	**47.48%**	0.36	0.46
Cost_single (2)	24.17%	29.72%	36.98%	0.28	0.44
Cost_single (5)	22.31%	26.21%	37.61%	0.33	0.39
Cost_multi (2)	28.24%	30.06%	37.65%	0.36	0.38
Cost_multi (5)	27.75%	30.03%	43.14%	0.36	0.46

results of frequent activities, again cost-sensitive approach performed better in most of the cases in comparison to oversampling and undersampling methods. There were few cases such as in House B, in R2 activities classification, oversampling approach in LSTM, and baseline model in BiLSTM network performed better than other approaches. However, the cost-sensitive approach performed equally in these cases, for example, the results of cost single (2) and single oversample (2) are almost equal in the LSTM model. Similarly, in the BiLSTM network of House B, the difference between baseline and cost-multi (5) is very less. In the CASAS-Kyoto dataset, the multi-undersampling method performed better in the BiLSTM network for both the residents. However, in per class F1-score results, the cost-sensitive method performed better in the classification of minority classes. Table 13, 14 and 15, presents per class F1-score results of cost-sensitive learning method for all the three datasets. As can been seen from Table 13 of House A, in R1 F1-score increases by 2%, 4%, and 2% in the LSTM model for minority classes such as Watching TV, Using Internet, and Talking on phone. Similarly, in the BiLSTM model of R1, F1-score of Using Internet class increases by 7% in cost-multi (2) in comparison to the baseline model of BiLSTM. In R2, the cost-multi (2) method of LSTM performed better in Watching TV, Napping, and Using Internet activities and in BiLSTM, an increase of 7% is observed in Napping activity. In the CASAS-Kyoto dataset, a cost-sensitive approach increases the F1-score of R1 and R2 minority classes such as Sweep kitchen floor, set dinning table for dinner, Playing checkers game, Retrieve dishes from cabinet and Other.

CASAS-Kyoto dataset shows improvement in class imbalance techniques (Table 11 and Table 12) in comparison to baseline model such as in LSTM model of CASAS, cost-sensitive method outperformed all the other methods and in BiLSTM model of CASAS, undersampling approach performed better, but the results of F1-score of cost-sensitive approach are almost similar to the undersampling method for each class. The micro average F1-score of both House A and House B improved a lot in frequent activity experiments, whereas in CASAS it did not show much improvement. It can be due to the curse of dimensionality in the SH datasets as not all sensors are relevant to the classification and high dimension deteriorates the performance of the classifier. Furthermore, it has been observed that the CASAS-Kyoto dataset showed a difference in the performance of the model with different class imbalance techniques, whereas in the ARAS dataset not much clear trend is observed. It can be attributed to the fact that the CASAS-Kyoto dataset is quite a balanced dataset whereas the ARAS dataset is highly imbalanced.

6 Conclusion

Multiple resident activity recognition is an important yet challenging research area that plays a crucial role in various applications such as smart environments, elderly care, assisted living, security, surveillance, and context-aware systems. In real-life applications, the frequency and duration of human activities are intrinsically imbalanced. In order to handle class imbalance problems in activity recognition systems, we applied both data level and algorithm level class imbalance techniques to deep learning networks for multi-class problems. The experiments are performed with separate activity labels and also with combined activity labels for both residents. In addition, sampling and cost-sensitive learning methods are evaluated at different sampling ratios and cost coefficients to explore model performance. The paper presents an extensive study on three SH datasets (ARAS and CASAS-Kyoto) with different imbalance techniques on LSTM and BiLSTM networks.

The results indicate that cost-sensitive learning improves the performance in balanced accuracy of both residents in comparison to sampling methods. We further extended our approach by performing experiments on frequent activities for all three datasets to explore performance and a similar trend was observed. The cost-sensitive learning approach showed improvements in the majority of the cases.

The proposed models are evaluated on two residents, but in the case of more than two residents (three or more residents), the model will perform training in a similar way as in the case of two residents. However, the performance of the model may be affected since it depends on the imbalanced class distribution, and also sensor readings would correspond to many residents, which becomes a complex problem. Generally, activity recognition of two residents is a more common approach than three or four residents.

To further improve the model performance on minority class, in our future work we will explore other deep learning models and hybrid approaches for handling class imbalance in the multi-resident setting. A further upcoming important aspect is in what is now called explainable AI [63], i.e. to focus on explainable and human interpretable models for the temporal SH datasets which will help in understanding the decision made by the network.

TABLE 13: F1-score results (per class) of House A

ID	Activities	LSTM					BiLSTM				
		Baseline	CS(2)	C(5)	CM(2)	CM(5)	Baseline	CS(2)	C(5)	CM(2)	CM(5)
	Going Out	0.86	0.80	0.80	0.80	0.83	0.84	0.88	0.86	0.92	0.84
	Sleeping	0.95	0.94	0.95	0.95	0.96	0.95	0.96	0.95	0.95	0.96
R1	Watching TV	0.74	0.72	0.68	0.73	0.76	0.74	0.73	0.73	0.76	0.75
	Using Internet	0.76	0.77	0.79	0.79	0.80	0.75	0.77	0.78	0.82	0.81
	Talking on phone	0.20	0.24	0.22	0.16	0.22	0.25	0.26	0.21	0.18	0.25
	Going Out	0.92	0.70	0.71	0.93	0.92	0.92	0.72	0.71	0.93	0.92
	Sleeping	0.91	0.77	0.78	0.91	0.92	0.94	0.78	0.77	0.92	0.94
R2	Watching TV	0.51	0.51	0.48	0.55	0.58	0.54	0.52	0.49	0.53	0.62
	Napping	0.09	0.07	0.17	0.13	0.11	0.11	0.17	0.28	0.17	0.18
	Using Internet	0.30	0.26	0.28	0.33	0.24	0.34	0.33	0.25	0.28	0.37

TABLE 14: F1-score results (per class) of House B

ID	Activities	LSTM					BiLSTM				
		Baseline	CS(2)	C(5)	CM(2)	CM(5)	Baseline	CS(2)	C(5)	CM(2)	CM(5)
	Other	0.38	0.33	0.32	0.38	0.32	0.49	0.37	0.50	0.43	0.31
	Going out	0.99	0.98	0.99	0.99	0.99	1.00	0.99	1.00	1.00	0.99
R1	Sleeping	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
	Watching TV	0.91	0.93	0.94	0.93	0.94	0.93	0.91	0.92	0.94	0.92
	Studying	0.78	0.78	0.76	0.82	0.78	0.83	0.82	0.82	0.80	0.77
	Going Out	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	Sleeping	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
R2	Watching TV	0.94	0.93	0.94	0.94	0.93	0.97	0.92	0.95	0.94	0.96
	Using Internet	0.61	0.52	0.59	0.54	0.47	0.59	0.64	0.56	0.53	0.53
	Reading Book	0.28	0.26	0.19	0.18	0.17	0.36	0.23	0.17	0.24	0.24

TABLE 15: F1-score results (per class) of CASAS-Kyoto

ID	Activities	LSTM					BiLSTM				
		Baseline	CS(2)	C(5)	CM(2)	CM(5)	Baseline	CS(2)	C(5)	CM(2)	CM(5)
	Fill medication dispenser	0.35	0.28	0.35	0.26	0.33	0.38	0.19	0.37	0.21	0.27
	Sweep kitchen floor	0.09	0.23	0.30	0.21	0.25	0.29	0.22	0.35	0.28	0.38
R1	Set dinning table for dinner	0.19	0.00	0.00	0.18	0.07	0.00	0.11	0.00	0.14	0.15
	Pack supplies in basket	0.12	0.03	0.03	0.09	0.12	0.12	0.18	0.07	0.07	0.08
	Other	0.35	0.33	0.30	0.41	0.42	0.40	0.47	0.40	0.43	0.46
	Hang up clothes	0.44	0.60	0.53	0.44	0.59	0.65	0.54	0.54	0.64	0.51
	Playing checkers game	0.17	0.16	0.15	0.39	0.38	0.30	0.30	0.20	0.30	0.25
R2	Retrieve dishes from cabinet	0.12	0.09	0.03	0.00	0.11	0.04	0.35	0.19	0.19	0.42
	Pack food & bring to front door	0.17	0.20	0.16	0.15	0.04	0.19	0.21	0.45	0.22	0.24
	Other	0.46	0.43	0.45	0.42	0.45	0.43	0.30	0.46	0.35	0.52

REFERENCES

[1] S. Ranasinghe, F. Al Machot, and H. C. Mayr, "A review on applications of activity recognition systems with regard to performance and evaluation," *International Journal of Distributed Sensor Networks*, vol. 12, no. 8, p. 1550147716665520, 2016.

[2] D. Singh, I. Psychoula, E. Merdivan, J. Kropf, S. Hanke, E. Sandner, L. Chen, and A. Holzinger, "Privacy-enabled smart home framework with voice assistant," in *Smart Assisted Living*. Springer, 2020, pp. 321–339.

[3] D. Singh, E. Merdivan, I. Psychoula, J. Kropf, S. Hanke, M. Geist, and A. Holzinger, "Human activity recognition using recurrent neural networks," in *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, 2017, pp. 267–274.

[4] E. Hoque and J. Stankovic, "Aalo: Activity recognition in smart homes using active learning in the presence of overlapped activities," in *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. IEEE, 2012, pp. 139–146.

[5] L. Chen, C. D. Nugent, and H. Wang, "A knowledge-driven approach to activity recognition in smart homes," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 6, pp. 961–974, 2011.

[6] J. Rafferty, C. D. Nugent, J. Liu, and L. Chen, "From activity recognition to intention recognition for assisted living within smart homes," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 3, pp. 368–379, 2017.

[7] A. Benmansour, A. Bouchachia, and M. Feham, "Multioccupant activity recognition in pervasive smart home environments," *ACM Computing Surveys (CSUR)*, vol. 48, no. 3, p. 34, 2016.

[8] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-smote: a new over-sampling method in imbalanced data sets learning," in *International conference on intelligent computing*. Springer, 2005, pp. 878–887.

[9] S.-J. Yen and Y.-S. Lee, "Cluster-based under-sampling approaches for imbalanced data distributions," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5718–5727, 2009.

[10] C. X. Ling and V. S. Sheng, "Cost-sensitive learning and the class imbalance problem," pp. 231–235, 2008.

[11] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 559–563, 2017.

[12] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," in *European conference on machine learning*. Springer, 2004, pp. 39–50.

[13] J. Błaszczyński and J. Stefanowski, "Neighbourhood sampling in bagging for imbalanced data," *Neurocomputing*, vol. 150, pp. 529–542, 2015.

[14] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, 2011.

[15] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[16] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

[17] A. Fernández, V. López, M. Galar, M. J. Del Jesus, and F. Herrera, "Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches," *Knowledge-based systems*, vol. 42, pp. 97–110, 2013.

[18] N. Thai-Nghe, Z. Gantner, and L. Schmidt-Thieme, "Cost-sensitive learning methods for imbalanced data," in *The 2010 International joint conference on neural networks (IJCNN)*. IEEE, 2010, pp. 1–8.

[19] D. Singh, E. Merdivan, S. Hanke, J. Kropf, M. Geist, and A. Holzinger, "Convolutional and recurrent neural networks for activity recognition in smart environment," in *Towards integrative machine learning and knowledge extraction*. Springer, 2017, pp. 194–205.

[20] N. Nguyen, S. Venkatesh, and H. Bui, "Recognising behaviours of multiple people with hierarchical probabilistic model and statistical data association," in *BMVC 2006: Proceedings of the 17th British Machine Vision Conference*. British Machine Vision Association, 2006, pp. 1239–1248.

[21] Y. Du, F. Chen, and W. Xu, "Human interaction representation and recognition through motion decomposition," *IEEE Signal Processing Letters*, vol. 14, no. 12, pp. 952–955, 2007.

[22] P. Natarajan and R. Nevatia, "Coupled hidden semi markov models for activity recognition," in *2007 IEEE Workshop on Motion and Video Computing (WMVC'07)*. IEEE, 2007, pp. 10–10.

[23] A. S. Crandall and D. J. Cook, "Using a hidden markov model for resident identification," in *2010 Sixth International Conference on Intelligent Environments*. IEEE, 2010, pp. 74–79.

[24] G. Singla, D. J. Cook, and M. Schmitter-Edgecombe, "Recognizing independent and joint activities among multiple residents in smart environments," *Journal of ambient intelligence and humanized computing*, vol. 1, no. 1, pp. 57–63, 2010.

[25] L. Wang, T. Gu, X. Tao, H. Chen, and J. Lu, "Recognizing multi-user activities using wearable sensors in a smart home," *Pervasive and Mobile Computing*, vol. 7, no. 3, pp. 287–298, 2011.

[26] D. Singh, J. Kropf, S. Hanke, and A. Holzinger, "Ambient assisted living technologies from the perspectives of older people and professionals," in *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, 2017, pp. 255–266.

[27] D. Singh, I. Psychoula, J. Kropf, S. Hanke, and A. Holzinger, "Users' perceptions and attitudes towards smart home technologies," in *International Conference on Smart Homes and Health Telematics*. Springer, 2018, pp. 203–214.

[28] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Information Fusion*, vol. 35, pp. 68–80, 2017.

[29] Z. Qin, Y. Zhang, S. Meng, Z. Qin, and K.-K. R. Choo, "Imaging and fusing time series for wearable sensor-based human activity recognition," *Information Fusion*, vol. 53, pp. 80–87, 2020.

[30] M. S. Zainudin, M. N. Sulaiman, N. Mustapha, and T. Perumal, "Activity recognition based on accelerometer sensor using combinational classifiers," in *2015 Ieee Conference on Open Systems (Icos)*. IEEE, 2015, pp. 68–73.

[31] C.-H. Lu and L.-C. Fu, "Robust location-aware activity recognition using wireless sensor network in an attentive home," *IEEE Transactions on Automation Science and Engineering*, vol. 6, no. 4, pp. 598–609, 2009.

[32] A. S. Crandall and D. J. Cook, "Resident and caregiver: Handling multiple people in a smart care facility." in *AAAI Fall Symposium: AI in Eldercare: New Solutions to Old Problems*, 2008, pp. 39–47.

[33] K.-C. Hsu, Y.-T. Chiang, G.-Y. Lin, C.-H. Lu, J. Y.-J. Hsu, and L.-C. Fu, "Strategies for inference mechanism of conditional random fields for multiple-resident activity recognition in a smart home," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, 2010, pp. 417–426.

[34] D. Cook, M. Schmitter-Edgecombe, A. Crandall, C. Sanders, and B. Thomas, "Collecting and disseminating smart home sensor data in the casas project," in *Proceedings of the CHI workshop on developing shared home behavior datasets to advance HCI and ubiquitous computing research*, 2009, pp. 1–7.

[35] R. Chen and Y. Tong, "A two-stage method for solving multi-resident activity recognition in smart environments," *Entropy*, vol. 16, no. 4, pp. 2184–2203, 2014.

[36] D. Liciotti, M. Bernardini, L. Romeo, and E. Frontoni, "A sequential deep learning application for recognising human activities in smart homes," *Neurocomputing*, 2019.

[37] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang, "Deep residual bidir-lstm for human activity recognition using wearable sensors," *Mathematical Problems in Engineering*, vol. 2018, 2018.

[38] X. Li, Y. Zhang, J. Zhang, S. Chen, I. Marsic, R. A. Farneth, and R. S. Burd, "Concurrent activity recognition with multimodal cnn-lstm structure," *arXiv preprint arXiv:1702.01638*, 2017.

[39] R. A. Hamad, L. Yang, W. L. Woo, and B. Wei, "Joint learning of temporal models to handle imbalanced data for human activity recognition," *Applied Sciences*, vol. 10, no. 15, p. 5293, 2020.

[40] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221–232, 2016.

[41] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[42] A. Fernández, S. García, and F. Herrera, "Addressing the classification with imbalanced data: open problems and new challenges on class distribution," in *International conference on hybrid artificial intelligence systems*. Springer, 2011, pp. 1–10.

[43] I. Mani and I. Zhang, "knn approach to unbalanced data distributions: a case study involving information extraction," in *Proceedings of workshop on learning from imbalanced datasets*, vol. 126, 2003.

[44] Z.-H. Zhou and X.-Y. Liu, "On multi-class cost-sensitive learning," *Computational Intelligence*, vol. 26, no. 3, pp. 232–257, 2010.

[45] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, vol. 16, pp. 3–17, 2014.

[46] S. Wang, Z. Li, W. Chao, and Q. Cao, "Applying adaptive over-sampling technique based on data density and cost-sensitive svm to imbalanced learning," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2012, pp. 1–8.

[47] M. Lin, K. Tang, and X. Yao, "Dynamic sampling approach to training neural networks for multiclass imbalance classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 4, pp. 647–660, 2013.

[48] B. Krawczyk and M. Woźniak, "Cost-sensitive neural network with roc-based moving threshold for imbalanced classification," in *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, 2015, pp. 45–52.

[49] N. V. Chawla, D. A. Cieslak, L. O. Hall, and A. Joshi, "Automatically countering imbalance and its empirical relationship to cost," *Data Mining and Knowledge Discovery*, vol. 17, no. 2, pp. 225–252, 2008.

[50] K. M. Chathuramali and R. Rodrigo, "Faster human activity recognition with svm," in *International Conference on Advances in ICT for Emerging Regions (ICTer2012)*. IEEE, 2012, pp. 197–203.

[51] C. Huang, Y. Li, C. Change Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5375–5384.

[52] S. Ando and C. Y. Huang, "Deep over-sampling framework for classifying imbalanced data," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2017, pp. 770–785.

[53] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and P. J. Kennedy, "Training deep neural networks on imbalanced data sets," in *2016 international joint conference on neural networks (IJCNN)*. IEEE, 2016, pp. 4368–4374.

[54] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, "Cost-sensitive learning of deep feature representations from imbalanced data," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 8, pp. 3573–3587, 2017.

[55] E. Lin, Q. Chen, and X. Qi, "Deep reinforcement learning for imbalanced classification," *Applied Intelligence*, pp. 1–15, 2020.

[56] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, p. 27, 2019.
[57] S. Guo, Y. Liu, R. Chen, X. Sun, and X. Wang, "Improved smote algorithm to deal with imbalanced activity classes in smart homes," *Neural Processing Letters*, vol. 50, no. 2, pp. 1503–1526, 2019.
[58] M. Oussalah, A. Hessami, B. M. Abidine, B. Fergani, and L. Fergani, "A new classification strategy for human activity recognition using cost sensitive support vector machines for imbalanced data," *Kybernetes*, 2014.
[59] H. Alemdar, H. Ertan, O. D. Incel, and C. Ersoy, "Aras human activity datasets in multiple homes with multiple residents," in *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. IEEE, 2013, pp. 232–235.
[60] D. Cook, *Center of advanced studies in adaptive system(CASAS)*, 2009 (accessed July 20, 2020). [Online]. Available: http://http://casas.wsu.edu/datasets/
[61] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[62] C. Olah, "Understanding lstm networks," 2015, accessed: 2020-11-10. [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/
[63] R. Goebel, A. Chander, K. Holzinger, F. Lecue, Z. Akata, S. Stumpf, P. Kieseberg, and A. Holzinger, "Explainable ai: the new 42?" in *Springer Lecture Notes in Computer Science LNCS 11015*. Cham: Springer, 2018, pp. 295–303.

Deepika Singh is a Computer Science PhD student at Technical University Graz and researcher at Human Centered AI Lab (Holzinger Group) at Medical University Graz, Austria. She received her Masters of Technology in Computer Science from Banasthali Vidhyapith, India in 2014 and Bachelors of Technology in Computer Science from Rajasthan Technical University, India in 2011. Her research interests include health informatics, machine learning and deep learning.

Erinc Merdivan completed his PhD in Computer Science(CS) at CentraleSupelec in 2019. He received his Master of Science in Computer Science and Engineering from Sabanci University, Istanbul in 2013 and the Bachelor of Science in Electronics Engineering from Sabanci University, Istanbul in 2011. His research interests include deep learning, NLP and deep reinforcement learning.

Johannes Kropf holds a PhD in technical mathematics and is a researcher at AIT since 2008. He works as a scientist and project manager in the Center for Health and Environment and is involved in various national and international R&D projects. His research work focuses on user behaviour modelling, data analytics and persuasive computing in ambient intelligent environments, especially in the Active and Assisted Living (AAL) domain.

Andreas Holzinger is Visiting Professor for explainable AI at the Alberta Machine Intelligence Institute of the University of Alberta, Canada since 2019 and head of the Human-Centered AI Lab at the Medical University Graz, Austria. He received his PhD in cognitive science from Graz University in 1998 and his second PhD in computer science from Graz University of Technology in 2003. He is IEEE Member since 2000.

CHAPTER 6

Privacy Preservation in Smart Home using Deep Learning

6.1 Introduction

There are numerous definitions and methodologies regarding privacy in machine learning ranging from strict to flexible. In general, strict notions of privacy are preferred as they are capable to confront strong adversaries. Although relaxed notions of privacy can be helpful as they reduce perturbations to data and allow more accurate analysis in machine learning methods. Since these methods require a large amount of data to provide accurate predictions and classifications, more individuals data are collected nowadays by IoT devices. And it is becoming more challenging to protect the identity of the individual reliably especially when multiple stakeholders and services need access to the user data. In such scenarios, relaxed methods of privacy can be beneficial where certain parts of data should not be communicated, but other parts of data are necessary for analysis. In this chapter, we present a method to address the aforementioned challenges in the ambient assisted living paradigm by developing and evaluating a novel encoder-decoder mechanism for data anonymization. This mechanism can be deployed in the smart home domain to enable users to choose which data and at what granularity will be shared with other services and stakeholders. The first publication presents an LSTM multiple encoder and multiple decoder model which is able to learn privacy operations such as disclosure, generalization, and deletion of data and thus can generate different data views depending on the access level of the end-user and the required access information. Furthermore, we extended our previous work in the second publication by investigating the single encoder and multiple decoder approach for privacy preservation. The experiments were conducted on a synthetic dataset modeled on the type of data that are typically collected in ambient assisted living environments to create a combination of personal, medical, and Internet of Things sensor data. Results show that the proposed mechanism learns privacy operations such as disclosure, deletion, and generalization and can perform encoding and decoding of the data with good recovery.

The content of this chapter is based on the publication:

Psychoula, I., Merdivan, E., **Singh, D.**, Chen, L., Chen, F., Hanke, S., Kropf, J., Holzinger, A. and Geist, M., 2018, March. A deep learning approach for privacy preservation in assisted living. In 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) (pp. 710-715). IEEE.

Contribution: Deepika Singh, Ismini Pyschoula and Erinc Merdivan worked together in designing and implementation of experiments. Data generation was conducted by Ismini Psychoula. The manuscript was written by both Ismini Psychoula and Deepika Singh. Ismini Psychoula presented the paper at PERCOM 2018, Athens, Greece. The other authors contributed to the scientific discussion and revision of the manuscript.

Psychoula, I., **Singh, D.**, Merdivan, E., Chen, L, 2021, January. Privacy Preservation with Autoencoder based De-Identification and Differential Privacy. Submitted

Contribution: Deepika Singh, Ismini Pyschoula and Erinc Merdivan worked together in designing the experiments. The first part of the experiments were performed by Deepika Singh and Erinc Merdivan. And, the experiments on Differential privacy was performed by Ismini Psychoula. The manuscript was mainly written by Ismini Psychoula and Deepika Singh partially contributed in some sections of the manuscript. The other authors contributed to the scientific discussion and revision of the manuscript.

6.2 Publication VIII: A deep learning approach for privacy preservation in assisted living

A Deep Learning Approach for Privacy Preservation in Assisted Living

Ismini Psychoula*, Erinc Merdivan†, Deepika Singh† §, Liming Chen*, Feng Chen*,
Sten Hanke†, Johannes Kropf†, Andreas Holzinger§, Matthieu Geist‡
*School of Computer Science and Informatics, De Montfort University, Leicester, UK
Email: {ismini.psychoula, liming.chen}@dmu.ac.uk
†Center for Health & Bioresources, AIT Austrian Institute of Technology, Wiener Neustadt, Austria
Email: {erinc.merdivan, deepika.singh, sten.hanke, johannes.kropf}@ait.ac.at
§Holzinger Group, Institute for Medical Informatics/Statistics, Medical University Graz, A-8036 Graz, Austria
Email: a.holzinger@hci-kdd.org
‡Universit de Lorraine & CNRS, LIEC, UMR 7360, Metz, F-57070 France
Email: matthieu.geist@univ-lorraine.fr

***Abstract*—In the era of Internet of Things (IoT) technologies the potential for privacy invasion is becoming a major concern especially in regards to healthcare data and Ambient Assisted Living (AAL) environments. Systems that offer AAL technologies make extensive use of personal data in order to provide services that are context-aware and personalized. This makes privacy preservation a very important issue especially since the users are not always aware of the privacy risks they could face. A lot of progress has been made in the deep learning field, however, there has been lack of research on privacy preservation of sensitive personal data with the use of deep learning. In this paper we focus on a Long Short Term Memory (LSTM) Encoder-Decoder, which is a principal component of deep learning, and propose a new encoding technique that allows the creation of different AAL data views, depending on the access level of the end user and the information they require access to. The efficiency and effectiveness of the proposed method are demonstrated with experiments on a simulated AAL dataset. Qualitatively, we show that the proposed model learns privacy operations such as disclosure, deletion and generalization and can perform encoding and decoding of the data with almost perfect recovery.**

I. Introduction

The dramatic demographic change in most western countries will increase the need for development of new Ambient Intelligence (AmI) technologies making use of Artificial Intelligence (AI) and machine learning (ML) [28]. The new EU Data Protection regulations applying from 2018 onwards [2] will make privacy aware machine learning necessary [23]. Consequently, issues of privacy, security, safety and data protection move more and more into the focus of AI and ML, thereby fostering an integrated ML approach [17], which emphasizes the importance of the human-in-the-loop. Currently, major threats to privacy come from personal data aggregation and the increasing power of data mining and pattern recognition techniques, as well as from healthcare data sharing and analysis. As the number of information sources increases the potential to combine these sources, profile the users and learn sensitive information about them also increases, which makes it a great threat to individual privacy. This an important issue especially in the field of AAL where the users are not always aware of the privacy risks they might face.

To address the threats mentioned previously we propose a model for the encoding and sharing of combined healthcare and AAL data. The model aims to achieve the privacy of input data before they are distributed to various stakeholders. To protect the privacy a Long Short-Term Memory (LSTM) encoder-decoder system is designed that allows the creation of different data views to correspond to the access level of the receiver.

The remainder of the paper is organized as follows. Section 2 presents an overview of existing privacy techniques and related work for deep learning in privacy. Section 3 introduces the proposed privacy model. Section 4 describes the case study and explains the performed experiments while Section 5 presents the results and the performance of the algorithm. Lastly, Section 6 includes the conclusion and suggestions for future work.

II. Related Work

This section is divided into three parts. The first part gives an overview of privacy definitions and identifiers. The second describes previous work done in regards to privacy preserving techniques while the third part gives an introduction to deep learning and overview of existing work in privacy protection with the use of deep learning techniques.

A. Private Data

The new EU General Data Protection Regulation (GDPR) [2] defines personal data as “any information relating to an identified or identifiable natural person” and specifically acknowledges that this includes both ‘direct’ and ‘indirect’ identification. The identification can be by means of “an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity”. While so far there is not one privacy definition yet that is able to encompass all the different aspects of privacy, there are guidelines that list the possible identifiers that could be used to identify a person from a group [3] [1]:

1) Names, Geographical subdivisions smaller than a state, Dates (other than year)
2) Phone & Fax Numbers
3) Electronic mail addresses
4) Social Security, Medical Record & Health plan beneficiary numbers
5) Account & Certificate/license numbers
6) Vehicle identifiers and serial numbers (including license plate numbers)
7) Device identifiers and serial numbers
8) Web Uniform Resource Locators (URLs) & Internet Protocol (IP) address numbers
9) Biometric identifiers, including finger, retinal and voice prints
10) Full face photographic images and any comparable images
11) Any other unique identifying number, characteristic or code

However, the above list of identifiers is not exhaustive, as technology advances more potential identifiers could emerge.

B. Privacy Preservation with Anonymization Methods

Privacy enhancing technologies protect the users' privacy based on technology, and can offer additional levels of protection than just relying on laws and policies. In order to address the privacy concerns of the users, several approaches have been proposed by the research community. These approaches include information manipulation, privacy and context awareness, access control and data anonymization. In the sections below the anonymization are further analyzed since they are the methods most commonly used for privacy preservation.

1) k-Anonymity: A very well-known method to anonymize data before releasing them is k-anonymity [29]. In a k-anonymized dataset, each record is indistinguishable from at least $k-1$ other records in regards to specific identifying attributes [31]. k-anonymity is achieved by suppressing (deleting an attribute value from the data and replacing it with a random value that matches any possible attribute value) or generalizing the attributes in the data, which means that an attribute is replaced with a less specific but semantically consistent value [29]. The utility and privacy of the data are connected. There is no way so far that can increase the data privacy without also decreasing the data utility [24]. The objective in these problems is to maximize utility by minimizing the amount of generalization and suppression. Achieving k -anonymity by generalization with this objective as a constraint is a Non-deterministic Polynomial-time hard (NP hard) problem which cannot be solved fully automatically [25] [18]. In most cases k -anonymity is able to prevent identity disclosure so that a record in a k-anonymized data set cannot be connected again to the corresponding record in the original data set. But in some cases, it may fail to protect against attribute disclosure.

2) l-diversity: This method was developed to address the weaknesses of k-anonymity, which as shown, does not guarantee privacy against adversaries that use background knowledge or in cases where data are lacking diversity. For l-diversity, the anonymization conditions are satisfied if, for each group of records sharing a combination of key attributes, there are at least l -"well-represented" values for each confidential attribute [22]. The disadvantage of this method is that it depends on the range of the sensitive attributes. If l-diversity is to be applied to a sensitive attribute that does not have many different values, artificial data will have to be inserted. The use of artificial data will improve the privacy but may result in problems with the analysis thus ruining the utility of the data. Also, this method is vulnerable to skewness and similarity attack so it cannot always prevent attribute disclosure.

3) t-closeness: As shown l-diversity might not always be sufficient in preventing attribute disclosure. Since it does not account for the semantic closeness of the sensitive values. A new method named t- closeness was proposed in [21] to address these problems. This method requires the distribution of the sensitive attributes in an equivalent class to be close to the distribution of the attribute in the overall table, which in turn means that the distance between the two distributions should be no more than a specified threshold t. While the authors in [21] describe ways to check t-closeness (using several distances between distributions), no computational procedure to enforce this property is given [10]. The authors proposing the t-closing method [21] mention that t-closeness limits the amount of useful information that is released. The only way to increase the utility of the data is to increase the threshold t, which in turn decreases the privacy protection.

As seen from the overview of the strengths and weaknesses of each technique k-anonymity and the other anonymization methods are not always successful in guarantying that no information is leaked while ensuring usable data levels. While the methods of k-anonymity and l-diversity do not always accomplish complete privacy, the method of t-closeness provides it. But sometimes it is at the expense of the correlations between confidential attributes and key attributes. Also, the computational method for a specific dataset to be anonymized is an additional problem of this method. The papers defining k-anonymity and l-diversity propose approaches based on generalization and suppression which a lot of times can cause numerical attributes to become categorical. In the case of t-closeness, a computational procedure to reach it is not described. Thus, some issues in this field are still open, both at a conceptual and computational level, which can be improved by defining better properties and by creating more effective methods.

C. Privacy Preservation with Deep Learning

Deep learning is a promising area of machine learning research with significant success in recent years. So far the applications of deep learning are being used in various systems such as image and speech recognition, data analysis, social media, bioinformatics, medicine, and healthcare. Usually, deep learning architectures are constructed as multi-layer neural networks. There are several different neural network architectures, such as the Recurrent Neural Network (RNN) [13], the feed-forward neural network [4] and the Deep Belief Network (DBN) [14]. Deep learning has the ability to transform original data into a higher level with more abstract expressions. That means that high-dimensional original data can be converted to low-dimensional data by training multiple neural networks on how to reconstruct the high-dimensional input data.

However, the existing literature on privacy protection mostly focuses on traditional privacy preserving methods, as described in the previous section, and not on deep learning. Differential privacy proposed by Dwork [11] is one of the few approaches of privacy protection that makes use of machine learning methods. Applications of Differential Privacy include boosting [12], principal component analysis [7], linear and logistic regression [5], [32] support vector machines [26], risk minimization [6], [30] and continuous data processing [27]. However, the most relevant work to this paper is that of Dai et al. [9] in which they used an Encoder-Decoder system to protect private information in videos by extracting the privacy region and scrambling it while encoding. The system allows the users to fully restore the original video only if they have a legitimate key, otherwise, they can only see the non private regions in the video.

III. LSTM Encoder-Decoder Model

Long Short Term Memory networks are a special kind of RNN, capable of learning long-term dependencies [16]. A basic sequence-to-sequence model, as introduced in [8], consists of two recurrent neural networks (RNNs). The first is an encoder that processes the input and the second a decoder that generates the output. The recurrently connected blocks in LSTM layers are known as memory blocks. Each block contains one or more memory cells which are composed of three units: an input gate, a forget gate and an output gate. These gates modulate the interactions between the memory cell and the environment. Figure 1 shows the single cell of LSTM memory block. LSTM can assist in error minimization because the error can be back-propagated through time and layers. By maintaining a more constant error, the recurrent network can continue to learn over many time steps, and thus be able to link causes and effects.

The model that was developed in this work uses the multi-layered Long Short-Term Memory (LSTM) encoder to map the input sequence to a vector of a fixed dimensionality, and then another LSTM is used to decode the target sequence from the vector. The encoder network is the part of the network that takes the input sequence and maps it to an encoded representation of the sequence. The encoded representation is then used by the decoder network to generate an output sequence. This makes the framework have a lock and key analogy, where only someone with the correct key (decoder) will be able to access the resources behind the lock (encoder). The multiple hidden layers of neural networks have characteristics that enable this kind of learning [15], along with the mapping characteristic of the encoder-decoder models which are able to create corresponding pairs could make them appropriate for privacy preserving frameworks.

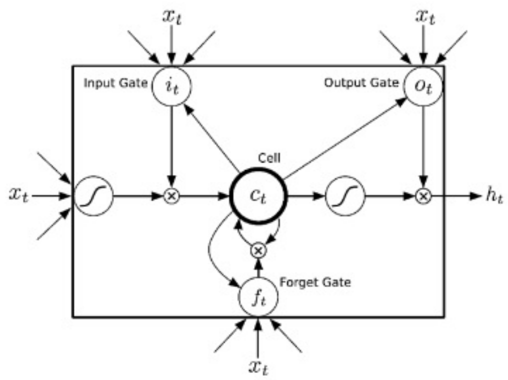


Fig. 1: LSTM single cell image [33].

IV. Experiment Design

The experiments were conducted using as basis the LSTM Encoder-Decoder model introduced in the previous section. In the following sections we describe the evaluation use case scenario and details of the simulated dataset used for the study followed by the modeling and training methodology.

A. *Use case*

John is 80 years old and lives in an ambient assisted living environment. He is widowed and was recently diagnosed with Alzheimer's disease. Currently, he lives alone but he likes to stay in touch with his family and friends. The AAL environment he lives in gives him independence and allows him to control his home automation system, for example, he can remotely open and close windows/doors, control the lighting, heating, and the alarm system. Also, it allows monitoring of his vital signs and offers him reminders about medication and appointments. The sensors deployed in the home send the collected information to the cloud offering access to, family members, care givers, doctors, and researchers.

In this use case scenario, four different views of the data are being created (Figure 2) depending on the access level of the receiver and the preferences of the user (Table I). The user in this scenario has a very close relationship with his family and trusts [34], [19] his caregiver so he has selected almost all his information to be accessible to them, especially because he feels safer knowing they will be able to help him in case of emergency. With regards to doctors, he has allowed only some basic personal and medical information to be visible to them so a different view is created for them. And finally for research purposes, the view that is created does not show any explicit personal data and most of the other sensitive attributes are generalized.

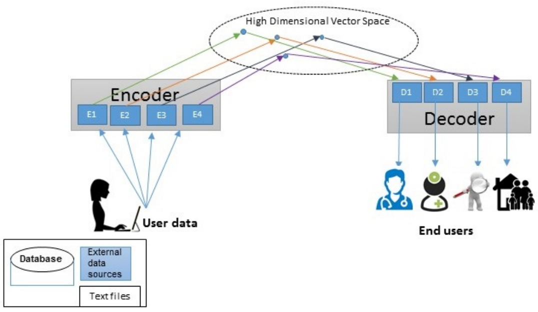


Fig. 2: Conceptual System Architecture

B. Dataset

For the purposes of this work data related to AAL were simulated. The type of data that were selected are divided into three categories: personal, medical and smart home sensor attributes. Each entry of the simulated data includes these three kinds of attributes. Personal attributes are those that can explicitly identify a person such as name, address and phone number. The second includes attributes that could potentially identify a person such as gender and birth date. Lastly, the third category includes sensitive medical information and sensor data, like blood pressure, medical history and presence sensors (Table I).

TABLE I: Simulated Data and Access to Information

Attribute	Family Member	Doctor	Caregiver	Researcher
Name	F	F	F	D
Age	F	F	G	G
Gender	F	F	F	F
Height	F	F	G	G
Weight	F	F	G	G
Address	F	G	F	G
Phone Number	F	F	F	D
Occupation	F	G	G	G
Marital Status	F	G	G	G
Timestamp	F	F	F	F
Blood Pressure	G	F	F	G
Glucose level	G	F	F	G
Disease	F	F	F	G
Wearable Pedometer	F	F	F	F
Presence Sensor	F	D	F	F
Temperature Sensor	F	F	F	G
Light Sensor	F	D	D	F
Window Sensor	F	D	F	D
External Door Sensor	F	D	F	D
Energy Consumption	G	D	D	G

Abbreviations F: fully disclosed, G: generalized, D: deleted

The simulation of the data was based on real world data collected from AAL environments and it included personal information, health care data as well as smart home sensor data. The data were simulated for 10000 users, which each user having 100 entries.

C. Model Configuration

As described previously, the proposed model makes use of the LSTM neural network architecture that learns to encode a variable-length input sequence into a fixed-length vector representation and to decode a given fixed-length vector representation back into a variable-length sequence. On this neural network three types of operations are applied to the encoder input by the decoder. These three operations are: 1) Disclosure which means keeping the data as it is 2) Deletion by removing the data or 3) Generalization which means replacing the value with a less specific but semantically consistent value. So the data from each entry can be fully disclosed to the receiver, generalized or deleted. Each value for a given attribute has different range aligned with real life values. Four separate views are created for different receivers, family member, doctor, caregiver, and researcher. Each receiver has a different decoder output due to their privacy clearances on patient information (Table I).

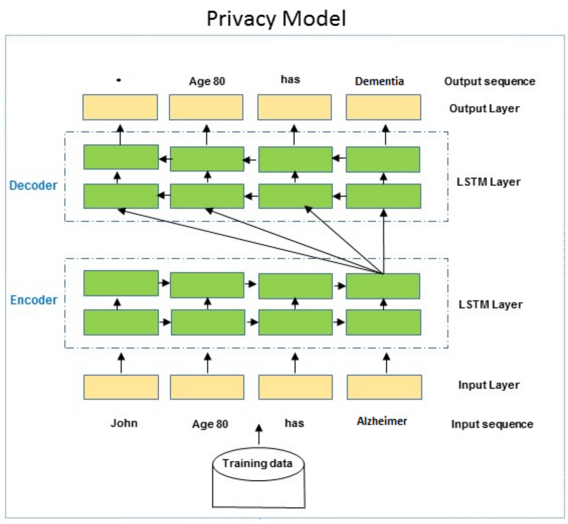


Fig. 3: Illustration of the proposed LSTM Encoder-Decoder Privacy Model

In Figure 3 the overall functionality of the encoder and decoder LSTM layers of the model are depicted. While in Figure 4 we show in more detail how the privacy operations work when the model reads an input sentence such as “John has Alzheimer” which will produce “* has Dementia” as the output sentence. In this instance the operation of deletion is applied on the Name attribute so ‘John’ is transformed to ‘*’ and the operation of generalization is applied on the Disease attribute which changes ‘Alzheimer’ to ‘Dementia’. The model stops making predictions after outputting the end-of-string token ‘eos’.

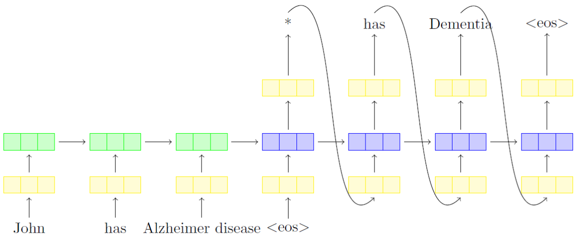


Fig. 4: Example of the model's Privacy Operations

D. Model Training and Testing

For the experiments the models were trained with 800.000 data entries and tested with 200.000 unseen entries. Each user entry consisted of different attribute as shown in Table I and their corresponding values, each entry had 160 characters at most. Different attributes are separated with '|' in order to easily distinguish between different attributes in an entry. A 40 character set was used as dictionary. Each sequence is maximum 160 characters long and ends with special token 'eos'. In order to handle different sequence lengths, we zero padded each entry to the maximum number of characters which in our case is 160. The encoder and decoder comprise an LSTM network which has 256 hidden units and is trained with Adam [20] with a learning rate of 0.0004.

V. Results

We qualitatively analyzed the trained model's results by comparing the decoded outputs with those of the expected model output after the privacy operations. The qualitative analysis shows that the LSTM Encoder Decoder is very good at learning the privacy operations of disclosure, generalization and deletion as well as at capturing the specified preferences in the access to information table. An example of the results of the model can be seen in the following figures.

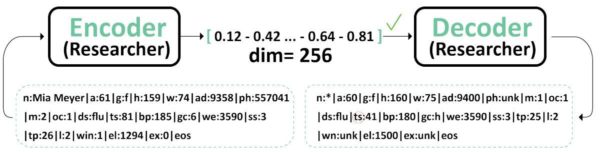


Fig. 5: Researcher view (With matching Encoder and Decoder)

In Figure 5, it can be seen that with the use of the right encoder-decoder mechanism the user's information can be transferred almost perfectly to the researcher with the appropriate privacy rules applied for the researcher's access level. And since only the encoded vector is shared it is not possible to get user information without the right decoder.

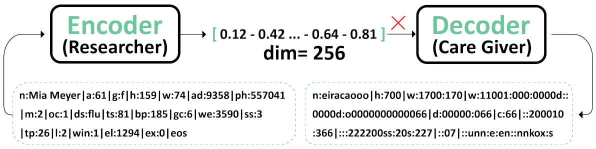


Fig. 6: Care Giver view (With not matching Encoder and Decoder)

Figure 6, showcases one of the most beneficial parts of our model. In this case, we explore the possibility that a Caregiver tries to access the data meant for the Researcher. Because the encoded vector of user information was shared, it would not be possible to decode it unless the right decoder was used. If the encoder and decoder do not match it is not possible to decode and access the user's private information. If the end receiver does not have the correct decoder it is almost impossible to find the right decoder weights since it is very high dimensional floating point vector.

The model has on average 1.5 character error per entry in testing. One or two characters error given the length of 160 characters per entry is very close to perfect recovery. Different decoders are trained for each view and all had the same very close error (1.5 char/entry) during testing. Decoders are shown to be capable of deleting, generalizing or keeping the information given by the encoder. The qualitative analysis of the trained model and the results shows that most of the time the 1 character mistake is in the attribute Timestamp. We attribute this error to the incremental nature of the Timestamp attribute, but this will have to be further investigated with additional experiments. Through these results, we show that the Encoder-Decoder model is able to learn operation rules in a privacy setting by disclosing, generalizing or deleting specific attributes. Thus, this model is able to learn the users' preferences in regards to the privacy policy and create sub-datasets for each receiver with the appropriate information for each one.

VI. Conclusion & Future Work

Our LSTM Encoder-Decoder model is able to learn privacy operations such as disclosure, generalization, and deletion of data, therefore it can generate different data views for data receivers in an encrypted way. It allows users to train independently on their own datasets and selectively share small subsets of their key attributes to specific receivers by creating different views for each one. This offers an attractive point in regards to the utility and privacy trade-off. Two goals are achieved in an end-to-end manner by using the LSTM based encoder and decoder. One is to get an encoded version of user information while the second one is to decode this encoded information according to privacy rules defined by the user. The encoded private information of the user can not be decoded unless the right decoder is used. If an adversary tried using a different decoder on encoded information the system would not disclose any information. Without the right decoder, it would not be possible to train a decoder on encoded information due to the very high dimensionality of the LSTM hidden state vector and possible values of each LSTM network parameter. This way the users preserve the privacy of their data while receiving the benefits of the AAL environment. Moreover, the model can handle raw data even in text format, which is very beneficial in the case of medical records which usually include a lot of doctors' and nurses' notes. This work is a preliminary experimental study in preserving privacy with the use of LSTM Encoders-Decoders. One of the model's limitations, that will be addressed in future work, is the tokenization of the attributes to improve the performance. Another limitation is the use of simulated data which means they do not contain missing or abnormal values to evaluate the robustness of the method. Future work will also include the expansion of the model to real world data, as well as more complex data formats such as meta-data and multimedia formats.

ACKNOWLEDGMENTS

This work has been funded by the European Union Horizon 2020 MSCA ITN ACROSSING project (GA no. 616757). The authors would like to thank the members of the project's consortium for their valuable inputs.

REFERENCES

[1] *Council of Europe, Committee of Ministers, Recommendation No. R (97) 5 on the Protection of Medical Data*. http://hrlibrary.umn.edu/instree/coerecr97-5.html.

[2] *EU General Data Protection Regulation*. http://www.eugdpr.org.

[3] *Health Insurance Portability and Accountability Act Of 1996*. https://www.gpo.gov/fdsys/pkg/PLAW-104publ191/html/PLAW-104publ191.htm.

[4] George Bebis and Michael Georgiopoulos. Feed-forward neural networks. *IEEE Potentials*, 13(4):27–31, 1994.

[5] Kamalika Chaudhuri and Claire Monteleoni. Privacy-preserving logistic regression. In *Advances in Neural Information Processing Systems*, pages 289–296, 2009.

[6] Kamalika Chaudhuri, Claire Monteleoni, and Anand D Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12(Mar):1069–1109, 2011.

[7] Kamalika Chaudhuri, Anand D Sarwate, and Kaushik Sinha. A near-optimal algorithm for differentially-private principal components. *Journal of Machine Learning Research*, 14(1):2905–2943, 2013.

[8] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[9] Feng Dai, Dongming Zhang, and Jintao Li. Encoder/decoder for privacy protection video with privacy region detection and scrambling. In *MMM (2)*, pages 525–527, 2013.

[10] Josep Domingo-Ferrer and Vicenç Torra. A critique of k-anonymity and some of its enhancements. In *ARES 2008*, pages 990–993. IEEE, 2008.

[11] Cynthia Dwork. Differential privacy. In *Encyclopedia of Cryptography and Security*, pages 338–340. Springer, 2011.

[12] Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. Boosting and differential privacy. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 51–60. IEEE, 2010.

[13] Christoph Goller and Andreas Kuchler. Learning task-dependent distributed representations by backpropagation through structure. In *Neural Networks, 1996., IEEE International Conference on*, volume 1, pages 347–352. IEEE, 1996.

[14] Geoffrey E Hinton. Deep belief networks. *Scholarpedia*, 4(5):5947, 2009.

[15] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[16] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[17] Andreas Holzinger, Randy Goebel, Vasile Palade, and Massimo Ferri. Towards integrative machine learning and knowledge extraction. In *Lecture Notes in Artificial Intelligence LNAI 10344*, pages 1–12. Springer, Cham, 2017.

[18] Andreas Holzinger, Markus Plass, Katharina Holzinger, Gloria Cerasela Crisan, Camelia-M. Pintea, and Vasile Palade. A glass-box interactive machine learning approach for solving np-hard problems with the human-in-the-loop. *arXiv:1708.01104*, 2017.

[19] Andreas Holzinger, Klaus Schaupp, and Walter Eder-Halbedl. An investigation on acceptance of ubiquitous devices for the elderly in an geriatric hospital environment: using the example of person tracking. In *Lecture Notes in Computer Science, LNCS 5105*, pages 22–29. Springer, Heidelberg, 2008.

[20] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[21] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, pages 106–115. IEEE, 2007.

[22] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3, 2007.

[23] Bernd Malle, Peter Kieseberg, Sebastian Schrittwieser, and Andreas Holzinger. Privacy aware machine learning and the right to be forgotten. *ERCIM News (special theme: machine learning)*, 107(3):22–23, 2016.

[24] Paul Ohm. Broken promises of privacy: Responding to the surprising failure of anonymization. 2009.

[25] Hyoungmin Park and Kyuseok Shim. Approximate algorithms for k-anonymity. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 67–78. ACM, 2007.

[26] Benjamin IP Rubinstein, Peter L Bartlett, Ling Huang, and Nina Taft. Learning in a large function space: Privacy-preserving mechanisms for svm learning. *arXiv preprint arXiv:0911.5708*, 2009.

[27] Anand D Sarwate and Kamalika Chaudhuri. Signal processing and machine learning with differential privacy: Algorithms and challenges for continuous data. *IEEE signal processing magazine*, 30(5):86–94, 2013.

[28] Deepika Singh, Johannes Kropf, Sten Hanke, and Andreas Holzinger. Ambient assisted living technologies from the perspectives of older people and professionals. In *Lecture Notes in Computer Science LNCS 10410*, pages 255–266. Springer, Cham, 2017.

[29] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[30] Martin J Wainwright, Michael I Jordan, and John C Duchi. Privacy aware learning. In *Advances in Neural Information Processing Systems*, pages 1430–1438, 2012.

[31] Xiaokui Xiao and Yufei Tao. Personalized privacy preservation. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 229–240. ACM, 2006.

[32] Jun Zhang, Zhenjie Zhang, Xiaokui Xiao, Yin Yang, and Marianne Winslett. Functional mechanism: regression analysis under differential privacy. *Proceedings of the VLDB Endowment*, 5(11):1364–1375, 2012.

[33] Shu Zhang, Dequan Zheng, Xinchen Hu, and Ming Yang. Bidirectional long short-term memory networks for relation classification. In *PACLIC*, 2015.

[34] Martina Ziefle, Carsten Rcker, and Andreas Holzinger. Medical technology in smart homes: Exploring the user's perspective on privacy, intimacy and trust. In *35th COMPSAC*, pages 410–415. IEEE, Munich, 2011.

6.3 Single Encoder and Multiple decoder Model

In this set of experiments, we created a single Encoder and multiple Decoders model to improve the flexibility and practicality of the model. since having a separate encoder for each end-user cannot be efficient. In addition, a single encoder model lowers the number of parameters to be trained significantly while allowing the addition of new decoders to the model. In this model, the third parties that have access to decoder will be able to access subsets of obfuscated and anonymized dataset. It still holds that the third parties that have no access to a decoder, will not be able to detect raw dataset. The model offers a cost-efficient way to transform the sensitive data before sharing it with third parties.

In the experiments, we use the previous experiment synthetic dataset which consists of patient-level sensitive features such as demographics, habits, and medical diagnoses. The synthetic data were modeled as it was not possible to obtain a large amount of real dataset of this type from any organization. In the synthetic dataset, each data entry includes personal, sensor, and medical data types. The personal data contain attributes like name, age, gender, height, weight, address, occupation, phone number, and marital status. The potentially identifiable attributes include gender and birth date. The sensor data consists of presence, temperature, light, window, door sensors, and energy consumption. The last data type category includes sensitive medical information such as blood pressure, glucose level, and any medical disease. For our obfuscation scheme the data from each entry can be generalized (G), deleted (D), or fully disclosed (F). Four separate views are created for the different stakeholders, which are family members, doctor, caregiver, and researcher and each stakeholder has a different decoder output due to their different access levels on user information. The model developed uses a Recurrent neural network (RNN) with Gated Recurrent Unit (GRU) [96] cell encoder to map the input sequence to a vector of fixed dimensionality, and then another RNN with GRU cell is used to decode the target sequence from the vector. The details of the model is shown in Figure 6.1

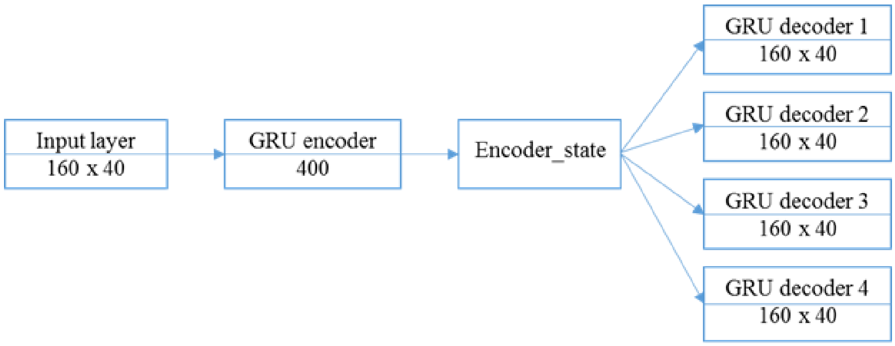


FIGURE 6.1: GRU Network for single encoder and multiple decoder

The model performs three operations in the encoder input which are Fully disclosed: keeping the data as it is; Deletion: removing the data and Generalization: replacing the value with a less specific but semantically consistent value. The network learns the type of operation to be performed based on the type of data and user preferences. The preferences are designed and set initially and then the model is trained on those parameters. For example, the data about energy consumption are generalized for the family member and the researcher but deleted for the doctor and caregiver. The model uses $800,000$ data entries for training and $200,000$ unseen entries for testing. Each user entry has 160 characters at most and different attributes are separated with "|" symbol to easily distinguish them and the entries end with the special token 'eos'. The dictionary has a set of 40 characters. Each entry is zero-padded to the maximum number of characters (160 in our experiment) to address the variable sequence lengths. In order to test if the model is resilient against missing data, we inserted missing values into the simulated data in the Glucose Levels attribute. The network is trained using Adam optimizer at a learning rate of 0.004. Figure 6.2 shows the reconstruction error of the network during training. The model training took around 12 hours on Zotac GeForce GTX 1080 Ti AMP Graphics Card (NVIDIA, 11GB GDDR5X, 352bit).

We also evaluated the model by using a 1D-CNN network for encoding. 1D-CNN network gained popularity due to its faster performance over RNN-based network. In this model, user data are tokenized into characters, and one-hot encoding is used for each character. Character level encoding is chosen since user data consists of many numeric

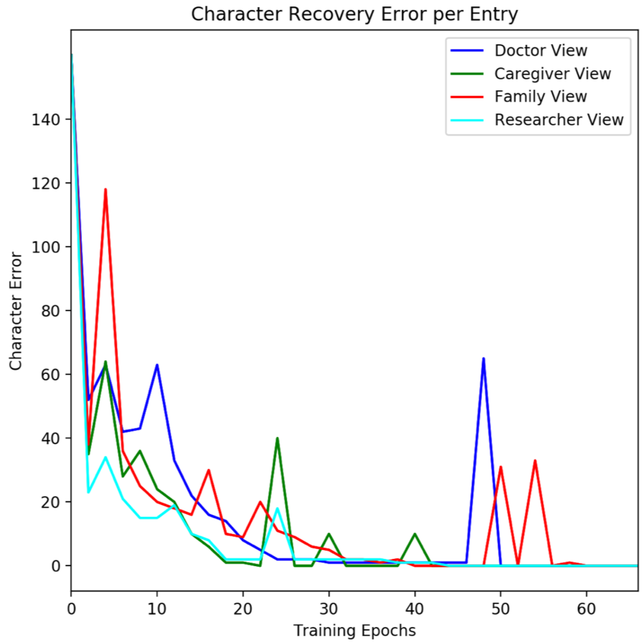


FIGURE 6.2: Character recovery error for GRU

values and text which may not exist as word vectors. The decoder is kept the same for both GRU and 1D-CNN encoders. The details of the 1D CNN network model is shown in Figure 6.3.

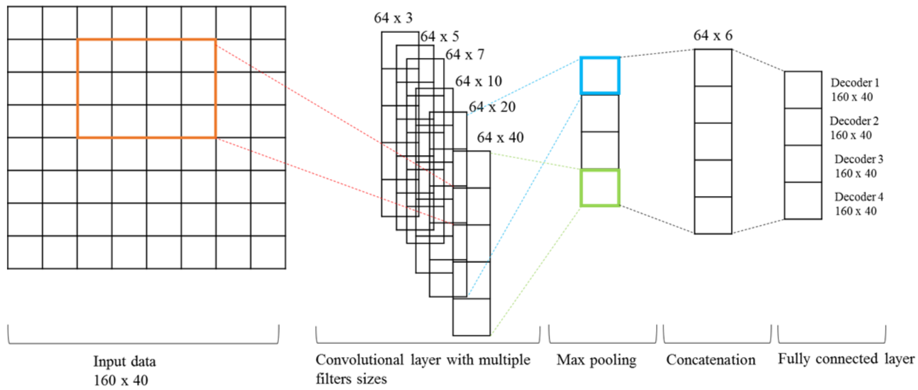


FIGURE 6.3: 1D-CNN Network for single encoder and multiple decoder

The initialization is done randomly with a Rectified Linear Unit (ReLu) activation function and the filter sizes used are 3, 5, 7, 10, 20, 40. As can be seen in Figure 6.4, the reconstruction error is decreasing faster and more uniformly in character recovery error than the GRU model, and the overall model converges faster. The training time was 10 hours on a LINUX machine with a Zotac GeForce GTX 1080 Ti AMP Graphics Card (NVIDIA, 11GB GDDR5X, 352bit).

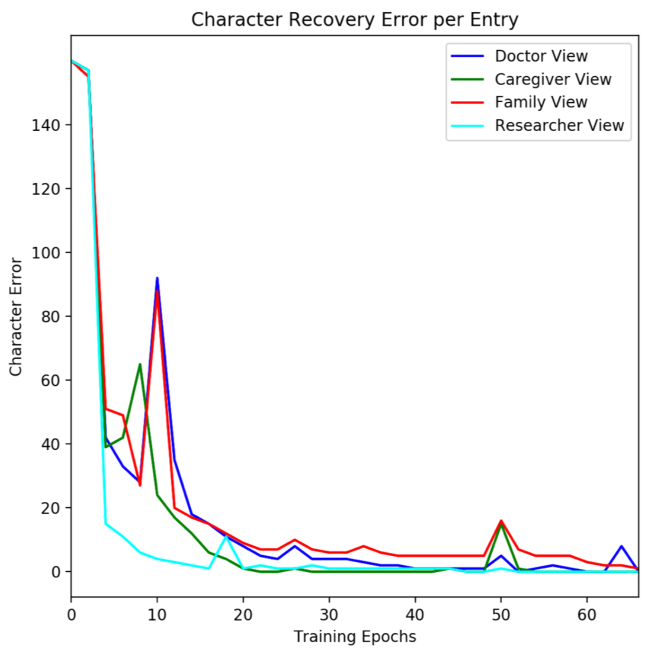


FIGURE 6.4: Character recovery error for 1D-CNN

6.3.1 Addition of Stakeholder on the Existing Anonymization

In order to examine the scalability and effectiveness of the mechanism, we tested the addition of an extra decoder to the existing encoder-decoders set. The addition of decoder will help in investigating how easy it would be to create another view of the data for new stakeholders or services requesting access to the data. In this case, the training of the model is first done with three stakeholders (Doctor, Caregiver, and Family Member), and then the Researcher view is added as an extra decoder. As can be seen from Figure 6.5, the reconstruction of the existing views is done with a minimum loss, and the added Researcher's Decoder quickly reaches to zero loss. The results show the ease in adding new stakeholders to the existing ones, thus scaling the mechanism for as many stakeholders and third parties as required. (The training time for the added stakeholder was 1.5 hours on a LINUX machine with Zotac GeForce GTX 1080 Ti AMP Graphics Card (NVIDIA, 11GB GDDR5X, 352bit.)

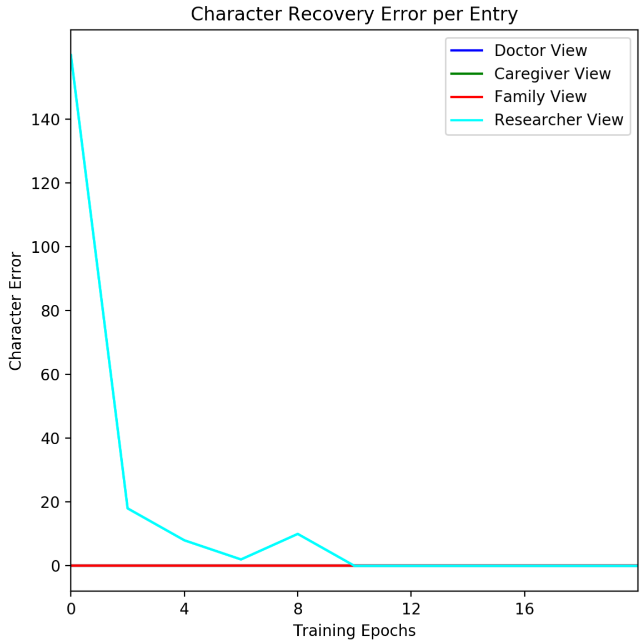


FIGURE 6.5: Addition of decoder with GRU

6.3.2 Conclusion

We developed a simple anonymizing encoder-decoder model for privacy-preserving which is able to learn privacy operations and data transformation and can generate different data subsets and views for external parties. This sequence to sequence model is trained similarly to machine translation task in which raw user data is translated into privacy enabled data. The advantage of the mechanism is that instead of sharing raw data, encoded data is shared with the stakeholders and the encoded user information cannot be observed unless the right decoder is used. In order to correctly decode the encoded data, an adversary should be able to guess the exact configurations of decoder parameters. The advantage of this approach compared to the classic encryption technique is that data analysis and classification can be done faster and more cost-effectively on the encoded data instead of encrypted data. Also, if the attacker cracked the data in the encryption model, it would be human-readable data whereas, in the encoder-decoder model, the attacker would not only need to have access to encoded data but also exact decoder parameter which includes hundreds or thousands of model weights are required to decode the data. This approach offers more flexibility than a split network by providing the option of decoding the data where it is deemed necessary and applicable to any datasets such as text, sensors, images, videos, and meta-data. In our experiments, we showed that the solution is efficient and practical to scale, especially if multiple stakeholders need to make use of it, by adding new decoders to the existing set.

CHAPTER 7

Dialogue Systems in Smart Home

7.1 Introduction

In the last decade, there has been an increase in research of computational systems using natural language for user interaction. Different systems have been developed in previous works to simulate the natural language communication between human and machine. These systems communicate with users in natural language (text, speech, or both) and are popularly known as conversational agents or dialogue systems. Dialogue systems are classified as open-domain dialogue systems (non-task oriented) and task-oriented dialogue systems. There exist various large size datasets for open-domain unstructured dialogue systems such as Ubuntu, Cornell, Twitter, and Open Subtitles. These datasets are developed by crawling online chats or movie subtitles. However, they are often low-quality datasets and are very noisy. Besides, there are no benchmark metrics to compare the performance of dialogue managers against each other. Generally, in the Ubuntu dataset, retrieval-based dialogue managers are compared with different recall metrics, or humans evaluate generator-based dialogue managers. But a human evaluation of a dialogue manager is an expensive process and subjective in small numbers. In addition, recall metrics are often considered from a small number of candidate replies, for example, the original answer is in among 10 candidate answers while the possible candidate answers are around 100k. Lack of benchmark metrics for translation of dialogue history-reply makes it difficult to evaluate and compare dialogue managers performance in the non-task oriented setting. Furthermore, there is no publicly available dataset with human annotations on the quality of dialogue-reply pairs which is required to develop and test such metrics. Therefore, in this work, we developed a human-annotated dialogue dataset from a subset of the Cornell movie dialogue dataset. The first publication of this chapter presents a detailed description of the developed dataset and preliminary results using supervised models and word-overlap metrics. The second publication of the chapter presents a detailed overview of the existing methods for training dialogue manager and proposed a new method where the text is processed as an image. The evaluation of the new image-based method is performed on Facebook bAbI Task 1 dataset in the Out Of Vocabulary setting.

The content of this chapter is based on the following publications:

Merdivan, E.*, **Singh, D.***, Hanke, S., Kropf, J., Holzinger, A. and Geist, M., 2020. Human annotated dialogues dataset for natural conversational agents. Applied Sciences, 10(3), p.762.

Contribution: Deepika Singh and Erinc Merdivan equally contributed to this work. The development of website, data collection and analysis was mainly performed by Deepika Singh. The experiments on the dataset was carried out together with Erinc Merdivan. The manuscript was mainly written by Deepika Singh together with continuous discussion with Erinc Merdivan. The remaining authors contributed in the revision of the manuscript.

Merdivan, E.*, **Singh, D.***, Hanke, S. and Holzinger, A., 2019. Dialogue systems for intelligent human computer interactions. Electronic Notes in Theoretical Computer Science, 343, pp.57-71.

Contribution: Deepika Singh and Erinc Merdivan equally contributed to this work. The designing of method is performed together with Erinc Merdivan. Erinc Merdivan implemented the experiments of this work and the manuscript was written by Deepika Singh. The other authors contributed to the scientific discussion and revision of the manuscript.

7.2 Publication IX: Human Annotated Dialogues Dataset for Natural Conversational Agents

Article

Human Annotated Dialogues Dataset for Natural Conversational Agents

Erinc Merdivan [1,2,*,†], Deepika Singh [1,3,*,†], Sten Hanke [4], Johannes Kropf [1], Andreas Holzinger [3] and Matthieu Geist [5]

1 AIT Austrian Institute of Technology, 2700 Wiener Neustadt, Austria; Johannes.Kropf@ait.ac.at
2 CentraleSupélec, Université de Lorraine, CNRS, LORIA, F-57000 Metz, France;
3 Holzinger Group, HCI-KDD, Institute for Medical Informatics/Statistics, Medical University Graz, 8036 Graz, Austria; andreas.holzinger@medunigraz.at
4 FH Joanneum Gesellschaft mbH, 8020 Graz, Austria; sten.hanke@fh-joanneum.at
5 Université de Lorraine, CNRS, LIEC, F-57000 Metz, France (now at Google Brain); matthieu.geist@univ-lorraine.fr
* Correspondence: merdivane@gmail.com (E.M.); deepika.singh@medunigraz.at (D.S.)
† These authors contributed equally to this work.

Received: 16 December 2019; Accepted: 16 January 2020; Published: date

Abstract: Conversational agents are gaining huge popularity in industrial applications such as digital assistants, chatbots, and particularly systems for natural language understanding (NLU). However, a major drawback is the unavailability of a common metric to evaluate the replies against human judgement for conversational agents. In this paper, we develop a benchmark dataset with human annotations and diverse replies that can be used to develop such metric for conversational agents. The paper introduces a high-quality human annotated movie dialogue dataset, HUMOD, that is developed from the Cornell movie dialogues dataset. This new dataset comprises 28,500 human responses from 9500 multi-turn dialogue history-reply pairs. Human responses include: (i) ratings of the dialogue reply in relevance to the dialogue history; and (ii) unique dialogue replies for each dialogue history from the users. Such unique dialogue replies enable researchers in evaluating their models against six unique human responses for each given history. Detailed analysis on how dialogues are structured and human perception on dialogue score in comparison with existing models are also presented.

Keywords: conversational agents; dialogue systems; chatbots

1. Introduction

The primary goal of intelligent dialogue systems in real-life applications is to enable efficient communication between humans and computers in a natural and coherent manner. A dialogue system requires a large amount of data to learn meaningful features and response generation strategies for building an intelligent conversational agent. Various methods, including deep learning techniques, have contributed immensely towards building dialogue systems in several real-life application domains such as natural language processing, recommender systems and question-answering systems.

It is still challenging to build a system that can understand the underlying semantics of the user input sequence, and generate coherent and meaningful responses due to limited and domain-specific dialogue datasets [1]. Previous works have developed variously structured [2,3] and unstructured [4] large datasets to train dialogue managers for the dialogue systems. One major challenge while training a dialogue manager is a lack of benchmark metrics which can be used to measure and compare performance of dialogue managers for non-task-oriented dialogue systems. Furthermore, there is no publicly available dataset with human annotations on the quality of dialogue–reply pairs to develop such metrics.

Dialogue has a certain level of abstraction due to natural language and human knowledge base; in other words, there is a very high response diversity which is unlikely to be captured by a single response [5]. Therefore, human perception of the dialogue is of utmost importance to provide diverse and unique dialogue replies depending on the dialogue history encapsulating individual's language, experience, knowledge and writing preferences. In addition, human-generated replies for the given dialogue context will help in developing robust dialogue managers, since it is not feasible to simulate good responses by using only templates. This is required in various domains but in the medical domain generally and in ambient assisted living specifically [6].

In order to address these challenges, we have developed a high-quality human annotated multi-turn movie dialogue dataset, HUMOD, from a subset of the Cornell movie dialogue dataset [7]. The collected dataset contains human annotations on fictional conversations of the movie scripts and diverse human generated replies. We have chosen Cornell movie dialogue dataset as our base dataset since it has diverse conversations on a wide span of topics, which are close to human spoken language [8], thus making the dataset ideal to train open domain dialogue systems. In [8], it is shown with quantitative and qualitative analyses that movie language is a potential source for teaching and learning spoken language features.

An example from the HUMOD dataset is shown in Figure 1. The dataset is created such that each dialogue context has two possible replies, similar to [9]—first is the actual reply of the dialogue context, which we named as positive reply and second is a reply that is sampled uniformly from the set of all possible replies, which is named as the candidate negative reply for the dialogue context. In Figure 1, a sample of the dataset is presented in three blocks, where the first block shows the 6-turn dialogue context, the second block contains a positive reply (actual reply from Cornell movie dataset) and candidate negative reply (sampled reply) and the third block shows the two human generated replies for the dialogue context. In [9,10], training is done using both positive and candidate negative replies as Next, Utterance Classification (NUC). In NUC, candidate replies can also be related to context; therefore, it is important to investigate the relevance of candidate negative reply together with actual reply. The ratings on both (positive and candidate negative) the dialogue replies from 1–5 (1: Irrelevant; 2: Weakly relevant; 3: Neutral; 4: Fairly relevant; 5: Highly relevant) are given by users. Furthermore, there are six (three for positive and three for candidate negative) human-generated replies for each dialogue context. Thus, six unique replies for each context of HUMOD dataset will help in evaluating the models against the provided human responses. The detailed descriptions of the dataset design and collection, analysis, and task formulation are discussed in Section 3. In addition, we provide benchmark performances of the supervised Hierarchical Attention Network (HAN) [11] model (with two different loss functions), Bidirectional Encoder Representations from the Transformers (BERT) [12] model and word-overlap metrics such as BLEU [13], ROUGE [14] and METEOR [15] to analyze the quality of dialogue replies for the given context. The dataset is also publicly available at: https://github.com/erincmer/HUMOD.

Dialogue Context

Speaker 1: Where did you meet Miss Lawson?

Speaker 2: At a dinner party -- about eight months ago.

Speaker 1: Did you ever see her again after that?

Speaker 2: Yes -- several times.

Speaker 1: What eventually happened to your relationship with Miss Lawson?

Speaker 2: We stopped seeing each other.

Dialogue Replies (Humans rated 1 – 5)

Positive: Why?

Negative: Don't you expect me to be a little hurt?

Human Generated Replies

#1: Why did you stop seeing each other?

#2: Would you consider seeing Miss Lawson again?

⋮

Figure 1. A sample of 6-turn dialogue context with positive (actual) reply and candidate negative (sampled) reply and two examples of human generated replies for the dialogue context.

2. Related Datasets

Dialogue systems have been categorized into two groups [16]: task-oriented systems and non-task oriented systems (also known as chatbots). The aim of task-oriented systems is to assist users to complete specific tasks by understanding the inputs from the user. Manual handcrafted rules in such systems make it not only expensive and time-consuming but also limit the systems to a particular domain. Non-task oriented dialogue systems can communicate with humans on open domains, thus they can be used in real-world applications.

Dialogue systems are mainly task-oriented, designed to complete specific tasks such as airline tickets booking [17], bus information searching [18], restaurant booking [10] or railroad goods shipping [19]. Such systems perform well when the task is simple and explicit intentions of the users are well-calibrated to the system capabilities. Moreover, some of the popular domain specific datasets are bAbI simulated restaurant booking dialogues [10], Movie dialog dataset [20] and Ubuntu Dialogue Corpus dataset [4]. The Ubuntu dataset consists of large-scale unstructured dialogues with a multi-turn conversation between two persons, where negative turns of the dialogue are created by sampling. Major drawbacks of the Ubuntu dataset are that it has no annotation on context–reply pairs and no different replies for the same context.

Many recent works use conversational models for open-domain datasets such as Twitter Dialog Corpus [21] and the Chinese Weibo dataset [22] that are posts and replies from social networking sites. PERSONA-CHAT dataset [23] introduces chit-chat dialogues between crowd-sourced participants based on their given profile. The DailyDialog [24] dataset uses manual labeling with three human experts to develop a dataset with communication intention and emotion information. Moreover, the other large-scale datasets that are often used to train neural network-based conversational models are movie-subtitles datasets such as OpenSubtitles [25], Cornell Movie-Dialogue Corpus [7], Movie-DiC [26] and Movie-Triples [27]. In this work, we selected 4750 dialogues from the Cornell movie dialogue dataset as our base dialogues which are then used for human annotation as explained in the next section. A comparison of the HUMOD dataset with the existing movie dialogue datasets is shown in Table 1.

In previous works of dialogue response evaluation, different unsupervised metrics have been adopted [28] to evaluate (or how not to evaluate) dialogue systems with human judgments in the Ubuntu and Twitter corpus dataset. In addition, supervised metrics [29] have been implemented that are trained to predict human judgments with or without reference reply. In this paper, we perform a correlation of human judgment with both supervised and unsupervised metrics in order to provide preliminary benchmark metrics on the HUMOD dataset.

Table 1. A comparison of existing movie dialogue datasets with the HUMOD dataset. (*) denotes that the HUMOD dataset can be extended by replacing the diverse replies with the original reply as explained in Figure 4.

Dataset	# of Dialogues	Human Annotated	Diverse Replies	Description
Cornell Movie-Dialogue [7]	220K	No	No	Conversation from the movie scripts.
Movie-DiC [26]	132K	No	No	American movie scripts.
Movie-Triples [27]	245K	No	No	Dialogues of three turns between two interlocutors.
OpenSubtitles [25]	36M	No	No	Movie subtitles which are not speaker-aligned.
HUMOD	28.5K *	Yes	Yes	Conversation from the movie scripts with 1 to 5 ratings and six diverse replies from humans.

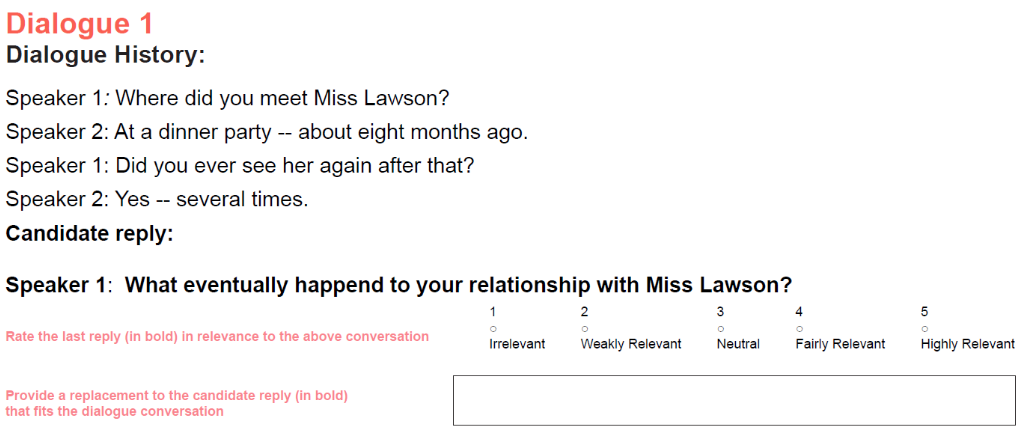


Figure 2. Screenshot of dialogue context with positive (actual) reply.

3. Human Annotated Movie Dialogues Dataset (HUMOD)

3.1. Dataset Design and Collection

We developed a website for data collection and used the crowdsourcing platform Amazon Mechanical Turk (AMT) (Full anonymity of the users were maintained and no ethical concerns were raised by the host institution) for collecting human annotations on selected Cornell movie dialogues. The dataset creation is performed in the following steps: First, we used simple random sampling to select 4750 dialogue contexts from the Cornell movie dataset in which each dialogue is divided into utterances. Since the dataset contains human–human dialogues, we assigned each utterance alternatively as Speaker 1 and Speaker 2 as shown in Figure 1. The dataset consists of multi-turn dialogues ranging from two to the maximum of seven turns. Then, each dialogue history is appended with an actual reply from the movie dialogue dataset and also with a candidate negative reply which is sampled uniformly from a possible set of replies, thus making the dataset size of 9500. At the beginning, basic demographics of the users were asked including Age range, Gender, and English proficiency level. These demographic variables are collected so that the dataset can be tailored as per the needs and preferences of researchers to develop conversational agents for different demographics. In the next step, each AMT user was given a task of 20 sampled dialogue conversations and was asked to rate the last reply of the dialogue conversation from 1–5, (where 1: Irrelevant reply; 2: Weakly relevant reply; 3: Neutral; 4: Fairly relevant reply; 5: Highly relevant reply) according to the dialog context and provide their relevant reply in replacement to the last reply. The total number of completed tasks were 1425.

Each dialogue–reply pair is rated by three different users, such that, for each dialog context, there are six unique responses, resulting in a total dataset size of 28,500. In order to maintain the high quality responses in the dataset, we selected responses of only those AMT users who have more than 80% hit approval rate. Each task was evaluated by a human expert and low quality tasks consisting of short or generic replies, erroneous replies consisting of gibberish text (user's reply for given dialogue context), user's replies in different languages (other than English) and the same ratings given to every dialogue–reply pairs were discarded.

Figures 2 and 3 present the screen shots from the website, showing how dialogue appears to the users with positive reply and candidate negative reply for the particular dialogue context. Since each task contains 20 randomly selected dialogue context-reply pairs, the chances of the same dialogue context appearing with both positive and candidate negative replies to the same user is very minimal.

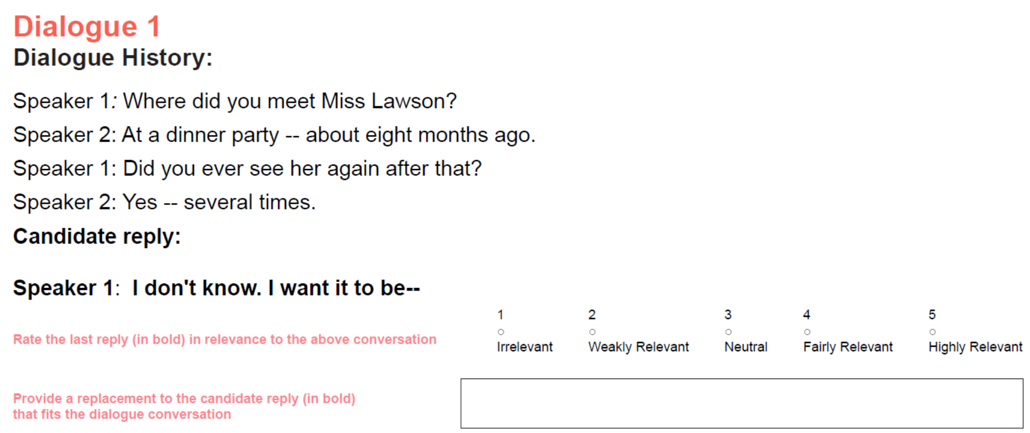


Figure 3. Screenshot of dialogue context with candidate negative (sampled) reply.

The six diverse replies for each dialogue context in the proposed dataset enable the extendability feature of HUMOD dataset. An example with reference to a dialogue context of how the dataset can be extended using human replies is shown in Figure 4. The seven-turn dialogue of the dataset is divided into two parts. The first part of the figure (upper block) contains the first four turns of the dialogue context and the second part (lower block) consists of all possible replies for the remaining turns of the seven-turn dialogue context. In the lower block, the second column displays the original reply from the Cornell movie dialogue dataset, and the remaining columns show the human-generated replies (four out of six) as a replacement to the original reply. Different combinations of replies in each turn can be used to augment the dataset with the caution of adding some noise. From the example in Figure 4, we analyze that, with five possible replies in each turn, a total of 125 diverse dialogues can be created, while many of the created dialogues are as coherent as the original one. This extension feature would enable to generate extra data inexpensively, which could help in training generative models to generalize better and learn to provide various replies for the same dialogue context. The extendability assumption comes from the fact that humans generated alternatives replies while being coherent with dialog context. However, it is not very beneficial to perform topic modelling or word commonality analysis since they do not reflect coherence of alternative reply in the given dialog context and only comparing to dialogue ground reply would lead to poor results.

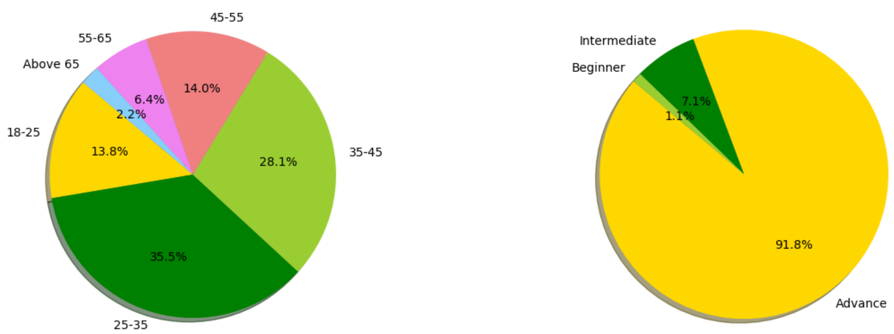
Dialogue Context
{Turn 1}Speaker 1: Where did you meet Miss Lawson?
{Turn 2} Speaker 2: At a dinner party -- about eight months ago.
{Turn 3}Speaker 1: Did you ever see her again after that?
{Turn 4}Speaker 2: Yes -- several times.

#Turns	Original reply from Cornell movie dataset	Human Generated Replies			
{Turn 5} Speaker 1:	What eventually happened to your relationship with Miss Lawson?	Where did you see her?	And so, what happened?	Where and when is the last time you saw Miss Lawson?	Did anything happen during those meetings?
{Turn 6} Speaker 2:	We stopped seeing each other.	She is my best friend but I am not able to contact her since a month	Nothing came of it.	After few dates, our relationship went sore when we spoke about our career.	I lost track of her.
{Turn 7} Speaker 1:	Why?	Why did you stop?	Would you consider seeing Miss Lawson again?	Don't you expect her to be a little hurt?	You guys were so good for each other.

Figure 4. The extendability approach of the HUMOD dataset.

3.2. Data Analysis

This section presents the description of the HUMOD dataset and insights on human annotations. The users' demographics obtained are shown in Figure 5. The aim is to achieve human perception towards dialogue across the age groups, and we managed to achieve diversity among users. The gender ratio obtained was 46% of male and 54% female, and users were also not restricted to their geography. The majority of the users (91.8%) were having advanced English level understanding as can been seen in Figure 5b, which makes the HUMOD dataset of high-quality.



(a) Age **(b)** English proficiency level

Figure 5. Users' demographics.

The number of turns per dialogue used in the dataset is given in Figure 6. Figure 7 shows the distribution of the number of words among all the obtained dialogue replies from the users. Figure 8 presents the evaluation of human generated replies on a selected HUMOD dataset. As can be seen from the figure, 2.8% of the AMT users rated Irrelevant, whereas, in the original dataset, the Irrelevant score was around 3.3% for the same dialogue–reply pairs. Figure 9 presents the histogram of user ratings for positive and negative replies. As can be seen from the figure, positive replies are rated high (4 or 5) by users, about 74% (10,610 out of 14,250) and negative replies are rated low (1 or 2), around 73% (10,404 out of 14,250). It is crucial to take into account the number of low scores in positive replies and high scores in negative replies which introduce noise in the dataset. Thus, it is important to get human judgments even for positive pairs. The negative training data can be even harmful for binary text classification [30], which may occur in the case of next utterance classification in dialogue settings. Therefore, another approach for training dialogue managers could be Positive-Unlabeled (PU) Learning [31]. In PU Learning, there is a positive set, and instead of negative there is an unlabeled dataset which can contain positive and negative samples. Existing methods [31,32] which perform well in PU Learning problems can also be applied to next utterance classification for dialogues.

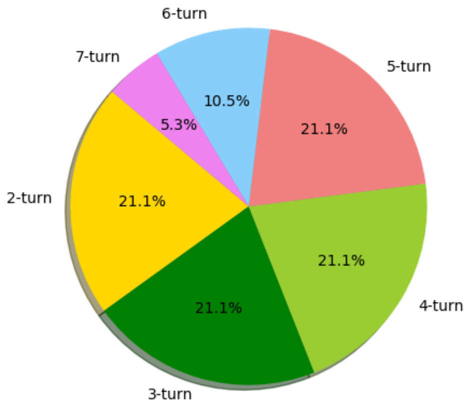


Figure 6. Number of turns in dialogue.

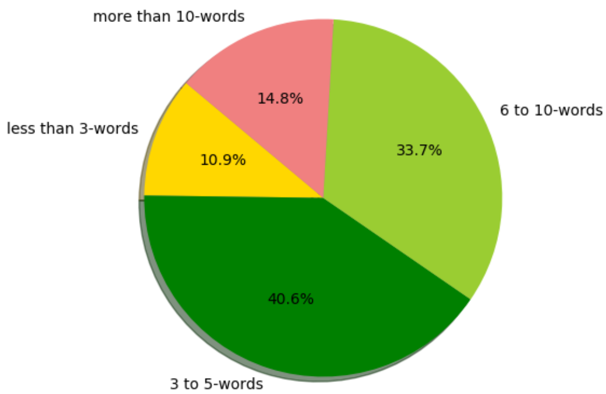


Figure 7. Number of words in human responses.

The human generated replies of HUMOD dataset are highly relevant since users were asked to provide relevant replies for the given dialogue context during data collection. However, in order to perform initial evaluation on human generated replies, we randomly selected 30 dialogue reply pairs with six diverse replies of different turns (for example, five dialogues for each turn i.e., 2-turn, 3-turn, 4-turn, 5-turn, 6-turn, 7-turn) from the HUMOD dataset. Each diverse reply set is rated by two AMT users. The number of human generated replies for evaluation could be higher with increasing cost, but a strong pattern of relevancy can be easily seen in these sampled human generated replies.

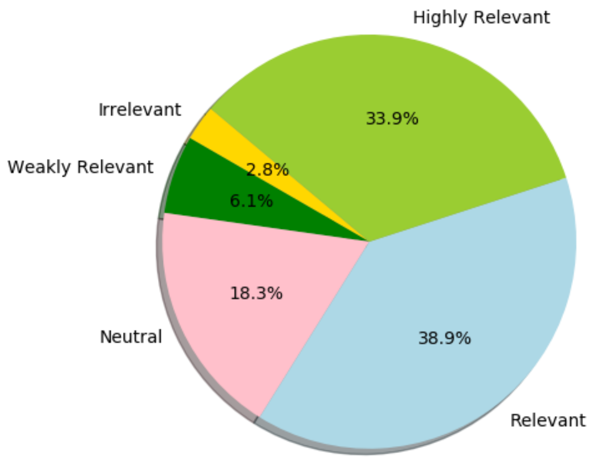


Figure 8. Evaluation of human responses on the selected HUMOD dataset.

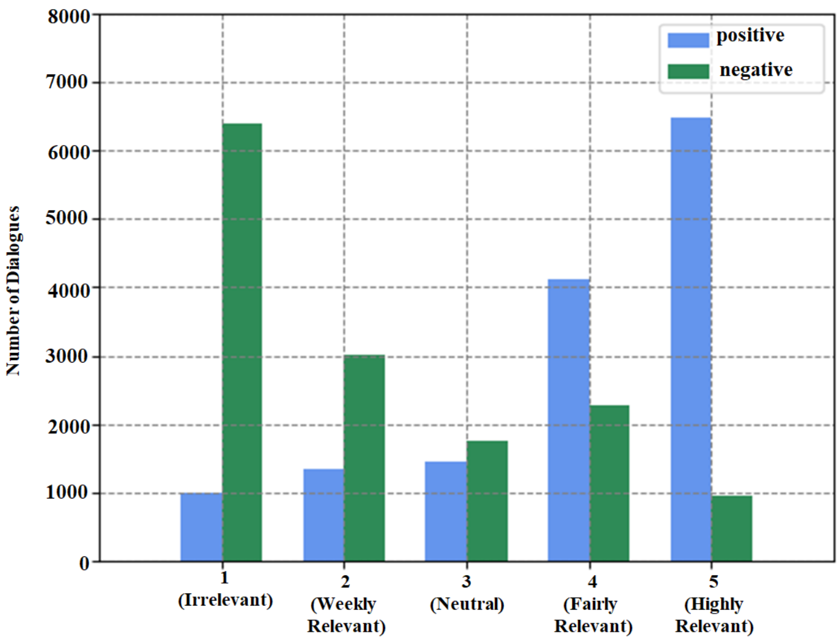


Figure 9. Human scores vs. Positive and Candidate Negative Dialogue Pairs.

To evaluate human agreement on their responses with each other, we calculated weighted Cohen's kappa score [33] between human ratings. We calculated weighted kappa score for different configurations of three ratings for each dialogue context and reply pair. From the three ratings of each reply, we calculated weighted kappa score for the closest two (as a majority voting) ratings, the highest two ratings, the lowest two ratings, and on the random selection of two ratings. For example, if a dialogue context-reply pair is rated (5, 4 and 1), we keep the closest two (5 and 4) and randomly assign them to Rater 1 and 2. Results for all different configurations can be seen in Table 2.

Table 2. Inter-annotator agreement.

Rater	Cohen's Kappa Score (κ)
Closest two ratings	0.86
Lowest two ratings	0.57
Highest two ratings	0.55
Random two ratings	0.42

3.3. Task Formulation

We provide a benchmark performance with supervised models and word-overlap metrics on the proposed HUMOD dataset. The task formulated is to predict human ratings for given dialogue history–reply pair from Irrelevant (1) to Highly-relevant (5). The dataset will enable researchers to test their dialogue reply metrics against human ratings.

4. Methods

This section provides details about the supervised models and word-overlap metrics, which are used to predict the human scores for dialogue replies. For supervised models, a network inspired from hierarchical attention networks (HAN) is designed as shown in Figure 10. Recurrent Neural Network (RNN) based models are implemented in a way such that it does not require a reference answer, since it makes them more applicable but difficult to train in comparison to the models which uses reference response. Our model is the same as the Automatic Dialogue Evaluation Model (ADEM) [29] model with an additional attention layer. In ADEM, training and evaluation was performed with and without reference reply. Although ADEM results without reference reply is lower in comparison to with the reference reply, we preferred to train without a reference as it is not practical to have reference reply in real life applications. In addition, we used a Bidirectional Encoder Representations from Transformers (BERT) supervised model for comparison with HAN and word-overlap metrics. Word-overlap metrics require reference answers unlike supervised models and only use reference reply without any context information, which makes them perform poorly. These metrics are easy and inexpensive to run without the need for training and can be applied as long as there is a reference text. Supervised models require training to predict the dialogue reply score which makes it difficult to generalize and expensive to use in different domains and tasks. We used metrics such as BLEU [13], METEOR [15] and ROUGE [14], where the first two are used widely in translation and the last one for assessing the quality of the summarized text.

4.1. BLEU

BLEU [13] score calculates the precision of n-grams of machine-generated dialogue replies in human replies. Often, the brevity penalty is used to avoid short sentences. From different n-gram choices, the most common one is n = 4 where the weighted average of different BLEU (1 to 4) scores are used to evaluate the machine-generated replies. BLEU first calculates a modified precision score for each n-gram length as below:

$$P_n(r,\hat{r}) = \frac{\sum_{ngr} Count_{matched}(ngr)}{\sum_k Count(ngr)}, \tag{1}$$

where ngr represents all possible n-grams of length n in hypothesis sentences. Later,

$$\text{BLEU-4} = b(r, \hat{r}) \exp\left(\sum_{n=1}^{4} 0.25 \log P_n(r, \hat{r})\right). \tag{2}$$

To avoid shorter sentences, the modified precision score is often multiplied with a brevity penalty to achieve the final score.

4.2. ROUGE

The ROUGE [14] score originally calculates the recall of n-grams of human dialogue replies in machine-generated dialogue replies. ROUGE can be extended to ROUGE-L, which is a combination of recall and precision ROUGE scores based on longest matching sequence (LCS). LCS only requires sentence level word order matches and it allows other words to appear between words of LCS. LCS does not need n-gram to be predefined:

$$R = \max_j \frac{l(r, \hat{r}_{i_j})}{|\hat{r}_{i_j}|}, \tag{3}$$

$$P = \max_j \frac{l(r_i, \hat{r}_{i_j})}{|r_{i_j}|}, \tag{4}$$

$$\text{ROUGE} = \frac{(1+\beta^2)RP}{R+\beta^2 P}. \tag{5}$$

4.3. METEOR

METEOR [15] score aligns human generated reply and machine-generated reply. This alignment is done word by word with an order of exact match, Porter stemming match or WordNet synonym match. It computes the parametric harmonic mean (F_{mean}) between unigram recall and unigram precision:

$$P = \frac{m}{t}, \tag{6}$$

$$R = \frac{m}{r}, \tag{7}$$

$$F_{mean} = \frac{P.R}{\alpha P + (1-\alpha)R}, \tag{8}$$

$$Pen = \gamma\left(\frac{ch}{m}\right)^{\theta}, \tag{9}$$

$$\text{METEOR} = (1 - Pen)F_{mean}, \tag{10}$$

where t and r are the total numbers of unigrams in the translation and the reference. m represents the number of mapped unigrams between reference and hypothesis sentences. Penalty term (Pen) is computed so that it takes into account matched unigrams between hypothesis and reference that are in the same order.

4.4. Hierarchical Attention Network-Based Models

In this work, we implemented two separate networks, one for dialogue context and another for the dialogue reply. Context encoder is a Hierarchical Attention Network (HAN) [11] as shown in Figure 10. It encodes dialogue context to a dialog context vector c. Reply encoder network is a biLSTM [34] network with attention mechanism on top to convert reply to a reply vector $\hat{r}$ (Figure 11).

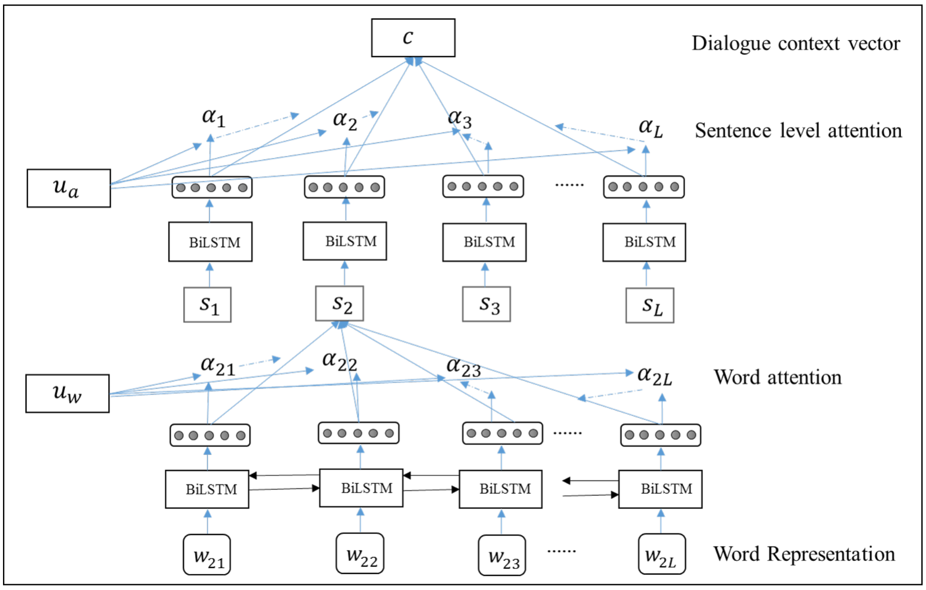


Figure 10. Hierarchical attention network for the dialogue context.

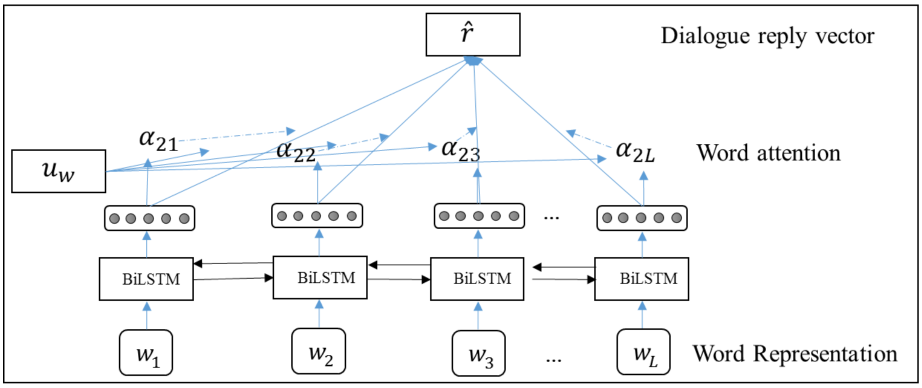


Figure 11. Dialogue reply encoder.

We implemented two different loss functions after the dialogue context and reply are encoded to vector representation. The first loss is the cross-entropy loss which classifies concatenated vectors of dialogue context and reply into five classes (1–5):

$$\text{HAN-R(MSE)}(c, \hat{r}) = \sum_i [FC([c_i, \hat{r}_i]) - h_i)]^2, \tag{11}$$

where h represents average human scores, c represents the context and $\hat{r}$ is reply vector. FC is the fully connected layer which outputs the score for the given context and reply.

Second loss, as in Equation (11), is trained with mean squared error against human score and uses a concatenated vector. We did not use cosine similarity between dialogue context and reply since the authors in [35] showed that ADEM, which is similar to ours, creates response embedding with very low vector spread in the embedding space. In dialogue, there are many alternative replies for the same context, and the same reply can be a good fit to very different contexts and common replies that occur a lot and fit to a lot of different contexts. Due to these observations, when a dialogue manager is trained using cosines distance and make embedding of context and reply similar, eventually very different texts become similar and may collapse to very small region in embedding space.

4.5. Bidirectional Encoder Representations from Transformers (BERT)

BERT [12] is a bidirectional language model that allows models to learn both left and right context in all layers. BERT is pretrained with two novel methods that are "Masked Language Model (MLM)" and "Next sentence prediction" and uses Transformer [36] with attention instead of recurrent networks like LSTM. The Next sentence prediction method is more related to our task since it trains BERT to learn the relationship between sentences, which may improve performance in case of dialogue context and reply scoring. The BERT model has achieved state-of-the-art performance on various Natural language processing (NLP) tasks and also outperformed human performance in question-answering tasks. Therefore, we fine-tuned BERT (from tensorflow-hub) on the HUMOD dataset as shown in Figure 12 in order to compare performance with other approaches. As can be seen from Figure 12, BERT takes input as tokenized dialogue context and dialogue reply separated with SEP token and outputs a final vector representation of dialogue context and reply pair. Later, this vector is classified to a relevance score of 1–5 with additional dense layer on top of BERT.

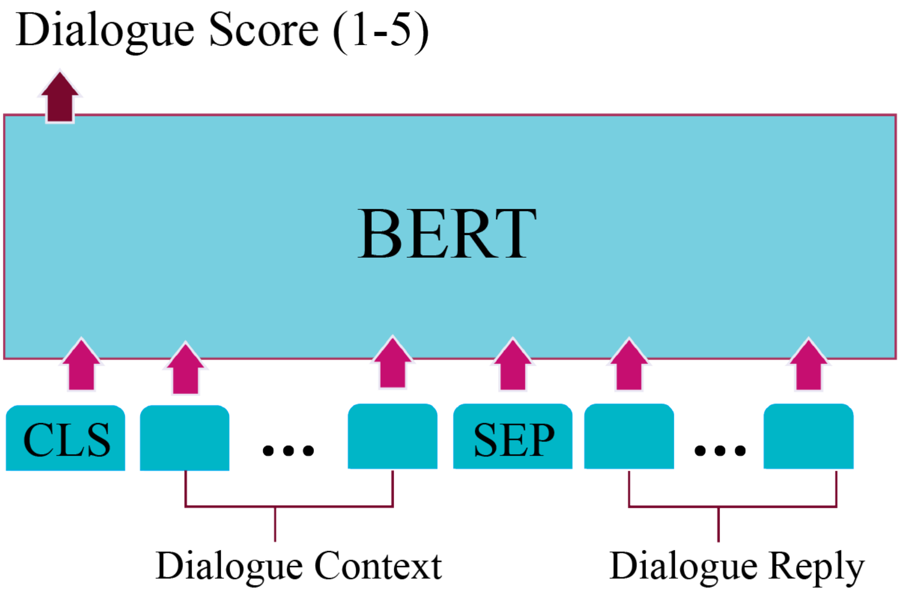


Figure 12. BERT used as a dialogue reply scorer.

5. Experiments

We performed a comparison of supervised and word-overlap metrics to see how they are correlated with human judgments. The dataset is divided into 8500 context-reply pairs for the train set and 1000 context-reply pairs for the test set.

No dialogue context is shared between train and test set. Since for each dialogue we have scores of three judges, we took the average score of three judges. Both supervised and word-overlap approaches are evaluated on the same test set. For word-overlap metrics, we normalized average human scores into the 0–1 range and calculated metrics using the NLGEval toolkit [37]. Both the context encoder and the reply encoder use 50-dimensional GloVe word vectors [38], and the dimension of 100 was chosen for each biLSTM hidden state for the HAN-based model. For the model which uses BERT, input is constructed as dialogue context and dialogue reply separated with a special token, and the final encoded vector is classified into one of the five classes.

6. Empirical Results

The preliminary results of HUMOD dataset with existing supervised and word-overlap metrics is shown in Table 3.

Table 3. Correlation of different models with human scores.

Models	Correlation (p-Value)
BLEU-4	0.055 (0.08)
ROUGE	−0.035 (0.26)
METEOR	−0.017 (0.59)
HAN-R(CE)	0.138 (<0.001)
HAN-R(MSE)	0.128 (0.003)
BERT	0.602 (<0.001)

We provided the benchmark results for the overall correlation of human judgment with different models. Although supervised models are correlated up to some degree, it is still far from applicable to use in dialogue reply scoring as widely as translation scores are used to evaluate machine translation models both in terms of human correlation and ease of use.

As can be seen from Table 3, the BERT model outperformed the word-overlap metrics and HAN model. Since the BERT model takes advantage of the language model and can be fined-tuned according to other dataset, in this experiment, we used BERT with pretrained weights and fine-tuned it for the HUMOD dataset, which may explain the performance difference of supervised models. In addition, correlation of BERT with human judges is performed to investigate the behavior of the network against different dialogue turns (shown in Table 4). We tested BERT performance on different turn dialogues to see how context length affects the performance of dialogue measure. 2-turn dialogues correlation is found to be lowest, which can be due to hardness to evaluate the dialogue reply score for very short dialogues since it contains very little context, which may increase humans' own judgement and bias. Similarly, for long-dialogue conversation, it is slightly harder for the system to contain the context and make a good understanding of the fitness of dialogue context and reply pairs.

Table 4. Correlation of BERT against different turns with human scores.

Dialogue Turn	Correlation (p-Value)
2-turn	0.52 (<0.001)
3-turn	0.58 (<0.001)
4-turn	0.67 (<0.001)
5-turn	0.61 (<0.001)
6-turn	0.66 (<0.001)
7-turn	0.59 (<0.001)

7. Conclusions and Future Work

This paper presents the human annotated movie dialogue dataset (HUMOD) for research to develop benchmark metrics for comparison of different models on human scores and generated replies. The detailed description of the dataset construction and statistics is provided. The availability of the HUMOD dataset opens up various possibilities for research and development of complex dialogue systems for real-life applications. Different replies for the same context as well as dialogue ratings can be used to develop a metric to compare methods such as BLEU. Another interesting usage of a unique diverse reply is to train generative models that generate diverse dialogue replies which may make dialogue managers more human-like in real-life applications. We present the preliminary results to provide baselines with supervised and word-overlap metrics. HAN provides better results in comparison to word-overlap metrics since these approaches do not use any context. The BERT model outperforms the HAN model and provides good correlation on the human dialogue score; however, it is harder to train and requires fine-tuning for each different dataset.

In future work, we will work on a metric that uses dialogue contexts and replies to produce a robust score without any training. We will also focus on generative models that can leverage all possible replies for a given context rather than a single context–reply pair. Another interesting approach is to generate diverse replies from single context–reply pairs and to use such created data as augmentation for dialogue managers. All this is relevant for the upcoming research stream of explainable AI, where NLU [39] plays a particular role, e.g., in the generation of human-understandable explanations [40], [41], where it is very useful for the development of future human-AI interfaces.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: title, Table S1: title, Video S1: title.

Author Contributions: Conceptualization, E.M. and D.S.; methodology, E.M. and D.S.; software, E.M. and D.S.; validation, E.M. and D.S.; formal analysis, E.M. and D.S; investigation, E.M. and D.S; resources, E.M. and D.S; data curation, E.M. and D.S; writing—original draft preparation, E.M. and D.S.; writing—review and editing, E.M., D.S., A.H. and M.G.; visualization, E.M. and D.S; supervision, S.H., J.K., A.H. and M.G.; project administration, S.H. and J.K.; funding acquisition, S.H., J.K. and A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded by the European Union Horizon2020 MSCA ITN ACROSSING project (GA no. 616757).

Acknowledgments: The authors would like to thank the members of the ACROSSING project's consortium for their valuable inputs. We are also thankful to all the participants of the dialogue dataset survey.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HUMOD	Human annotated movie dialogue dataset
NUC	Next utterance classification
AMT	Amazon mechanical turk
HAN	Hierarchical attention network
BERT	Bidirectional encoder representations from transformers
NLP	Natural language processing
LSTM	Long short term memory
LCS	Longest matching sequence
PU	Positive-unlabeled
CE	Context encoder
MSE	Mean squared error

References

1. Petukhova, V.; Gropp, M.; Klakow, D.; Schmidt, A.; Eigner, G.; Topf, M.; Srb, S.; Motlicek, P.; Potard, B.; Dines, J.; et al. *The DBOX Corpus Collection of Spoken Human-Human and Human-Machine Dialogues*; Technical Report; European Language Resources Association (ELRA): Paris, France, 2014.
2. Henderson, M.; Thomson, B.; Williams, J.D. The second dialog state tracking challenge. In Proceedings of the SIGDIAL 2014 Conference, Philadelphia, PA, USA, 18–20 June 2014.
3. Williams, J.; Raux, A.; Ramachandran, D.; Black, A. The dialog state tracking challenge. In Proceedings of the SIGDIAL 2013 Conference, Metz, France, 22–24 August 2013.
4. Lowe, R.; Pow, N.; Serban, I.; Pineau, J. The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. In Proceedings of the SIGDIAL 2015 Conference, Prague, Czech Republic, 2–4 September 2015.
5. Artstein, R.; Gandhe, S.; Gerten, J.; Leuski, A.; Traum, D. Semi-formal evaluation of conversational characters. In *Languages: From formal to natural*; Springer: Berlin/Heidelberg, Germany, 2009.
6. Holzinger, A. User-Centered Interface Design for disabled and elderly people: First experiences with designing a patient communication system (PACOSY). In *Computer Helping People with Special Needs, ICCHP 2002, Lecture Notes in Computer Science (LNCS 2398)*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 34–41, doi:10.1007/3-540-45491-8_8.
7. Danescu-Niculescu-Mizil, C.; Lee, L. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics. Association for Computational Linguistics, 2011.

8. Forchini, P. Spontaneity reloaded: American face-to-face and movie conversation compared. In Proceedings of the Corpus Linguistics Conference 2009 (CL2009), Liverpool, UK, 20–23 July 2009.
9. Lowe, R.; Serban, I.V.; Noseworthy, M.; Charlin, L.; Pineau, J. On the evaluation of dialogue systems with next utterance classification. In Proceedings of the SIGDIAL 2016 Conference, Los Angeles, CA, USA, 13–15 September 2016.
10. Bordes, A.; Boureau, Y.L.; Weston, J. Learning end-to-end goal-oriented dialog. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017.
11. Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; Hovy, E. Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL), San Diego, CA, USA, 12–17 June 2016.
12. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Volume 1 (Long and Short Papers).
13. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting on Association For Computational Linguistics (ACL), Philadelphia, PA, USA, 7–12 July 2002.
14. Lin, C.Y. Rouge: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*; Association for Computational Linguistics: Barcelona, Spain, 2004.
15. Banerjee, S.; Lavie, A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, Ann Arbor, MI, USA, June 2005.
16. Chen, H.; Liu, X.; Yin, D.; Tang, J. A survey on dialogue systems: Recent advances and new frontiers. *ACM SIGKDD Explor. Newsl.* **2017**, *19*, 2.
17. Zue, V.; Seneff, S.; Polifroni, J.; Phillips, M.; Pao, C.; Goodine, D.; Goddeau, D.; Glass, J. PEGASUS: A spoken dialogue interface for online air travel planning. *Speech Commun.* **1994**, *15*, 331–340.
18. Raux, A.; Langner, B.; Bohus, D.; Black, A.W.; Eskenazi, M. Let's Go Public! Taking a spoken dialog system to the real world. In Proceedings of the Ninth European Conference on Speech Communication and Technology, Lisbon, Portugal, 4–8 September 2005.
19. Allen, J.F.; Miller, B.W.; Ringger, E.K.; Sikorski, T. A robust system for natural spoken dialogue. In Proceedings of the 34th annual meeting on Association for Computational Linguistics (ACL), Santa Cruz, CA, USA, 23–28 June 1996.
20. Dodge, J.; Gane, A.; Zhang, X.; Bordes, A.; Chopra, S.; Miller, A.; Szlam, A.; Weston, J. Evaluating prerequisite qualities for learning end-to-end dialog systems. In Proceedings of the International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2–4 May 2016.
21. Ritter, A.; Cherry, C.; Dolan, W.B. Data-driven response generation in social media. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP), Edinburgh, UK, 27–31 July 2011.
22. Wang, H.; Lu, Z.; Li, H.; Chen, E. A dataset for research on short-text conversations. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP), Washington, DC, USA, 18–21 October 2013.
23. Zhang, S.; Dinan, E.; Urbanek, J.; Szlam, A.; Kiela, D.; Weston, J. Personalizing Dialogue Agents: I have a dog, do you have pets too? *arXiv* **2018**, arXiv:1801.07243.
24. Li, Y.; Su, H.; Shen, X.; Li, W.; Cao, Z.; Niu, S. DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset. In Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), Taipei, Taiwan, 27 November–1 December 2017.
25. Tiedemann, J. Parallel Data, Tools and Interfaces in OPUS. In Proceedings of the International Conference on Language Resources and Evaluation (LREC), Istanbul, Turkey, 21–22 May 2012.
26. Banchs, R.E. Movie-DiC: a movie dialogue corpus for research and development. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL), Jeju Island, Korea, 8–14 July 2012.

27. Serban, I.V.; Sordoni, A.; Bengio, Y.; Courville, A.C.; Pineau, J. Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence AAAI, Phoenix, AZ, USA, 12–17 February 2016.
28. Liu, C.W.; Lowe, R.; Serban, I.V.; Noseworthy, M.; Charlin, L.; Pineau, J. How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. *arXiv* **2016**, arXiv:1603.08023.
29. Lowe, R.; Noseworthy, M.; Serban, I.V.; Angelard-Gontier, N.; Bengio, Y.; Pineau, J. Towards an automatic Turing test: Learning to evaluate dialogue responses. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL), Vancouver, BC, Canada, 30 July–4 August 2017.
30. Li, X.L.; Liu, B.; Ng, S.K. Negative training data can be harmful to text classification. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP), Cambridge, MA, USA, 9–11 October 2010.
31. Li, X.L.; Liu, B. Learning from positive and unlabeled examples with different data distributions. In Proceedings of the European Conference on Machine Learning (ECML). Springer, Porto, Portugal, 3–7 October 2005.
32. Merdivan, E.; Loghmani, M.R.; Geist, M. Reconstruct & Crush Network. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017.
33. Cohen, J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol. Bull.* **1968**, *70*, 213.
34. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **2005**, *18*, 602–610.
35. Sai, A.B.; Gupta, M.D.; Khapra, M.M.; Srinivasan, M. Re-evaluating ADEM: A Deeper Look at Scoring Dialogue Responses. *AAAI* **2019**, *30*, 1.
36. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
37. Sharma, S.; El Asri, L.; Schulz, H.; Zumer, J. Relevance of Unsupervised Metrics in Task-Oriented Dialogue for Evaluating Natural Language Generation. *arXiv* **2017**, arXiv:1706.09799.
38. Pennington, J.; Socher, R.; Manning, C. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014.
39. Holzinger, A.; Kieseberg, P.; Weippl, E.; Tjoa, A.M. Current Advances, Trends and Challenges of Machine Learning and Knowledge Extraction: From Machine Learning to Explainable AI. In *Springer Lecture Notes in Computer Science LNCS 11015*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 1–8, doi:10.1007/978-3-319-99740-7_1.
40. Hudec, M.; Bednárová, E.; Holzinger, A. Augmenting Statistical Data Dissemination by Short Quantified Sentences of Natural Language. *J. Off. Stat.* **2018**, *34*, 981, doi:10.2478/jos-2018-0048.
41. Holzinger, A.; Kickmeier-Rust, M.; Mueller, H. KANDINSKY Patterns as IQ-Test for machine learning. In *Springer Lecture Notes LNCS 11713*; Springer Nature: Cham, Switzerland, 2019; pp. 1–14, doi:10.1007/978-3-030-29726-8_1.

7.3 Publication X: Dialogue systems for intelligent human computer interactions

ScienceDirect

Electronic Notes in Theoretical Computer Science 343 (2019) 57–71

Electronic Notes in Theoretical Computer Science

www.elsevier.com/locate/entcs

Dialogue Systems for Intelligent Human Computer Interactions

Erinc Merdivan[1]

AIT Austrian Institute of Technology GmbH, Wiener Neustadt, Austria
CentraleSupelec, Metz, France

Deepika Singh[2]

AIT Austrian Institute of Technology GmbH, Wiener Neustadt, Austria
Holzinger Group, HCI-KDD, Institute for Medical Informatics/Statistics, Medical University Graz, Austria

Sten Hanke[3]

AIT Austrian Institute of Technology GmbH, Wiener Neustadt, Austria

Andreas Holzinger[4]

Holzinger Group, HCI-KDD, Institute for Medical Informatics/Statistics, Medical University Graz, Austria

Abstract

The most fundamental communication mechanism for interaction is dialogues involving speech, gesture, semantic and pragmatic knowledge. Various researches on dialogue management have been conducted focusing on standardized model for goal oriented applications using machine learning and deep learning models. The paper presents the overview on existing methods for dialogue manager training; their advantages and limitations. Furthermore, a new image-based method is used in Facebook bAbI Task 1 dataset in Out Of Vocabulary setting. The results show that using dialogue as an image performs well and helps dialogue manager in expanding out of vocabulary dialogue tasks in comparison to Memory Networks.

Keywords: dialogue system, image-based method, chatbots

[1] Email: erinc.merdivan@ait.ac.at
[2] Email: deepika.singh@ait.ac.at
[3] Email: sten.hanke@ait.ac.at
[4] Email: andreas.holzinger@medunigraz.at

https://doi.org/10.1016/j.entcs.2019.04.010
1571-0661/

1 Introduction

The Information and Communication Technology (ICT) with which we interact in daily life is more distributed and embodied into the environment (the so called intelligent space) [1]. Especially, when designing ICT solutions for elderly people, who are very often critical towards new technology, distributed system can be even more challenging. To improve the Human Computer Interaction (HCI) with ICT solutions, a directed natural interaction and an emotional intelligence is very important [2].

User studies regarding elder behavior change over the ageing process [3] identified that *"a skill that many elderly people retain, even with significant cognitive degradation, is the ability to communicate in a multimodal face-to-face fashion. The skills for this type of interaction are acquired in infancy and early childhood and comprise of tacit, crystallized knowledge in older adulthood"*. Face-to-face interaction incorporates a wide range of non-verbal and paraverbal ways to carry semantic content complementary to the speech. It allows persons with disabilities to compensate some perception channels (e.g., hearing) by utilizing other channels of communication (e.g., body gestures). Face-to-face dialog is also characterized by well-established repair mechanisms of understanding thus enabling the listener to request a repetition or clarification by the speaker (e.g., communicated by a head nod). Moreover, face-to-face dialog has built-in mechanisms for constraining the interactants focus of attention. This focus is important as some elderly people have difficulty dividing their attention or handling distractions.

This kind of face-to-face interaction can be provided utilizing avatars. Avatars have the potential to impersonate the used technology and thus increase the acceptance of the software [4]. The interaction with avatars is able to provide multiple advantages. Avatars can, for instance, provide gestures, which in turn are able to increase the understanding of the presented information. Furthermore, the visual enrichment of verbal information i.e., adding a lip synched animated character to audio speech output, can increase the intelligence and enhance the robustness of the information transmission as known from natural speech. Therefore, a consistency between the visual and vocal output is of uttermost important [5].

A growing amount of Embodied Conversational Agents (ECAs) make use of Natural Language Processing (NLP) for implementing the intelligent dialogue component. NLP-based systems are of major interest in human-machine interactions for multimodal interfaces and are preferred widely for Natural and Spoken Language Understanding (NLU or SLU), Natural Language Generation (NLG) and Dialogue Processing (DP) tasks.

Complete dialogue solution as shown in figure 1, consists of many parts, each of which specializes in certain task. Automatic Speech Recognition (ASR) module is responsible for converting spoken users utterance into text. Natural Language Interpreter converts textual information to meaningful features so that Dialogue State Tracker (DST) can process this features and update the current dialogue state. DST outputs current dialogue state so that Dialogue Response Selection(DRS) module (which is trained to output a response to user utterance) can generate textual reply to user. Later this textual reply is converted to speech by text-to-speech (TTS) synthesizer. Since ASR and TTS are not related to dialogue manager directly, they can be considered as complementary modules to complete dialogue solution [14].

A dialogue manager is a component of a dialogue system, which is responsible for the transmission of information among participants in human-machine interaction. Dialogue manager can be divided into two main groups: Chatbots [16] and frame-based dialogues [17]. In chatbots, an agent is often trained to work without any knowledge of the dialogue structure mostly referred as open dialogue, whereas in frame-based dialogues; dialogue frames are structured by experts with slots and values that each slot can take. In this work, we mainly focus on chatbots and how they are trained with rule-based, sequence to sequence learning, reinforcement learning and hierarchical reinforcement learning based manager models. All these methods can be used separately or together to leverage each other's strengths. The paper gives an overview on the ECAs from the various projects; existing models applied to dialogue manager training and its advantages and drawbacks. Furthermore, it includes results on dialogue dataset with new image based method where dialogue is processed as images to train dialogue manager.

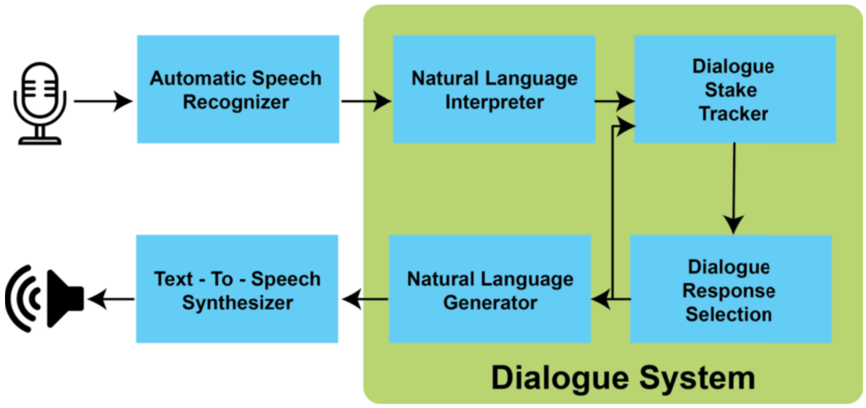


Fig. 1. Schematic diagram of Complete Dialogue System

The paper is structured into six sections. First section provides an introduction of the Dialogue system followed by Embodied Interaction. The third section describes the existing rule based and machine learning based models in dialogue systems. Later, a new proposed methodology for the dialogue system is explained. The next section consists of the preliminary results and their discussions; and the last section concludes the findings of the work.

2 Embodied Interaction Experience

Several working groups and projects have been shown that embodied interactions can help to increase the interaction with ICT systems [1]. Bickmore et al. developed a virtual laboratory to explore the longitudinal usage of a virtual exercise coach. Older adult participants interacted with an agent from their home once a day for up to 120 days. The results showed that users who interacted with an agent that used variable dialogue exercised significantly more than those interacting with an ECA with non-variable dialogue. It has been shown that the natural dialogue is an important part in natural conversation, specifically how an ECA is communicating and is structuring the dialogue with the user. In this way ECAs are capable of understanding the user's voice, gestures and emotions and plan based on this understanding multimodal utterances from propositional representations of meaning [8]. In Embodied Interactions an intelligent dialogue manager component is responsible for interpreting the users speech commands and generating the appropriate verbal and non-verbal behaviour of the agent. Studies performed in our projects: Miraculous-Life[5] [12] , CaMeLi[6] [11] and Ibi [13] have shown that the embodiment (the avatar) should express appropriate empathetic feedback (i.e., emotional facial expressions and voice intonation) based on behaviour and dialogue patterns established between older adults and human caregivers over the lifetime. In terms of dialogue communication, the avatar should allow smooth turn-taking by different modes like waiting, listening and talking. Figure 2 shows a screenshot of the Miraculous-Life animated avatar. In terms of interaction, the results indicated that the communication of older adults with the agent should be based on friendly and natural dialogues and it is critical that the agent has a clearly understandable speech output [9] and [10]. As some older adults experienced difficulties hearing and understanding the agents' speech, pronunciation and intonation has to be improved and more preferences are given to the agent which can adapt to the native languages of the user [15]. It has been observed that interaction with ECAs is still not entirely multimodal, rather more speech command based, which is also a cause of the limited flexibility of existing dialogue components. Furthermore, the most reported reason was that the ECAs did not fully reach users original expectations. Among the main reasons for this we note the mismatched expectations related with verbal communication capabilities. In fact, the majority of the users in the target group, on facing a human-like character expected a more natural interaction in terms of speech dialogues. This includes the Automatic Natural Speech Recongition but also the flexibility of the dialogue itself. Participants easily got frustrated after a few unexpected verbal behaviour by the agent. On the other hand, we noticed that users from the target groups face some challenges regarding training the interaction with the agent. This led to a higher number of repetitions than desirable. From these consideration, two main conclusions have been drawn.

(i) An ideal solution would be to have more flexibility and variety in the speech commands.

(ii) it is of utmost importance that all the interaction components run as robustly as possible, are fault-tolerant, and support repair mechanisms.

[5] http://www.miraculous-life.eu/

[6] http://www.aal-europe.eu/projects/cameli/

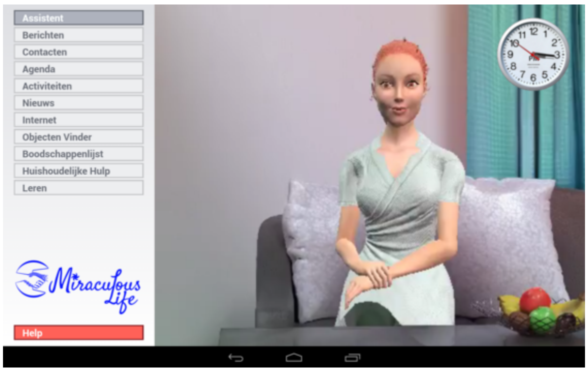


Fig. 2. The Miraculous-Life user interface including the animated avatar

3 The existing dialogue systems

The section provides the overview of the existing work on training dialogue managers which includes: Rule-based methods; Sequence-to-Sequence methods; Reinforcement learning based methods; and Hierarchical reinforcement learning methods.

3.1 Rule-based methods

The first chatbot developed by using rule-based system was ELIZA [18], which uses pattern matching based on user replies. In rule-based systems, human dialogues are modeled as set of states and dialogue manager has to choose replies for the conversation from the given set of rules [19]. This model has been used in many different applications such as restaurant booking or online psychological therapy chatbot [18]. In this method, a human, who is usually a domain expert analyzes the dialogue flow between human agents and tries to come up with predefined dialogue states and possible replies for each state based on patterns. The advantages of rule-based methods are that dialogue managers have control on selecting replies for the conversation and these selected replies from the full set of replies ensure that the user is not upset or offended, thus keeping the system consistent.

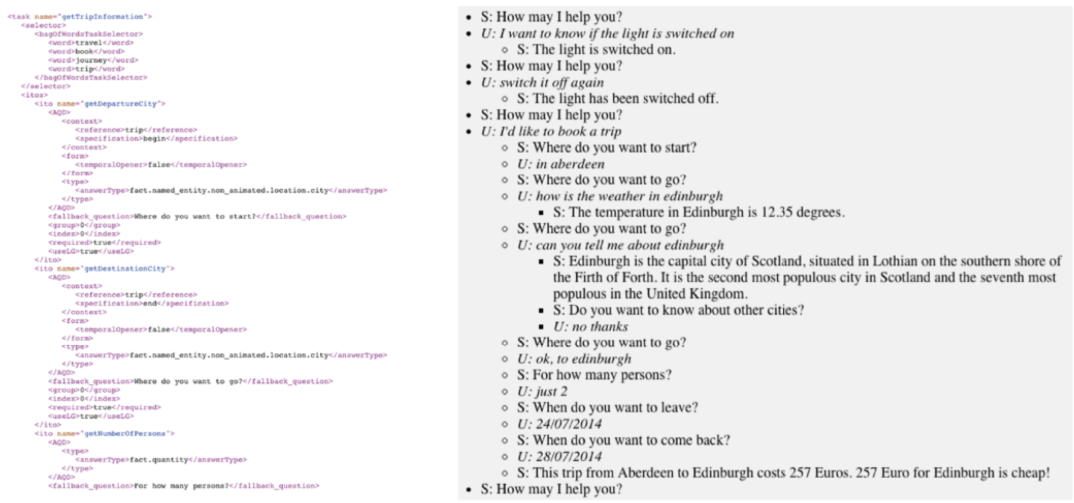


Fig. 3. Rule based system

NADIA [20] is a recent example of such system, where expert can define a dialogue structure as seen on the left side in figure 3 (in .xml format). In NADIA dialogue systems, expert can define structure with questions and answers and dialogue manager would use these hard coded rules in order to engage in a conversation. On the right side in figure 3, a sample conversation of NADIA with user can be seen, which consists of answering user's questions while making a trip reservation.

If real human dialogue flow does not have many different states and/or replies, rule-based systems usually outperform machine learning models [21]. However, most of the time in real life, human language can get very complex and it becomes very easy to run out of dialogue states designed by the expert. In such scenarios, it is not possible to use rule-based models other than giving generic answer to user such as "I don't know what you are asking", which may frustrate the user after certain amount of time.

3.2 Sequence-to-Sequence based methods

Sequence to Sequence methods or widely referred as seq2seq are deep learning methods which transform a given sequence from one domain to another. For seq2seq models usually a dictionary is defined with all the words from which a dialogue manager can choose. These dictionaries can be very large depending on the complexity of dialogue. Dialogue manager outputs a probability distribution over this dictionary of words. In end-to-end supervised learning models, in each time step dialogue manager chooses the word with highest probability conditioned on dialogue history which consists of every word chosen before given time step and in some cases with some additional information. Sequence- to-sequence methods rely on recurrent neural networks, 1-D convolutional units or the recent feed-forward neural networks. Initially, they were applied on translation tasks such as English to French and achieved astonishing results with very little or none natural language processing of sentences [22]. Later, it has been adopted on dialogue tasks, where dialogue history is treated as a sequence and mapped to another sequence which is a dialogue response [23, 24]. In translation tasks, dataset consists of pair of sentences in different languages, whereas in the dialogue tasks, there can be many more dialogue replies for the same history.

Although, seq2seq suffers from one to many mapping which has been applied to train dialogue management tasks. It has given good results on chatbots and also in domain specific tasks such as restaurant booking. Since these methods give most probable answer, it is often difficult to customize unless additional information is given and may overfit in generating certain frequent replies in the dataset. A solution could be by adding embeddings about user information in order to train network to take user preferences into account. [25].

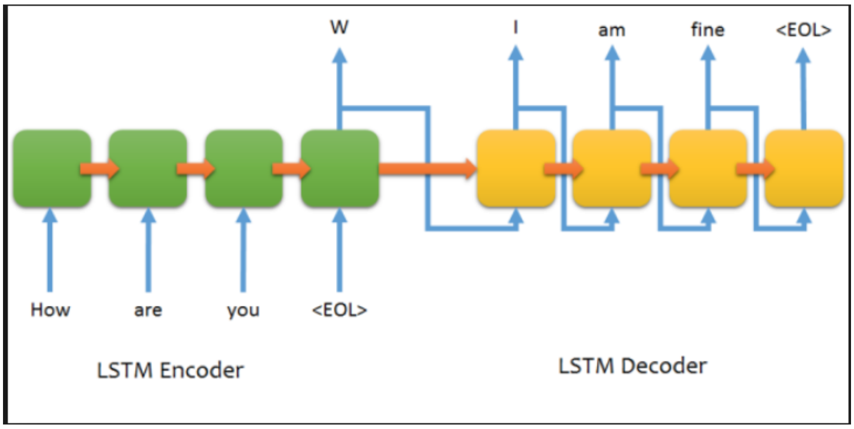


Fig. 4. Sequence-to-Sequence learning [22]

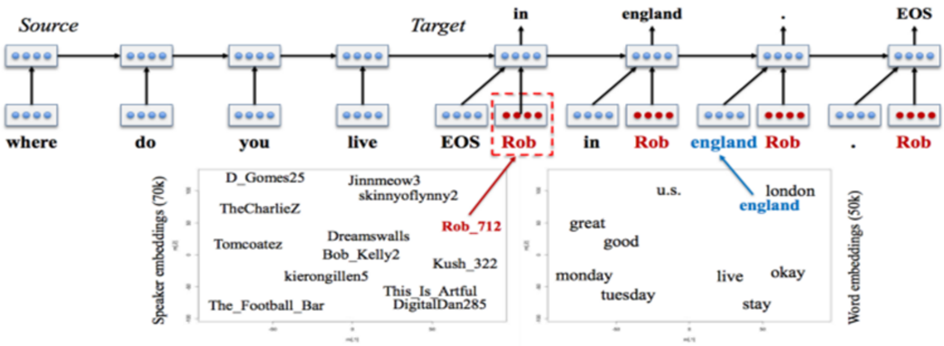


Fig. 5. Persona based neural response generation [25]

The figure 4 demonstrate seq2seq learning applied on generating response for the given history. LSTM encoder takes either full history or last reply and converts it into encoded feature vector where LSTM decoder takes this vector and outputs a possible reply condition on encoded feature vector. In figure 5, seq2seq model is changed to also accept and embedding of user, 'Rob' in this case, which allows network to generate replies based on user persona.

3.3 Reinforcement learning based methods

Reinforcement learning (RL) method is a machine learning method where agent learns to maximize notion of reward in given environment by learning which actions to take in each state [34]. Reinforcement learning methods gained popularity in recent years due to revisiting existing RL algorithms with new deep learning methods. Deep reinforcement learning algorithms outperform humans in certain game tasks such as in Atari games, chess and GO [27].

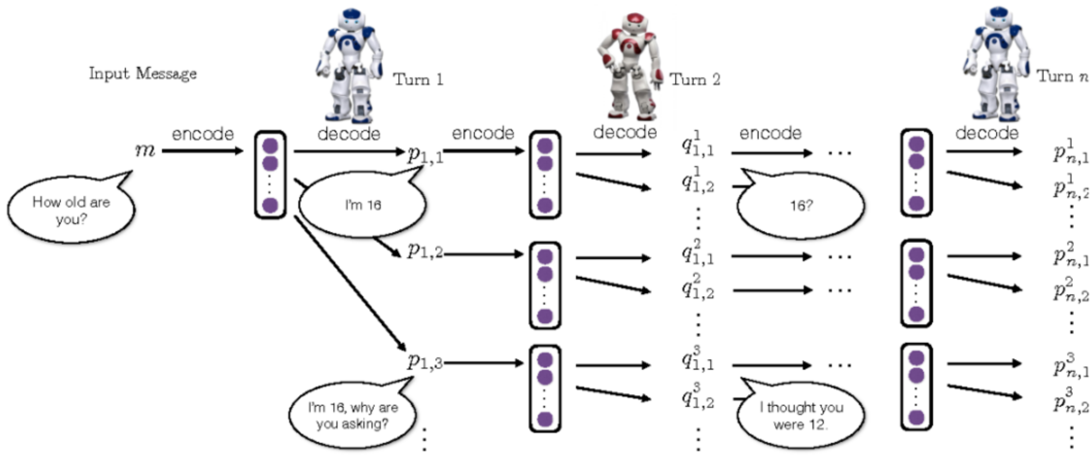


Fig. 6. Dialogue simulation between agents [29]

Dialogue management can be formulized as a Markov Decision Process which is defined as a tuple (S,A,T,R,γ). S and A represent set of states and actions; T represent the transition probabilities between s and s' given action a is taken in state s. R is the reward function such as R:S x A $\rightarrow \mathbb{R}$ which gives the reward value for an action a taken in state s. γ is the discount factor which is used to calculate the expected discounted cumulative reward $\mathbb{E}[\sum_{i\geq 1} \gamma^{i-1} R_{t+i}]$. policy π of RL agent is a mapping between states and actions. Policy of an agent can either be stochastic or deterministic and actions policy maps to can be either continuous valued actions or discreet valued actions. RL agent dialogue manager is trained in order to learn a policy which maximizes the expected discounted cumulative return.

Deep RL algorithms can be even trained playing against each other, to achieve post human performance even when the state space is very high. RL methods require a scalar reward which is more natural for dialogue tasks than supervised models. Since, it is difficult to get a supervised dialogue dataset; it would be more natural for RL dialogue manager to discover itself the dynamics of dialogue by following a reward function. However, in dialogue tasks it is often hard to design such a reward for RL methods to trained on dialogue tasks [28]. Heuristic rewards can be designed by human experts in order to train dialogue manager with RL based methods [29] or automated dialogue reply measures can also be used for training [30] [31]. The figure 6 shows a reinforcement learning framework for neural response generation by simulating dialogues between the two agents, integrating the strengths of neural seq2seq systems and reinforcement learning for dialogue.

In game setting tasks, it is also important to note that action space is very low whereas in dialogue tasks they are high dimensional. This high dimensionality is due to large number of possible sentences in any language. Such high number of possible actions increases the complexity of RL methods to apply on dialogue tasks. Often less importance is given to number of training samples needed for RL methods. For even simple data efficient methods it may require 10K of training data which is impractical to collect from humans and also impractical to simulate due to complexity and variety in human dialogues [32].

3.4 Hierarchical Reinforcement learning-based methods

Dialogue modelling can be modeled as an Markov Decision Process (MDP), however, if dialogue task is complex and natural language is used, then state and action spaces are of very high dimension. Traditional RL suffers from curse of dimensionality and different solutions are investigated by researchers. One of solutions, is using hierarchical reinforcement learning, where agent learns to abstract problem. Agent in HRL setting learns to abstract state and/or action instead of learning primitive actions for each state. The agent usually has a different hierarchical levels of policies where high level policy controls which low level policy would be chosen; while each low level policy is optimized for a different and simpler task. Furthermore, the agent makes a decision on which policy it should choose from and then follow this sub-policy which is different from higher level policy until termination condition is satisfied [33].

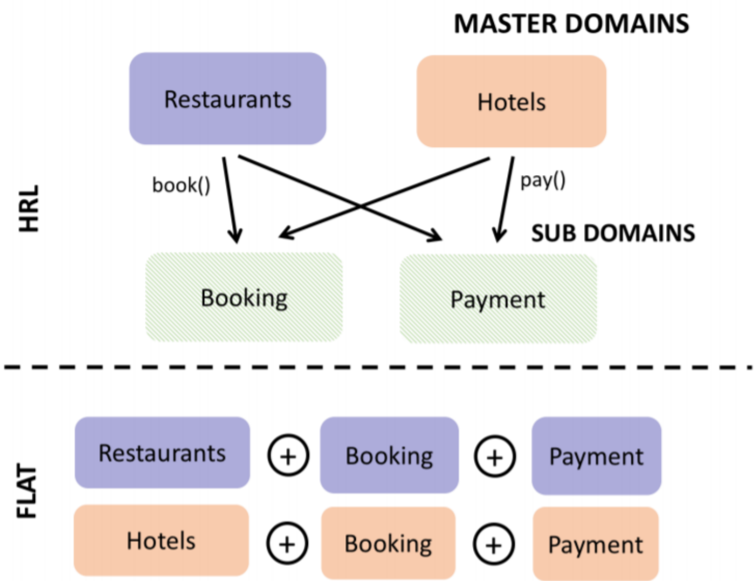


Fig. 7. Comparison of two analyzed architecture [36]

HRLs are mainly applied on special version of MDP call Semi-MDPs which allows actions to last for different amount of time between state transitions in order to model temporally-extended action set [34]. First evaluation of HRL methods on SMDP-based dialogue agent setting on a realistic environment, showed that semi-learnt or hiearchial behaviours outperformed fully-learnt agents behaviour in a travel planning domain with real users [35]. Difference between flat(traditional RL) and hierarchical RL for Booking and Payment tasks/sub-domains shared between two master domains is shown in figure 7, where flat RL learns only about primitive actions and needs to output an action for each time step wheras HRL learns hierarchical abstraction and may learn primitive as well as composite actions.

In more recent work, [36] introduced hierarchical model which is trained by hierarchical reinforcement learning with the Gaussian Process as function approximators; while [37] uses deep Q-networks as function approximators and both HRL models outperform standard RL models over composite tasks. HRL based methods on dialogue management although outperforms standard RL, have open research questions concerning handling deeper hierarchies; designing reward for lower level hierarchies; and also automatically dividing complex task into simpler subtasks which is possible with deeper hierarchies than few levels of hierarchy.

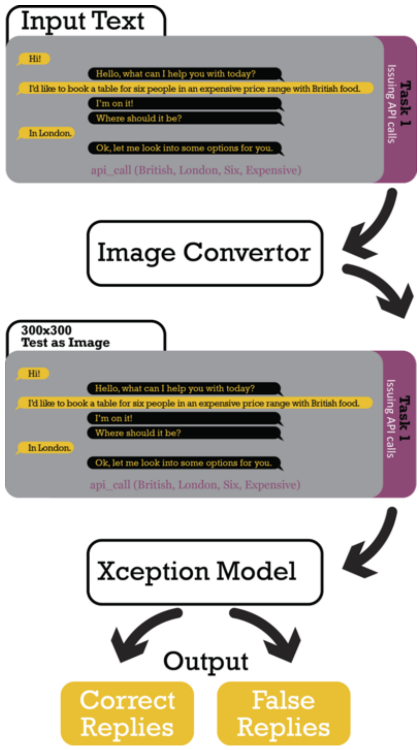


Fig. 8. Image-based methodology for dialogue manager

4 Image-based method

Recently, a novel method has been proposed to train dialogue managers [38]. In this method dialogue is converted to an image by rendering the dialogue history with an reply as an image. Processing sequential data as an image is quite common in audio processing of spectrograms. Although, audio is sequential and recurrent neural networks or 1-D convolutional neural networks are more suited, hence, they can also be processed as an image. Significant research works have been conducted using spectrograms and implementation of vision based methods in audio domain [39–41].

In dialogue domain (figure 8), same idea has been applied where manager is trained by creating positive samples combining dialogue history with correct reply and negative samples combining dialogue history with false replies samples from dataset of possible replies. As shown in figure 9, each dialgoue in this image are 300x300 black and white images and processed as an image using Xception vision model [42]. The above two blocks of the figure consists of correct dialogues where dialogue histories are appended with correct replies, given as an image. And, in lower two blocks, the same dilaogue histories are appended with false replies, which are sampled from the all possible candidate replies. It can be seen that dialogue manager task is to provide correct information to the user, who asked about restaurant address. In above two images (dialogue with correct replies) dialogue manager returns address of correct restaurant, however, in lower two images (dialogue with false replies) dialogue manager replies with irellevant replies such as offering new restaurant on left and making api call on right down. In such a setup, dialogue manager investigates dialogue only visually and learns which dialogues are correct and which dialogues are false. However, this methodology has never been applied in dialogue modeling.

Fig. 9. Dialogue as an image

5 Experiment and Results

We used Xception vision model on Facebook bAbI dataset Task 1 [21]. In this approach, we created a train set consisting of correct and wrong dialogue history and reply pairs. Later, we converted this textual dialogue to an image and processed it as a binary classification problem. For testing we rank all candidate answers for given dialogue history and choose the highest. This work is an extension of the work done [38] in Out Of Vocabulary (OOV) setting.

In natural language domain dictionary, it is important when manager does not have certain words in its dictionary than understanding the text is not possible. MemNets performance drops when tested on OOV, since language based models fails when they do not find OOV words during training. Adding human generated features can overcome this issue, however, it is very task dependent and requires human effort. On the other hand, it is different in image domain situations since there is no need for dictionary instead the network should only learn to match patterns of images. Nevertheless, other issues such as font type and font size may arise in image domain with image based methods. However, it is out of the scope of this work and therefore it is not required to have a method to adopt different font types and sizes since the aim is to learn dialogue flow.

In Facebook bAbI dataset Task 1 there are 1000 training and 1000 test dialogues. We sampled 10 negative responses for each training dialogue and we chose the correct answer from 4212 candidates answers for test dialogues. Experiments are done on a single GPU (1080ti) and took a week's time to train the model. As reported in Table 1, the results show that using image domain also helps the dialogue manager to expand to Out Of Vocabulary (OOV) dialogue tasks naturally where OOV response accuracy matches in vocabulary results.

Table 1
Facebook bAbI Dialog Task 1

Metrics	**Xception (%)**	**Memory Networks w/o Match Type (%)**	**Memory Networks w/ Match Type (%)**
Per-response Accuracy	85.7	99.9	100.0
Per-response Accuracy (Out-Of-Vocabulary)	86.6	72.3	100.0

6 Conclusion

A new image-based approach to train dialogue managers has been implemented on Facebook bAbI dialog task 1. Our method outperforms the Memory Networks (without any extra processing such as match type) for dialogue response accuracy in OOV dialog tasks using 10% of training dataset. Performance is consistent between in vocabulary and OOV. Furthermore, the experiments on English and other languages with large dataset are needed to fully capture the benefits of image based processing. It will also be interesting to perform both image-based and language-based methods simultaneously. Further work will be performed by implementing the improved dialogue components and testing it with other embodied interaction components.

Acknowledgement

This work has been funded by the European Union Horizon2020 MSCA ITN ACROSSING project (GA no. 616757). The authors would like to thank the members of the project's consortium for their valuable inputs.

References

[1] Sten Hanke, Christiana Tsiourti, Miroslav Sili, Eleni Christodoulou, *Embodied Ambient Intelligent Systems*, Ambient Intelligence and Smart Environments, Volume **20**: Recent Advances in Ambient Assisted Living Bridging Assistive Technologies, e-Health and Personalized Health Care, 65 85, (2015)

[2] P.H. Robert, A. König, H. Amieva, S. Andrieu, F. Bremond, R. Bullock, M. Ceccaldi, B. Dubois, S. Gauthier, P.-A. Kenigsberg, S. Nave, J.M. Orgogozo, J. Piano, M. Benoit, J. Touchon, B. Vellas, J. Yesavage and V. Manera, *Recommendations for the use of Serious Games in people with Alzheimers Disease*, related disorders and frailty, in Front Aging Neurosci., 2. Mrz (2014)

[3] T. W. Bickmore, L. Caruso, E. Clough-Gorr, *Acceptance and usability of a relational agent interface by urban older adults*, In CHI05 extended abstracts on Human factors in computing systems, ACM, Press, Portland, OR, USA, 1212-1215, (2005)

[4] A. H. Martin M. Morandell, B. Wöckl, S. Dittenberger und S. Fagel, *Avatars@Home: Interfacing the Smart Home for Elderly People*, in HCI and Usability for e-Inclusion: 5th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2009, Linz, Austria, November 9-10, 2009 Proceedings, Linz, Austria, Springer Berlin Heidelberg,353-365, (2009)

[5] S. Fagel, A. Hilbert, C. Mayer, M. Morandell, M. Gira und M. Petzold,*Avatar User Interfaces in an OSGi-based System for Health Care Services*, Global Health 2013, The Second International Conference on Global Health Challenges, 1-4, 17 November (2013).

[6] R. Sharma, M. Yeasin, N. Krahntoever, I. Rauschert, I. Brewer, A.M. Maceachren and K. Sengupta, *Speech-gesture driven multimodal interfaces for crisis management*, Proceedings of the IEEE, Volume 91(9), p. 13271354, (2003).

[7] T.H. Bui, *Multimodal Dialogue Management - State of the art*, Centre for Telematics and Information Technology University of Twente. Retrieved from http://eprints.ewi.utwente.nl/5708/, 2006

[8] S. Kopp, L. Gesellensetter, N. C. Krämer, I. Wachsmuth,*A Conversational Agent as Museum Guide Design and Evaluation of a Real-World Application Intelligent Virtual Agents*, Lecture Notes in Computer Science Volume **3661**, p. 329-343, (2005).

[9] M. Sili, J. Bobeth, E. Sandner, S. Hanke, S. Schwarz, C. Mayer, *Talking Faces in Lab and Field Trials - A View on Evaluation Settings and User Involvement Results of Avatar Based User Interaction Techniques in Three Ambient Assisted Living Projects*, International Conference on Human Aspects of IT for the Aged Population, pp. 134-144, Springer, Cham, (2015).

[10] D. Cereghetti, C. Wings, J. Meijers, D6.4 Pilot acceptance evaluation results, public deliverable, http://www.miraculous-life.eu/public-deliverables last seen: 10.11.2018

[11] Christiana Tsiourti, Joo Quintas, Maher Ben-Moussa, Sten Hanke, Niels Alexander Nijdam, Dimitri Konstantas, *The CaMeLi FrameworkA Multimodal Virtual Companion for Older Adults*, Proceedings of SAI Intelligent Systems Conference, pp. 196-217, Springer, Cham, 2016/9/21

[12] Sten Hanke, Emanuel Sandner, Samat Kadyrov, Andreas Stainer-Hochgatterer, *Daily life support at home through a virtual support partner*, IET Digital Library, 2016/1/1

[13] Miroslav Sili, Jan Bobeth, Emanuel Sandner, Sten Hanke, Stephanie Schwarz, Christopher C. Mayer, *Talking Faces in Lab and Field Trials - A View on Evaluation Settings and User Involvement Results of Avatar Based User Interaction* Techniques in Three Ambient Assisted Living Projects., pp. 134-144, HCI (25) 2015

[14] Serban, Iulian Vlad, Ryan Lowe, Peter Henderson, Laurent Charlin, and Joelle Pineau, *A survey of available corpora for building data-driven dialogue systems* URL:https://arxiv.org/pdf/1512.05742.pdf (2015).

[15] Singh, Deepika, Ismini Psychoula, Johannes Kropf, Sten Hanke, and Andreas Holzinger, *Users Perceptions and Attitudes Towards Smart Home Technologies*, In International Conference on Smart Homes and Health Telematics, 203-214. Springer, Cham, (2018).

[16] Shawar, Bayan Abu, and Eric Atwell, *Chatbots: are they really useful?*,In Ldv Forum, vol. **22**, no. **1**,(2007),29-49.

[17] Masche, Julia, and Nguyen-Thinh Le, *A Review of Technologies for Conversational Systems*, International Conference on Computer Science, Applied Mathematics and Applications. Springer, Cham, (2017).

[18] Weizenbaum, Joseph, *ELIZAa computer program for the study of natural language communication between man and machine* Communications of the ACM **9.1** (1966): 36-45.

[19] Webb, Nick, *Rule-based dialogue management systems* In Proceedings,(2000),164-169.

[20] Berg, Markus M,*NADIA: A Simplified Approach Towards the Development of Natural Dialogue Systems*, In International Conference on Applications of Natural Language to Information Systems, pp. 144-150. Springer, Cham, (2015).

[21] Bordes, Antoine, Y-Lan Boureau, and Jason Weston *Learning end-to-end goal-oriented dialog*, URL:https://arxiv.org/abs/1605.07683

[22] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le, *Sequence to sequence learning with neural networks* Advances in neural information processing systems, (2014).

[23] Vinyals, Oriol, and Quoc Le,*A neural conversational model*,URL:https://arxiv.org/pdf/1506.05869.pdf (2015).

[24] Sordoni, Alessandro, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Margaret Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan, *A neural network approach to context-sensitive generation of conversational responses*,URL:https://arxiv.org/pdf/1506.06714.pdf(2015)

[25] Li, Jiwei, Michel Galley, Chris Brockett, Georgios P. Spithourakis, Jianfeng Gao, and Bill Dolan, *A persona-based neural conversation model*,URL:https://arxiv.org/pdf/1603.06155.pdf(2016)

[26] Sutton, Richard S., and Andrew G. Barto, *Reinforcement learning: An introduction*(2011).

[27] Mnih, Volodymyr, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves et al,*Human-level control through deep reinforcement learning* Nature 518, no. **7540** (2015), 529.

[28] Liu, Chia-Wei, Ryan Lowe, Iulian V. Serban, Michael Noseworthy, Laurent Charlin, and Joelle Pineau, *How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation*,URL:https://arxiv.org/pdf/1603.08023.pdf (2016).

[29] Li, Jiwei, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, and Dan Jurafsky, *Deep reinforcement learning for dialogue generation*, URL:https://arxiv.org/pdf/1606.01541.pdf(2016).

[30] Merdivan, Erinc, Mohammad Reza Loghmani, and Matthieu Geist, *Reconstruct and Crush Network*, Advances in Neural Information Processing Systems, (2017).

[31] Lowe, Ryan, Michael Noseworthy, Iulian V. Serban, Nicolas Angelard-Gontier, Yoshua Bengio, and Joelle Pineau, *Towards an automatic Turing test: Learning to evaluate dialogue responses*,URL:https://arxiv.org/pdf/1708.07149(2017).

[32] Pietquin, Olivier, Matthieu Geist, Senthilkumar Chandramohan, and Herv Frezza-Buet, *Sample-efficient batch reinforcement learning for dialogue management optimization*ACM Transactions on Speech and Language Processing (TSLP) 7, no. **3** (2011),7.

[33] Barto, Andrew G., and Sridhar Mahadevan,*Recent advances in hierarchical reinforcement learning*, Discrete event dynamic systems 13 **1-2** (2003),41-77.

[34] Sutton, Richard S., Doina Precup, and Satinder Singh. *Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning* Artificial intelligence 112, no. **1-2** (1999), 181-211.

[35] Cuayhuitl, Heriberto, Steve Renals, Oliver Lemon, and Hiroshi Shimodaira, *Evaluation of a hierarchical reinforcement learning spoken dialogue system* Computer Speech and Language 24, no. **2** (2010), 395-429.

[36] Budzianowski, Pawe, Stefan Ultes, Pei-Hao Su, Nikola Mrki, Tsung-Hsien Wen, Inigo Casanueva, Lina Rojas-Barahona, and Milica Gai, *Sub-domain modelling for dialogue management with hierarchical reinforcement learning* URL:https://arxiv.org/pdf/1706.06210.pdf,(2017).

[37] Peng, Baolin, Xiujun Li, Lihong Li, Jianfeng Gao, Asli Celikyilmaz, Sungjin Lee, and Kam-Fai Wong,*Composite task-completion dialogue policy learning via hierarchical deep reinforcement learning*,URL:https://arxiv.org/pdf/1704.03084.pdf, (2017).

[38] Merdivan, Erinc, Anastasios Vafeiadis, Dimitrios Kalatzis, Sten Henke, Johannes Kropf, Konstantinos Votis, Dimitrios Giakoumis et al, *Image-based Natural Language Understanding Using 2D Convolutional Neural Networks*,URL:https://arxiv.org/pdf/1810.10401.pdf,(2018).

[39] Abdel-Hamid, Ossama, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu, *Convolutional neural networks for speech recognition*, IEEE/ACM Transactions on audio, speech, and language processing 22, no. **10** (2014),1533-1545.

[40] Phan, Huy, Lars Hertel, Marco Maass, and Alfred Mertins, *Robust audio event recognition with 1-max pooling convolutional neural networks*,URL:https://arxiv.org/pdf/1604.06338.pdf,(2016).

[41] Hershey, Shawn, Sourish Chaudhuri, Daniel PW Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal et al,*CNN architectures for large-scale audio classification* In Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on, 131-135. IEEE, (2017).

[42] Chollet, Franois,*Xception: Deep learning with depthwise separable convolutions*,URL:https://arxiv.org/pdf/1610.02357.pdf,(2017).

CHAPTER 8

Conclusion and Future Work

Today, technology is incredibly prevalent in our lives, especially in the home environment due to rapid development in smart devices, voice recognition systems, home automation, and environment control. These technologies help mainly older people in living an active and independent life. Aging leads to many health problems such as sedentariness, inability to perform daily living activities, cognitive diseases, dementia, and sometimes physical injuries. In order to monitor the health and maintain a healthy life of older people, the concept of the smart home has emerged tremendously due to its varied applications. However, there are barriers and challenges in the wide adoption of smart home by the end-users. This thesis aims to address these challenges and proposed an open smart home framework that is scalable, extensible, and adaptable to various real-life applications. In this thesis, we first present a conceptual framework of a smart home system which allows integration and communication between all the major modules i.e activity recognition; privacy-preserving; and dialogue systems in a HOMER middleware. HOMER is an open-source platform based on the Apache Karaf OSGi framework and enables modularity by encapsulating its functionalities in terms of OSGi bundles. The components of the framework are in form of OSGi bundles which can also be installed remotely and updated without rebooting the whole system. The bundles of the framework are interconnected with the Event Admin component which enables the communication between the bundles based on a subscribe and publish mechanism. In the following chapters of this thesis, the works have been performed in developing each module of the smart home system using machine learning/deep learning algorithms, together with user studies to identify end users' needs and concerns in a smart home system.

The second chapter of the thesis presents different user studies which identify users' perception towards AAL technologies, their requirements, and concerns in a smart home. The first study was conducted at Zuyderland hospital and nursing home with older people, caregivers, and professionals of the care giving organization. The participants showed a lot of interest in the smart home system and but the needs and problems were varied for each individual. The older people who were living by themselves were least interested in home automation whereas people who were suffering from chronic diseases or any other illness find it beneficial. All the residents and professionals agreed in having an activity monitoring system, however robots and cameras were not preferred by any of the residents. The second user study of the thesis is an online survey with 234 participants and focuses on understanding the attitudes and perceptions of future smart home users (from age group under 18-70 years). Various aspects are investigated in the study such as users' attitudes in-home, outdoor activities, and views regarding personal AI assistants; facilities participants would want in their smart home; perceived benefits and drawbacks of smart home; attitudes towards home monitoring; and data sharing. The findings of the study provide insights into the needs and concerns of the upcoming smart home users and would help in developing an acceptable smart home solution with minimal or no concerns. Finally, the third user study aims to understand users' privacy concerns in IoT based applications. In this study, both interviews with the elderly and an online survey were conducted. The findings of the study show that the users who are aware of privacy risks would desire to have more fine-grained control over the use of their personal data, but are willing to share the data when there are benefits and clear terms of use. The results of the studies can be used to provide a basis for building privacy-preserving modules with user preferences and data control in smart home and AAL environments in general.

Since activity monitoring is perceived as a crucial component for continuous health monitoring in the smart home system. Therefore, the next part of the thesis presented the work in developing and improving human activity recognition. We developed different deep learning neural networks on smart home benchmark datasets for a single resident and multiple resident activity recognition. Results show that deep learning models outperformed the traditional machine learning models on raw sensor datasets and does not require preprocessing and feature engineering in the datasets. Furthermore, we investigated model performance in different house and user settings, which is important for the models to be applied widely. In multiple resident activity recognition, we explored and evaluated different class imbalance approaches on deep learning models and showed the importance of evaluation metrics in activity recognition. A model may achieve very high accuracy in some cases due to class imbalance but the actual performance of the model can only be shown through balanced accuracy, exact match ratio (EMR), and F1-score. Among all the class imbalance approaches, cost-sensitive learning improves the performance of the model in comparison to sampling methods.

Another important component of a smart home framework is the privacy-preserving module. Based on the user studies, monitoring of private activities and data sharing with stakeholders or third parties are perceived as one of the major drawbacks of smart home technologies. Therefore, we developed a new privacy-preserving mechanism to transform data before sharing them with other services or stakeholders. These transformations aim to remove patterns or obfuscate sensitive data that could identify a user or infer private activities and at the same time keeping the utility of the data needed for specific services. The developed model is based on the deep learning encoder-decoder model and creates different views for different users such that only particular information will be shared as per user preferences. The proposed mechanism is easily generalizable and scalable as it can be applied to data of new users or add different transformations for new services with minimal cost.

Lastly, this thesis provides a detailed overview of the existing methodologies for training dialogue manager and proposed a new image based method where dialogue can be processed as an image instead of text. Moreover, a new dialogue dataset is collected which addresses two major issues: first is diversity in dialogue replies, although the same dialogue context can have various good replies, but the existing datasets consist of only one reply for each dialogue context. The second issue is the lack of benchmark dialogue metrics, therefore in this work, we provide benchmark performances with the deep learning models such as Hierarchical Attention Network (HAN), Bidirectional encoder representations from Transformers (BERT), and word overlap metrics (BLEU, ROUGE, and METEOR) to evaluate the quality of dialogue replies for the given context. Such benchmarks find utility in many other applications such as comparing different dialogue generator models, training new algorithms, or designing new dialogue managers.

Often in research, the solution to one problem may point to even more interesting solutions and ideas or new research problems. During this research, several open issues and challenges have been identified that can be improved further. Firstly, the sample size of online user studies is relatively small, especially from Africa, South America, and Australia. It would be better to have more participants and interviews with older people to have a general opinion which could be adopted by the research community for developing assistive smart solutions. However, the web link to the user study is still open for voluntary participants on social websites. Next, in human activity recognition, we focused on supervised and deep learning algorithms. During the studies, we also collected our own multiple occupancy dataset in the AIT office environment. However, there were many issues such as sensor faults, false sensor readings, and manual labeling of activities which was a very time consuming and costly process. So, the collection of the real dataset was not successful. A possible extension in activity recognition work could be the use of unsupervised methods or certain deep learning pre-trained methods that are successful in NLP tasks, which may also benefit in activity recognition model. Also, a further upcoming aspect is explainable AI and interpretable models that can help in understanding the decision made by the network. In privacy-preserving deep encoder-decoder mechanism, model performance can be evaluated and compare on more datasets including image and video data. In addition, the proposed model does not integrate encryption. One possible idea could be the use of an encryption key as an input to the encoder and decoder and only encode or decode the data if the correct key is used. In the thesis, we collected an open domain dialogue dataset and provided preliminary results with supervised models. In future work, it would be interesting to investigate the diversity of replies in different dialogue settings and how humans react to diverse replies in task-oriented dialogues. Also, improvement is required in investigating new algorithms for dialogues training as it is not very practical to train a separate model for each different dialogue dataset. Finally, work is needed to implement the proposed smart framework by integrating all the modules and evaluate it for different real-world applications.

Bibliography

[1] (2020) Un desa. Accessed: 2020-10-15. [Online]. Available: https://www.un.org/development/desa/pd/

[2] D. J. Cook, "How smart is your home?" *Science*, vol. 335, no. 6076, pp. 1579–1581, 2012.

[3] S. Hanke, H. Meinedo, D. Portugal, M. Belk, J. Quintas, E. Christodoulou, M. Sili, M. S. Dias, and G. Samaras, "Cogniwin–a virtual assistance system for older adults at work," in *International Conference on Human Aspects of IT for the Aged Population*. Springer, 2015, pp. 257–268.

[4] A. Costa, V. Julián, and P. Novais, "Advances and trends for the development of ambient-assisted living platforms," *Expert Systems*, vol. 34, no. 2, p. e12163, 2017.

[5] (2019) Agewell. Accessed: 2021-01-15. [Online]. Available: http://www.aal-europe.eu/projects/agewell/

[6] (2016) Active@home. Accessed: 2021-01-15. [Online]. Available: http://www.aal-europe.eu/projects/activehome/

[7] (2016) Acrossing project. Accessed: 2021-01-15. [Online]. Available: https://cordis.europa.eu/project/id/676157

[8] (2019) Smart home market size, share & industry analysis. Accessed: 2021-01-15. [Online]. Available: https://www.fortunebusinessinsights.com/industry-reports/smart-home-market-101900

[9] D. Singh, I. Psychoula, E. Merdivan, J. Kropf, S. Hanke, E. Sandner, L. Chen, and A. Holzinger, "Privacy-enabled smart home framework with voice assistant," in *Smart Assisted Living*. Springer, 2020, pp. 321–339.

[10] D. Singh, J. Kropf, S. Hanke, and A. Holzinger, "Ambient assisted living technologies from the perspectives of older people and professionals," in *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, 2017, pp. 255–266.

[11] D. Singh, I. Psychoula, J. Kropf, S. Hanke, and A. Holzinger, "Users' perceptions and attitudes towards smart home technologies," in *International Conference on Smart Homes and Health Telematics*. Springer, 2018, pp. 203–214.

[12] I. Psychoula, D. Singh, L. Chen, F. Chen, A. Holzinger, and H. Ning, "Users' privacy concerns in iot based applications," in *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 2018, pp. 1887–1894.

[13] D. Singh, E. Merdivan, I. Psychoula, J. Kropf, S. Hanke, M. Geist, and A. Holzinger, "Human activity recognition using recurrent neural networks," in *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, 2017, pp. 267–274.

[14] D. Singh, E. Merdivan, S. Hanke, J. Kropf, M. Geist, and A. Holzinger, "Convolutional and recurrent neural networks for activity recognition in smart environment," in *Towards integrative machine learning and knowledge extraction*. Springer, 2017, pp. 194–205.

[15] I. Psychoula, E. Merdivan, D. Singh, L. Chen, F. Chen, S. Hanke, J. Kropf, A. Holzinger, and M. Geist, "A deep learning approach for privacy preservation in

assisted living," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 2018, pp. 710–715.

[16] E. Merdivan, D. Singh, S. Hanke, J. Kropf, A. Holzinger, and M. Geist, "Human annotated dialogues dataset for natural conversational agents," *Applied Sciences*, vol. 10, no. 3, p. 762, 2020.

[17] E. Merdivan, D. Singh, S. Hanke, and A. Holzinger, "Dialogue systems for intelligent human computer interactions," *Electronic Notes in Theoretical Computer Science*, vol. 343, pp. 57–71, 2019.

[18] A. Muñoz, J. C. Augusto, A. Villa, and J. A. Botía, "Design and evaluation of an ambient assisted living system based on an argumentative multi-agent system," *Personal and Ubiquitous Computing*, vol. 15, no. 4, pp. 377–387, 2011.

[19] A. Sixsmith, S. Mueller, F. Lull, M. Klein, I. Bierhoff, S. Delaney, P. Byrne, S. Sproll, R. Savage, and E. Avatangelou, "A user-driven approach to developing ambient assisted living systems for older people: The soprano project," in *Intelligent technologies for bridging the grey digital divide*. IGI Global, 2011, pp. 30–45.

[20] M. Pieper, M. Antona, and U. Cortés, "Introduction to the special theme ambient assisted living," *Ercim News*, vol. 87, pp. 18–19, 2011.

[21] (2012) Ambient assisted living joint programme. Accessed: 2020-12-24. [Online]. Available: http://www.aal-europe.eu/

[22] S. Blackman, C. Matlo, C. Bobrovitskiy, A. Waldoch, M. L. Fang, P. Jackson, A. Mihailidis, L. Nygård, A. Astell, and A. Sixsmith, "Ambient assisted living technologies for aging well: a scoping review," *Journal of Intelligent Systems*, vol. 25, no. 1, pp. 55–69, 2016.

[23] Washington state university, wsu casas. Accessed: 2020-12-24. [Online]. Available: http://casas.wsu.edu/

[24] (2011) Dem@care, dementia ambient care: multi-sensing monitoring for intelligent remote management and decision support. Accessed: 2020-12-24. [Online]. Available: https://www.demcare.eu/

[25] Companionable, integrated cognitive assistive and domotic companion robot systems for ability and security. Accessed: 2020-12-24. [Online]. Available: https://www.smart-homes.nl/en/project/companionable-1/

[26] The european ambient assisted living innovation alliance, alliance2. Accessed: 2020-12-24. [Online]. Available: http://www.aaliance.eu/

[27] F. Furfari, M.-R. Tazari, and V. Eisemberg, "universaal: an open platform and reference specification for building aal systems," *ERCIM News*, vol. 87, pp. 44–45, 2011.

[28] M. C. Mozer, "The neural network house: An environment hat adapts to its inhabitants," in *Proc. AAAI Spring Symp. Intelligent Environments*, vol. 58, 1998.

[29] S. K. Das, D. J. Cook, A. Battacharya, E. O. Heierman, and T.-Y. Lin, "The role of prediction algorithms in the mavhome smart home architecture," *IEEE Wireless Communications*, vol. 9, no. 6, pp. 77–84, 2002.

[30] S. Helal, W. Mann, H. El-Zabadani, J. King, Y. Kaddoura, and E. Jansen, "The gator tech smart house: A programmable pervasive space," *Computer*, vol. 38, no. 3, pp. 50–60, 2005.

[31] J. Doyle, A. Kealy, J. Loane, L. Walsh, B. O'Mullane, C. Flynn, A. Macfarlane, B. Bortz, R. B. Knapp, and R. Bond, "An integrated home-based self-management system to support the wellbeing of older adults," *Journal of ambient intelligence and smart environments*, vol. 6, no. 4, pp. 359–383, 2014.

[32] G. P. Moschis, "Marketing to older adults: an updated overview of present knowledge and practice," *Journal of Consumer Marketing*, 2003.

[33] T. Heart and E. Kalderon, "Older adults: are they ready to adopt health-related ict?" *International journal of medical informatics*, vol. 82, no. 11, pp. e209–e231, 2013.

[34] R. Steele, A. Lo, C. Secombe, and Y. K. Wong, "Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare," *International journal of medical informatics*, vol. 78, no. 12, pp. 788–801, 2009.

[35] A.-S. Melenhorst, W. A. Rogers, and E. C. Caylor, "The use of communication technologies by older adults: exploring the benefits from the user's perspective," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 45, no. 3. SAGE Publications Sage CA: Los Angeles, CA, 2001, pp. 221–225.

[36] G. Demiris, M. J. Rantz, M. A. Aud, K. D. Marek, H. W. Tyrer, M. Skubic, and A. A. Hussam, "Older adults' attitudes towards and perceptions of 'smart home'technologies: a pilot study," *Medical informatics and the Internet in medicine*, vol. 29, no. 2, pp. 87–94, 2004.

[37] S. J. Czaja, W. R. Boot, N. Charness, and W. A. Rogers, *Designing for older adults: Principles and creative human factors approaches*. CRC press, 2019.

[38] A. Morris, J. Goodman, and H. Brading, "Internet use and non-use: views of older users," *Universal access in the information society*, vol. 6, no. 1, pp. 43–57, 2007.

[39] V. G. Sanchez, C. F. Pfeiffer, and N.-O. Skeie, "A review of smart house analysis methods for assisting older people living alone," *Journal of Sensor and Actuator Networks*, vol. 6, no. 3, p. 11, 2017.

[40] J. J. J. van Dongen, I. G. J. Habets, A. Beurskens, and M. A. van Bokhoven, "Successful participation of patients in interprofessional team meetings: A qualitative study," *Health Expectations*, vol. 20, no. 4, pp. 724–733, 2017.

[41] G. Demiris, D. P. Oliver, G. Dickey, M. Skubic, and M. Rantz, "Findings from a participatory evaluation of a smart home application for older adults," *Technology and health care*, vol. 16, no. 2, pp. 111–118, 2008.

[42] P. Visutsak and M. Daoudi, "The smart home for the elderly: Perceptions, technologies and psychological accessibilities: The requirements analysis for the elderly in thailand," in *2017 XXVI International Conference on Information, Communication and Automation Technologies (ICAT)*. IEEE, 2017, pp. 1–6.

[43] A. Holzinger, K. Schaupp, and W. Eder-Halbedl, "An investigation on acceptance of ubiquitous devices for the elderly in a geriatric hospital environment: using the example of person tracking," in *International Conference on Computers for Handicapped Persons*. Springer, 2008, pp. 22–29.

[44] L. Chen and C. D. Nugent, "Sensor-based activity recognition review," in *Human Activity Recognition and Behaviour Analysis*. Springer, 2019, pp. 23–47.

[45] B. Schiele, "Wearable sensing to annotate meeting recordings," in *Proceedings of the 6th IEEE International Symposium on Wearable Computers*, 2002, p. 186.

[46] P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, and T. Starner, "Recognizing workshop activity using body worn microphones and accelerometers," in *International conference on pervasive computing*. Springer, 2004, pp. 18–32.

[47] D. J. Patterson, L. Liao, D. Fox, and H. Kautz, "Inferring high-level behavior from low-level sensors," in *International Conference on Ubiquitous Computing*. Springer, 2003, pp. 73–89.

[48] D. Ashbrook and T. Starner, "Using gps to learn significant locations and predict movement across multiple users," *Personal and Ubiquitous computing*, vol. 7, no. 5, pp. 275–286, 2003.

[49] M. Sung, R. DeVaul, S. Jimenez, J. Gips, and A. Pentland, "Shiver motion and core body temperature classification for wearable soldier health monitoring systems," in *Eighth international symposium on wearable computers*, vol. 1. IEEE, 2004, pp. 192–193.

[50] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *International conference on pervasive computing*. Springer, 2004, pp. 1–17.

[51] N. Ravi, N. Dandekar, and P. Mysore, "P., and littman," *ML," Activity recognition from accelerometer data." AAAI*, vol. 5, pp. 1541–1546, 2005.

[52] D. L. Vail, M. M. Veloso, and J. D. Lafferty, "Conditional random fields for activity recognition," in *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, 2007, pp. 1–8.

[53] M. Richardson and P. Domingos, "Markov logic networks," *Machine learning*, vol. 62, no. 1-2, pp. 107–136, 2006.

[54] D. Riboni, C. Bettini, G. Civitarese, Z. H. Janjua, and R. Helaoui, "Fine-grained recognition of abnormal behaviors for early detection of mild cognitive impairment," in *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2015, pp. 149–154.

[55] J. Ye, G. Stevenson, and S. Dobson, "Usmart: An unsupervised semantic mining activity recognition technique," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 4, no. 4, pp. 1–27, 2014.

[56] Y. Bengio, "Deep learning of representations: Looking forward," in *International Conference on Statistical Language and Speech Processing*. Springer, 2013, pp. 1–37.

[57] Q. Yang, "Activity recognition: linking low-level sensors to high-level intelligence." in *IJCAI*, vol. 9, 2009, pp. 20–25.

[58] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[59] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[60] R. Xu, Q. Zeng, L. Zhu, H. Chi, X. Du, and M. Guizani, "Privacy leakage in smart homes and its mitigation: Ifttt as a case study," *IEEE Access*, vol. 7, pp. 63 457–63 471, 2019.

[61] C. Wilson, T. Hargreaves, and R. Hauxwell-Baldwin, "Benefits and risks of smart home technologies," *Energy Policy*, vol. 103, pp. 72–83, 2017.

[62] D. Barua, J. Kay, and C. Paris, "Viewing and controlling personal sensor data: what do users want?" in *International Conference on Persuasive Technology*. Springer, 2013, pp. 15–26.

[63] P. Klasnja, S. Consolvo, T. Choudhury, R. Beckwith, and J. Hightower, "Exploring privacy concerns about personal sensing," in *International Conference on Pervasive Computing*. Springer, 2009, pp. 176–183.

[64] H. Lee and A. Kobsa, "Understanding user privacy in internet of things environments," in *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*. IEEE, 2016, pp. 407–412.

[65] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[66] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, pp. 3–es, 2007.

[67] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *2007 IEEE 23rd International Conference on Data Engineering*. IEEE, 2007, pp. 106–115.

[68] F. K. Dankar and K. El Emam, "The application of differential privacy to health data," in *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, 2012, pp. 158–166.

[69] C. Liu, S. Chakraborty, and P. Mittal, "Dependence makes you vulnberable: Differential privacy under dependent tuples." in *NDSS*, vol. 16, 2016, pp. 21–24.

[70] Z. Sun, Y. Wang, M. Shu, R. Liu, and H. Zhao, "Differential privacy for data and model publishing of medical data," *IEEE Access*, vol. 7, pp. 152 103–152 114, 2019.

[71] R. Cramer, I. B. Damgård, and J. B. Nielsen, *Secure multiparty computation: an information-theoretic approach.* Cambridge University Press, 2015.

[72] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, 2015, pp. 1310–1321.

[73] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308–318.

[74] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, and K. Talwar, "Semi-supervised knowledge transfer for deep learning from private training data," *arXiv preprint arXiv:1610.05755*, 2016.

[75] P. Xie, M. Bilenko, T. Finley, R. Gilad-Bachrach, K. Lauter, and M. Naehrig, "Crypto-nets: Neural networks over encrypted data," *arXiv preprint arXiv:1412.6181*, 2014.

[76] J. W. Bos, K. Lauter, and M. Naehrig, "Private predictive analysis on encrypted medical data," *Journal of biomedical informatics*, vol. 50, pp. 234–243, 2014.

[77] J. Gao, M. Galley, and L. Li, "Neural approaches to conversational ai," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 1371–1374.

[78] S. Mallios and N. Bourbakis, “A survey on human machine dialogue systems,” in *2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE, 2016, pp. 1–7.

[79] I. V. Serban, R. Lowe, P. Henderson, L. Charlin, and J. Pineau, “A survey of available corpora for building data-driven dialogue systems,” *arXiv preprint arXiv:1512.05742*, 2015.

[80] J. Weizenbaum, “Eliza—a computer program for the study of natural language communication between man and machine,” *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[81] K. M. Colby, “Modeling a paranoid mind,” *Behavioral and Brain Sciences*, vol. 4, no. 4, pp. 515–534, 1981.

[82] J. L. Hutchens and M. D. Alder, “Introducing megahal,” in *New Methods in Language Processing and Computational Natural Language Learning*, 1998.

[83] O. Vinyals and Q. Le, “A neural conversational model,” *arXiv preprint arXiv:1506.05869*, 2015.

[84] R. Lowe, N. Pow, I. Serban, L. Charlin, and J. Pineau, “Incorporating unstructured textual knowledge sources into neural dialogue systems,” in *Neural information processing systems workshop on machine learning for spoken language understanding*, 2015.

[85] I. Serban, A. Sordoni, R. Lowe, L. Charlin, J. Pineau, A. Courville, and Y. Bengio, “A hierarchical latent variable encoder-decoder model for generating dialogues,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[86] J. D. Williams and G. Zweig, “End-to-end lstm-based dialog control optimized with supervised and reinforcement learning,” *arXiv preprint arXiv:1606.01269*, 2016.

[87] Z. Yan, N. Duan, J. Bao, P. Chen, M. Zhou, Z. Li, and J. Zhou, "Docchat: An information retrieval approach for chatbot engines using unstructured documents," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, pp. 516–525.

[88] Y. Wu, Z. Li, W. Wu, and M. Zhou, "Response selection with topic clues for retrieval-based chatbots," *Neurocomputing*, vol. 316, pp. 251–261, 2018.

[89] S. McGlashan, N. M. Fraser, N. Gilbert, E. Bilange, P. Heisterkamp, and N. J. Youd, "Dialogue management for telephone information systems," in *Third Conference on Applied Natural Language Processing*, 1992, pp. 245–246.

[90] A. Simpson and N. M. Eraser, "Black box and glass box evaluation of the sundial system," in *Third European Conference on Speech Communication and Technology*, 1993.

[91] A. L. Gorin, G. Riccardi, and J. H. Wright, "How may i help you?" *Speech communication*, vol. 23, no. 1-2, pp. 113–127, 1997.

[92] H. Chen, X. Liu, D. Yin, and J. Tang, "A survey on dialogue systems: Recent advances and new frontiers," *Acm Sigkdd Explorations Newsletter*, vol. 19, no. 2, pp. 25–35, 2017.

[93] A. Celikyilmaz, L. Deng, and D. Hakkani-Tür, "Deep learning in spoken and text-based dialog systems," in *Deep Learning in Natural Language Processing*. Springer, 2018, pp. 49–78.

[94] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, "Convolutional neural networks for human activity recognition using mobile sensors," in *6th International Conference on Mobile Computing, Applications and Services*. IEEE, 2014, pp. 197–205.

[95] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," *arXiv preprint arXiv:1604.08880*, 2016.

[96] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.

Appendix A

Interview Questions

Questionnaire to the Elderly

Date of Interview: _____/_____/_____ (DD/MM/YYYY)

Place: ___________

Identification code: ______ (To be filled by interviewer)

Questions:

1. General Information:

Name: ________________

Date of Birth: _____/_____/_____ (DD/MM/YYYY)

Sex: M ☐ F ☐

Marital status:

☐ Married (living with husband/wife)

☐ Full time relationship

☐ Separated (married, but living separately)

☐ Divorced

☐ Single

☐ Widowed

☐ Do not want to answer the question

Please provide the highest education you gained?

☐ Masters

☐ Bachelors

☐ Senior Secondary

☐ Secondary

☐ Primary

☐ No education

☐ Do not want to answer the question

Please provide your work status?

- ☐ Retired then last work you did ________________
- ☐ Still working full time
- ☐ Still working part time
- ☐ Not employed
- ☐ Others (please specify) _______________

What are your personal income sources?

- ☐ Work
- ☐ Pension
- ☐ Help from relatives
- ☐ Help from welfare society
- ☐ Others (please specify) _______________

What are your hobbies: _______________________________

2. Home Activities

With whom do you live?

- ☐ No one
- ☐ Spouse/Partner
- ☐ Children
- ☐ Grandchildren
- ☐ Caregivers
- ☐ Brothers/Sisters
- ☐ Others (please specify) _______________

What do you prefer to do in your free time?

- ☐ Sleep
- ☐ Go outside for walk

- ☐ Watch TV/Movies
- ☐ Listen Music
- ☐ Gardening
- ☐ Read books or novels
- ☐ go out for shopping
- ☐ Cook
- ☐ Others (please specify) _______________

Which activity you like to do the most?

- ☐ Watching TV
- ☐ Cooking
- ☐ Sleeping
- ☐ Listening music
- ☐ Gardening
- ☐ Reading books, newspapers, etc.
- ☐ Playing sports
- ☐ Others (please specify) _______________

In which activity you need help from your caregiver/partner?

- ☐ Housekeeping
- ☐ Get dressed/undressed/showering
- ☐ Grooming
- ☐ Preparing meals
- ☐ Gardening
- ☐ Visiting the hospital
- ☐ House hold activities like turning off/on lights of rooms,
- ☐ Others (please specify) _______________

Which activity you always forgot inside your home?

- ☐ Turning stove on/off
- ☐ Lights on/off
- ☐ Bathroom/Kitchen water tap
- ☐ Door lock/unlock
- ☐ Your keys/mobile phone
- ☐ Taking medicines
- ☐ Others (please specify) _______________

Do you feel safe when alone in your home?

- ☐ Yes seldom
- ☐ Yes, very often
- ☐ Yes, sometimes
- ☐ Do not feel safe, only during the nights
- ☐ Not feel safe, only during day time
- ☐ Not at all
- ☐ Don’t know

Would you feel safer if you would have a home security system?

- ☐ Yes, always
- ☐ Yes, sometimes
- ☐ Yes, very often
- ☐ Seldom
- ☐ Not at all
- ☐ Don’t know

Would you be interested in having a system which can detect people visiting at your home when you are not present?

- ☐ Yes, always
- ☐ Yes, sometimes
- ☐ Yes, very often
- ☐ Seldom
- ☐ Not at all
- ☐ Don't know

3. Technology

Which electronic gadget(s) are you using?

- ☐ Smart mobile phone
- ☐ Smart watch
- ☐ Tablets
- ☐ Computers
- ☐ Others (please specify) _______________
- ☐ I don't want to use any of electronic devices.

Do you find difficulty in using electronic gadgets?

- ☐ Yes, always
- ☐ Yes, sometimes
- ☐ Yes, very often
- ☐ Seldom
- ☐ Not at all
- ☐ Don't know

Do you use internet?

- ☐ Yes, always
- ☐ Yes, sometimes
- ☐ Yes, very often

- ☐ Seldom
- ☐ Not at all
- ☐ Don’t know

Would you be interested in having a system which monitors your health status like blood pressure, heart rate, body temperature?

- ☐ Yes, always
- ☐ Yes, sometimes
- ☐ Yes, very often
- ☐ Seldom
- ☐ Not at all
- ☐ Don’t know

Would you be interested in having a system which keep reminding you for your medication?

- ☐ Yes, always
- ☐ Yes, sometimes
- ☐ Yes, very often
- ☐ Seldom
- ☐ Not at all
- ☐ Don’t know

Would you be interested in having a device which can monitors your health status like daily intake of food, calories count, and water consumption per day?

- ☐ Yes, always
- ☐ Yes, sometimes
- ☐ Yes, very often
- ☐ Seldom
- ☐ Not at all
- ☐ Don’t know

Would you feel comfortable if a robot would be around you in the house and communicating to you?

☐ Yes, always

☐ Yes, sometimes

☐ Yes, very often

☐ Seldom

☐ Not at all

☐ Don’t know

Would you like to have system which notifies your kin/caregivers immediately if you have some sort of accident inside home (e.g. fall down, get hit by furniture, slipping in bathroom etc.), so that help can be arranged?

☐ Yes, always

☐ Yes, sometimes

☐ Yes, very often

☐ Seldom

☐ Not at all

☐ Don’t know

Are you afraid of falls and their consequences?

☐ Yes, always

☐ Yes, sometimes

☐ Yes, very often

☐ Seldom

☐ Not at all

☐ Don’t know

Would you feel safer/ more confident having a fall detection system automatically calling for help?

- [ ] Yes, always
- [ ] Yes, sometimes
- [ ] Yes, very often
- [ ] Seldom
- [ ] Not at all
- [ ] Don’t know

If you would have a wearable fall detection system, which ways of wearing it seem to be comfortable for you? (Multiple selections possible)

- [ ] On the belt
- [ ] Stuck to your hip with a patch
- [ ] Necklace
- [ ] Wrist

Would you like to have a smart house which can provide assistance in your household activities like:

Lock/unlock the door

Controlling lighting of the house,

Turning off/on TV, music player, AC

Window shield going up/down?

- [ ] Yes, always
- [ ] Yes, sometimes
- [ ] Yes, very often
- [ ] Seldom
- [ ] Not at all
- [ ] Don’t know

Social Engagement

Do you like to go outside your apartment?

- ☐ Yes, very often
- ☐ Yes, sometimes
- ☐ Seldom
- ☐ Not at all

How many times you go outside your apartment?

- ☐ Never
- ☐ Once in a day
- ☐ Twice
- ☐ More than three times
- ☐ Others (please specify) _______________

Which time you prefer to go outside?

- ☐ Morning
- ☐ Evening
- ☐ Afternoon
- ☐ Night
- ☐ Special occasions
- ☐ Never
- ☐ Others (please specify) _______________

How much you time you spend outside your home for some outdoor activities?

- ☐ Less than 30 minutes
- ☐ About 30 minutes
- ☐ 1 hour
- ☐ More than 1 hour
- ☐ I never go out

Do you feel comfortable in going out alone outside?

- ☐ Yes, always
- ☐ Yes, sometimes
- ☐ Yes, very often
- ☐ Seldom
- ☐ Not at all
- ☐ Don’t know

Would you like you have a companion with you when you go out?

- ☐ Yes, always
- ☐ Yes, sometimes
- ☐ Yes, very often
- ☐ Seldom
- ☐ Not at all
- ☐ Don’t know

Would you like keep a device with you which can assist you in outdoor activities and keep track of your location when you are out?

- ☐ Yes, always
- ☐ Yes, sometimes
- ☐ Yes, very often
- ☐ Seldom
- ☐ Not at all
- ☐ Don’t know

Dank u wel

Appendix B

Online Questionnaire

Smart Home Technologies Survey

Welcome!
This is an anonymous, confidential survey, which is held as part of the EU H2020 ACROSSING project (http://www.acrossing-itn.eu/).
ACROSSING aims to address existing challenges of Smart Home research by contributing towards an open smart home technology infrastructure by interlinking disciplines from sensing technologies, context inferences and interaction and considering key principles of social impact, ethics, security and privacy. This survey focuses on the identification of needs, requirements and barriers in the adoption of Smart Home technologies.

Answering this survey will take approximately 10 minutes.

This survey does not require you to disclose your identity.

* Required

1. Gender *

Mark only one oval.

- Male
- Female
- Other: ____________

2. Age *

Check all that apply.

- [ ] Under 18
- [ ] 18-24
- [ ] 25-35
- [ ] 35-55
- [ ] 55-70
- [ ] Over 70

3. Which of these best describes your location? *

Mark only one oval.

- Asia
- Africa
- North America
- South America
- Europe
- Australia/Oceania

4. What is the level of your highest education title *

Mark only one oval.

- I have not completed school
- I have completed school
- I have a university degree
- I have a postgraduate degree (MSc/PhD)
- Other: ______________________

5. How would you classify yourself as a computer user? *

Mark only one oval.

- Beginner - I am just getting used to using a computer.
- Basic Knowledge - I can perform basic functions
- Moderate - I have been using computers for a little while and know my way around
- Expert - I have intense knowledge of computers and their functions
- Other: ______________________

Home Activities

6. I feel safe inside the home when i am alone *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

7. I can do all my daily living activities by myself like bathing, cooking, grooming etc. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

8. I always forget to turn off the lights, kitchen appliances, take medicine etc. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

9. I would feel safer if i had a home security system in my home *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

10. I would like to put cameras outside my house, to know who visited my home *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

11. I do not mind if cameras are installed inside my home *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

12. I am always physically active inside my home *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

Social Engagement

13. I go outside daily to do some physical activity like walking, exercising etc. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

14. I do not feel comfortable going out alone (e.g walking, shopping etc) *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

15. I would not mind if I was monitored when outside the home. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

16. I would not mind if monitoring data is shared with my caregivers or relatives, so that they can track my location in case of emergency. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

Smart Home Technologies

17. Do you know what a smart home is? *

Mark only one oval.

◯ Yes
◯ No
◯ Kind of

18. Do you own any of these IoT Devices (Please select all that apply) *

Check all that apply.

- [ ] Smart phone
- [ ] Smart Watch
- [ ] Smart TV
- [ ] Fitness Bracelet (e.g. FitBit)
- [ ] Tablet
- [ ] Presence Sensors
- [ ] Sleep Monitors
- [ ] Monitoring Cameras
- [ ] Smart Blood Pressure Cuff
- [ ] Voice Assistant (e.g Amazon Echo, Google Home etc)
- [ ] Thermostats

Other: [ ] ______________________

19. According to you, what are the facilities your smart home must have? (Please select all that apply) *

Check all that apply.

- [ ] Automatic lighting and heating control
- [ ] Automatic locking/unlocking of doors and windows
- [ ] Electric appliances controlling e.g kitchen stove, refrigerator, TV, air conditioner etc
- [ ] Security and safety with cameras installed to monitor visitors
- [ ] Monitoring resident's health status (heart rate, blood pressure, step counts)
- [ ] Monitoring resident sleep and wake patterns
- [ ] Medicine reminder system
- [ ] Fall detection system
- [ ] Emergency alarm system to inform and call children, relatives or doctors in case of need

Other: [ ] ______________________

20. Suppose the smart home is monitoring your daily living activities (time spent in kitchen, in sleeping, cooking, eating, walking) and your health status, which data/information you would like to share with your doctor? (Please select all that apply) *

Check all that apply.

- [ ] Dietary routines and calorie count
- [ ] Time spent inside and outside the home
- [ ] Physical activity level e.g step count, walking, running, sitting
- [ ] Exercise routines
- [ ] Blood pressure
- [ ] Heart rate
- [ ] Stress level
- [ ] Sleep and awake patterns during night time

Other: [ ] ______________________

21. According to you, what are the potential benefits of smart home technologies? (Please select all that apply) *

Check all that apply.

- [ ] Save time
- [ ] Save energy
- [ ] Save money
- [ ] Provide comfort and make things effortless
- [ ] Provide peace of mind
- [ ] Improve quality of life
- [ ] Improves health
- [ ] Provide safety
- [ ] Increase property value

Other: [ ] ______________________

22. According to you, what can be the drawbacks of smart home technologies? (Please select all that apply) *

Check all that apply.

- [ ] Increase dependence on technology
- [ ] Disrupts daily routines
- [ ] Make less physically active
- [ ] Monitors private activities
- [ ] Increase dependence on outside experts
- [ ] Requires timely maintanance
- [ ] Non-essential luxuries

Other: [ ] ______________________

23. Would you have any concerns about living in a smart home? *

____
____
____
____
____

Personal Assistant

24. I would like to have a robot or virtual assistant in my home to help in daily living activities like ordering food, recommending exercise etc. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

25. Do you prefer to talk or touch the interface? *

Mark only one oval.

- Talk
- Touch
- Both
- Neither
- Other: ______________________

26. I would talk to the interface only if it speaks my native language *

Mark only one oval.

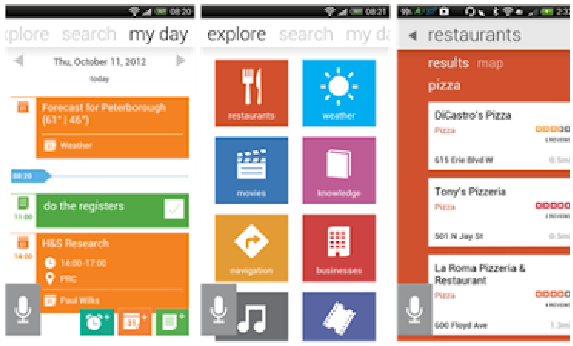
	1	2	3	4	5	
Strongly Disagree						Strongly Agree

27. I do not want the interface to talk to me unless I start to talk with it *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

28. Would you prefer an avatar (e.g. Image 1) or an interface (e.g. Image 2) *



Mark only one oval.

- ◯ Avatar
- ◯ Interface
- ◯ Both
- ◯ Neither
- ◯ Other: ____________________

29. I would like the voice of the system to change according to my preferences such as age, gender and accent *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

General attitudes towards monitoring

30. I do not mind being monitored unobtrusively in my home. *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	Strongly Agree

31. I do not mind being monitored as long as the data collected is useful for my doctor *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	Strongly Agree

32. I do not care who has access to information from in-home activity or computer use monitoring *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	Strongly Agree

33. I would not mind being videotaped to monitor my movement around the house *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	Strongly Agree

34. I prefer to live in my own house as long as possible, as technologies allow this, even if it is at the expense of my privacy *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

Attitudes towards sharing information

35. I would want information about my activity sent to me if there was a change in my activity (e.g. physical activity, sleep patterns) *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

36. I would want information about my activity sent to me if the changes suggest that I might have a chronic disease *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

37. I would want information about my activity sent to a family member if the changes suggest that I might have a chronic disease *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

38. I would want information about my activity sent to my doctor if the changes suggest I might have a chronic disease *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

39. I am willing to have information from activity monitoring (e.g. sleeping, eating, exercising) shared with my family *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

40. I am willing to have information from activity monitoring (e.g. sleeping, eating, exercising) shared with my doctor *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

41. Imagine there is a new inexpensive thermostat sensor for your house that can learn about your temperature preferences and movements around the house and potentially save money on your energy bill. It is programmable remotely in return for sharing data about some of the basic activities that take place in your house like when people are there and when they move from room to room. Do you think that is acceptable or not? Why do you think so? *

42. Imagine that in the future a smart fridge would be able to know when you run out of food and order for you all the groceries you need. The smart fridge will keep track of your shopping habits and might give them to third parties. Do you think that is acceptable or not? Why do you think so? *

Privacy and Security Concerns

43. I am concerned information from my smart home could be given to people/organizations that would use it in a way that would harm me *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

44. I am concerned about privacy in relation to in-home activity monitoring *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

45. It bothers me that my data might be visible and/or accessible by others *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

46. Suppose there was a leak of information that gave access to all the information in your smart home. Specifically someone else would have access to: medical information, banking details, location, movement from room to room, energy consumption, monitoring cameras, home security system, exercise (e.g. FitBit data, shopping habits). (Please select for each item what would concern you) *

Check all that apply.

	Financial risk of the content	Potential of embarrassment upon content exposure	Ease of access to the content by others (e.g friends, family, third parties)
Medical Information	☐	☐	☐
Location	☐	☐	☐
Movement from room to room	☐	☐	☐
Energy Consumption	☐	☐	☐
Monitoring Cameras	☐	☐	☐
Home Security System	☐	☐	☐
Exercise Data	☐	☐	☐
Shopping Habbits	☐	☐	☐
Banking Details	☐	☐	☐

47. How private do you consider each of these information items from the above scenario (Please select for each item) *

Mark only one oval per row.

	1- Not Private	2 - Neutral	3 - Somewhat Private	4 - Private	5 - Very Private
Medical Information	⬭	⬭	⬭	⬭	⬭
Location	⬭	⬭	⬭	⬭	⬭
Movement from room to room	⬭	⬭	⬭	⬭	⬭
Energy Consumption	⬭	⬭	⬭	⬭	⬭
Monitoring Cameras	⬭	⬭	⬭	⬭	⬭
Home Security System	⬭	⬭	⬭	⬭	⬭
Exercise Data	⬭	⬭	⬭	⬭	⬭
Shopping Habbits	⬭	⬭	⬭	⬭	⬭
Banking Details	⬭	⬭	⬭	⬭	⬭

48. If you have any remarks (additional comments) that you would like to make, please write them below.