DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

ABOU-BEKR BELKAID UNIVERSITY OF TLEMCEN

THESIS

Submitted to:

FACULTY OF SCIENCES - DEPARTMENT OF INFORMATICS

To obtain the degree of:

Doctorate

In Informatics, Specialty of Artificial Intelligence and knowledge Engineering

Presented by:

Mme BENMANSOUR Asma spouse BELKHERROUBI

On the topic:

Human Activity Recognition in the Context of Multi-occupant Smart Homes

---

Publicly defended on 9/11/2016 at Tlemcen in front of the jury composed of:

CHIKH Mohamed Amine	Pr. at Tlemcen University	Chairman
FEHAM Mohammed	Pr. at Tlemcen University	Supervisor
BOUCHACHIA Abdelhamid	Pr. at Bournemouth University, U.K.	Co-Supervisor
HAFFAF Hafid	Pr. at Oran University	Examiner
MOURID Tahar	Pr. at Tlemcen University	Examiner
KECHAR Bouabdellah	MCA at Oran University	Examiner

Acknowledgements

I express my deep gratitude to my co supervisor Mr BOUCHACHIA Abdelhamid for his scientific mind and pedagogy, for lending me his most precious time and helping me by his valuable guidance.

I thank my supervisor Mr FEHAM Mohammed who welcomed me into the STIC laboratory, helped me and encouraged me throughout the completion of this thesis.

I thank my father Mr BENMANSOUR Djamal who represents for me a model as a researcher and who helped me and supported me in order to complete this work.

I thank Mr CHIKH Mohamed Amine who has accepted to be chairman of the jury, I thank Mr HAFFAF Hafid, Mr KECHAR Bouabdellah and Mr MOURID Tahar who have accepted to evaluate this work.

"Satisfaction lies in the effort, not in the attainment,

full effort is full victory"

Mahatma Gandhi

To my sweet daughter whose love and joy allowed me to always move forward in my work

To my dear parents who are always there for me

To my loving grandmother and my beloved late grandfather who represent models of labor and perseverance

To my dear husband for his support and patience with me

To my brother and my sister to whom I wish happiness, health and success

To my husband's family who have always encouraged and supported me

List of abbreviations

ADL Activities of Daily Living

AHMM Abstract Hidden Markov Model

AHRI Aware Home Research Initiative

ANN Artificial Neural Network

ARAS Activity Recognition with Ambient Sensing

BN Bayesian Network

CASAS Center for Advanced Studies in Adaptive Systems

CHMM Coupled Hidden Markov Model

CHSMM Coupled Hidden-Semi Markov Model

CLA Combined Label of Activity pairs

CL-HMM Hidden Markov Model-based Combined Label

CLO Combined Label of Observation pairs

CRF Conditional Random Field

CSS Contact Switches Sensors

CV Cross Validation

DBN Dynamic Bayesian Network

DCS Door Contact Sensor

DOMUS Domotics and Mobile computing Research

DT Decision Tree

EPs Emerging Patterns

ESN Echo State Network

FCRF Factorial Conditional Random Field

GR Growth Rate

GUI Graphical User Interface

HAR Human Activity Recognition

HDP-HMM Hierarchical Dirichlet Process-Hidden Markov Model

HHMM Hierarchical Hidden Markov Model

HMM Hidden Markov Model

HSMM Hidden-Semi Markov Model

ICS Institute of Computer Software

IHMM Infinite Hidden Markov Model

KNN K-Nearest Neighbor

LHMM Linked Hidden Markov Model

LPW Low Powered Wireless

MAP Maximum-A-Priori

MC Markov Chain

MIT Massachusetts Institute of Technology

MLP Multi-Layer Perceptron

MM Markov Model

NB Naive Bayes

PCNP Personal Computer Networking Protocols

PHMM Parallel Hidden Markov Model

PIR Passive Infrared sensors

PLC Power Line Communication

RBFN Radial basis function network

RFID Radio Frequency Identification

SH Smart Homes

TDNN Time-Delay Neural Network

TV Television

UMTS Universal Mobile Telecommunications System

WSU Washington State University

Contents

List of Figures

List of Tables

I. Introduction

I.1 Context

Be it for the developed or the less developed countries in the world, the continued increase of the aging population is a serious problem. In 2020, the number of older adults aged 60 and over is supposed to reach 1 billion, and perhaps 2 billion by 2050 (Bloom et al., 2010). One in ten older people generally lives alone all over the world. Some of them suffer from physical (e.g. reduced mobility) or cognitive diseases (e.g. dementia, Alzheimer) which reduce their ability to live independently and sometimes keep them in risk situations (e.g. forgetting the stove on). Because there is a lack of infrastructures designed to manage the elderly population, a rising of healthcare costs and a shortage of nursing staff assistance, recent Smart Homes (SHs) research has focused on maintaining them at home by developing assisted living technologies which help them in the completion of their Activities of Daily Living (ADL). In this context, Human Activity Recognition (HAR) aims to recognize the ADLs of occupants at home.

SH technology aims to support people to have a better quality of life and to ensure elderly to live comfortably and independently (Demiris et al., 2004). The SH technology is considered as a way to reduce living and care costs and to improve the quality of life for people with care needs. It has been applied for many purposes (Miskelly, 2001) like energy saving, security and safety, fall detection, light management, smoke and fire detection etc. using various solutions such as video monitoring, alarms, smart planners and calendars, reminders, etc. Equipped with sensors, actuators and eventually cameras to collect different types of data about the home and the occupants, SHs can enable automatic systems or caregivers to control the environment on behalf the occupants, predict their actions and track their health condition.

A SH system incorporates different components structured in layered architecture as illustrated in figure 1. Each layer of the system has its own function and comes with its own challenges to be dealt with. The home is first equipped with sensors designed to perceive the state of both the environment and the occupant while he/she goes to perform his/her ADLs. The Data is collected as the physical layer by these sensors and transmitted through the communication layer to the processing unit in the reasoning layer. The communication layer plays the role of connecting all of the components such as sensors, actuators, gateway and storage hardware and consists of the different communication technologies such as : Low Powered Wireless (LPW) networks, Power Line Communication (PLC), Personal Computer Networking Protocols (PCNP) and Universal Mobile Telecommunications System (UMTS). In the reasoning layer (i.e. the "smart part" of a SH system) data potentially undergoes a pre-processing step for cleaning, the main step is, however, analysis which encompasses: HAR, behavior patterns discovery, detecting abnormal behavior, etc. The outcome of the analysis, allows to assess the physical and cognitive capabilities of the occupant, in order to, alerts or warnings the stakeholders (i.e. caregivers, occupant's relatives) though the interface layer.

This system architecture allows to determine the type of services and assistance required for the occupant over time and consequently to enhance their well being and independent living.

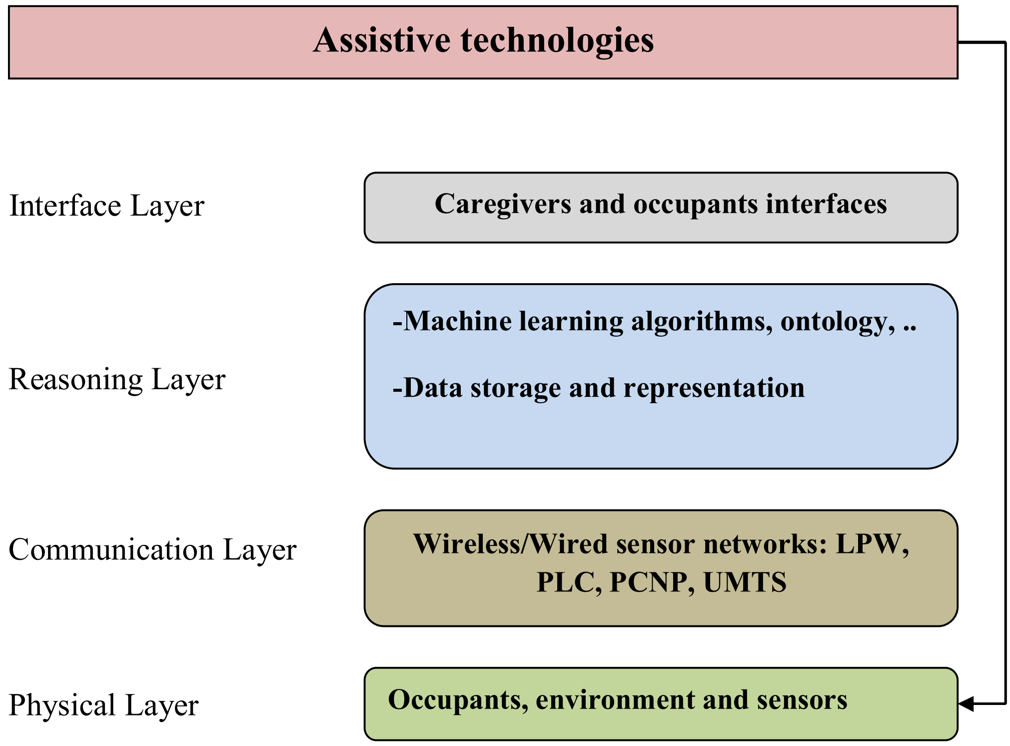

Figure 1: The layered architecture of an SH

So far research related to HAR has devoted a particular attention to the issue of monitoring of a single occupant in a SH assuming that in general elderly individuals live alone (Khan et al., 2012) (Hu et al., 2009) (Riboni et al., 2011) (Kasteren et al., 2008) (Kasteren et al., 2010) (Kasteren et al., 2011) (Sarkar et al., 2010) (Nait Aicha et al., 2013) (Gu et al., 2009a). However, the monitoring process is continuous and sometimes scenarios in which multiple people are simultaneously present within the home may take place even though the house is usually inhabited by a single occupant (e.g. receive visits from family members or professional health care givers).The SH solution for maintaining older people at home should not only focus on recognizing ADLs of single inhabitants. Extending HAR systems to multiple occupants referred to as multi-occupancy is necessary and should contribute to the facilitation of the deployment of these systems in real-world environments.

Multi-occupancy has not been studied much so far, because the field is still young and because many outstanding challenges in single occupancy remain unresolved, such as the recognition of complex activities (Liu et al., 2015, p. 2) and interleaved activities (Meditskos et al., 2015). Several recent papers highlight the challenges encountered in this field as shown in (Amiribesheli et al., 2015) and (Ni et al., 2015). In a recent work, Amiribesheli et al. (Amiribesheli et al., 2015) discuss the challenges related to data processing (i.e. maintaining the security, privacy and reliability of an activity data) and to activity recognition modeling (i.e. recognizing interleaved and concurrent activities, imbalanced data, online activity learning, applicability and adaptability of the activity model, scalability of the activity model).

Ni et al. (Ni et al., 2015) discuss eight challenges to solve before improving the quality of life in a SH for an elderly. Clearly, HAR in a multi-occupant environment only represents one of the challenges to face among many others relevant to HAR in a single occupant environment. Recently an increasing interest has been witnessed acknowledging the prominence of multi-occupancy as a research area in the context of SHs and activity recognition.

Different types of sensors have been used for multi-occupant activity recognition in SHs, but most of the work has considered video cameras and computer vision techniques to develop HAR systems (Nguyen et al., 2006) (McCowan et al., 2005) (Du et al., 2006) (Natarajan and Nevatia, 2007) (Du et al., 2007). The use of camera is, however, not suitable due to privacy concerns and vision based studies are out of the scope of this thesis. Recently, many studies have been interested in the use of pervasive sensors to recognize multi-occupant ADLs. In this context, we clearly distinguish two main types of developments:(i) those based on wearable sensors such as accelerometer, gyroscope, etc (Wang et al., 2009) (Wang et al., 2011) and (ii) those based on infrastructure sensors, such as motion, reed switches, etc (Hsu et al., 2010) (Chiang et al., 2010) (Cook et al., 2010) (Singla et al., 2010) (Prossegger and Bouchachia, 2014) (Alemdar et al., 2013). A lot of studies have been conducted using wearable sensors, where the identification of the person triggering the sensor individuals is straightforward. The disadvantage of wearable sensors is that they cause inconvenience and are impractical for situations in which individuals are opposed to wear the sensors, forget to wear them like elderly people with cognitive impairment. Moreover, pervasive infrastructure sensors offer the advantage of being non obtrusive to people as they are seamlessly placed in the environment. They could be either wall-mounted (e.g. motion sensors placed on the ceiling) or placed on objects (e.g. reed switches placed on doors). Using these non-intrusive sensors allows the occupants to live as normally as possible and not feel restrained by the technology that surrounds them while they perform their ADLs at home. In this thesis, we focus only on the latter technology.

The remainder of this chapter is organized as follows. Section I.2 describes the types of activities of daily living. Section I.3 highlights the problem of multi-occupancy focusing on two aspects: data association and interaction. Section I.4 presents our research focus and questions. Section I.5 gives an overview on the different chapters of the thesis.

I.2 Activities of Daily Living

Activity recognition is the process of automatically identifying human actions from the data captured by various types of sensors. It is relevant to many real-world applications such as surveillance, assisted living, and healthcare. Modeling simple activities has been the focus of most of the activity recognition research, while complex activities have only recently started to attract attention from the ambient intelligence and pervasive computing communities (Kim et al., 2010) (Liu et al., 2015, p. 2). Complex activities are common and can be performed by either single persons or by a group of people. We can therefore distinguish different types of activities:

- **Complex activity** consists of many sub-activities as fine-grained activities. For instance, the activity "*cooking soup*" could be modeled as a sequence of sub-activities: *measure water, pour water into a pot, add content of the bag, cook, and serve in a bowl*.

- **Simple activity** is usually an atomic activity which cannot consist of simpler activities, for instance, *pour water*.

Moreover we can distinguish two types of ADLs:

- **Basic ADLs** refer to self-care tasks (e.g. eating, moving, dressing, bathing and showering, grooming and toilet hygiene) (Roley et al., 2008);

- **Instrumental ADLs** are not essential for basic living, but they let an individual live independently in a community (e.g. doing housework, meeting with people, doing shopping, taking medicine, using of technology, using transportation) (Bookman et al., 2007).

Most of the state-of-the-art research has investigated monitoring and assisting people in single-occupancy living spaces. Nevertheless, living spaces are usually inhabited by more than a single person; hence, designing solutions for handling multi-occupancy is of prominent importance. In fact, recently, multi-occupancy research has gained more attention. However, the pace of research is slow and many outstanding problems are still ahead. The reason for this is that there have been numerous other challenges with single-occupancy to deal with before tackling multi-occupancy.

The research work published on multi-occupancy is mainly related to activity modeling and data association. The challenge is to find suitable models to address the problem of data association, to build activity recognizers that capture the various interactions between occupants. Data association is about the identification of the occupants, by whom each sensor is triggered. That is about mapping sensed data to the occupant who actually caused the generation of the data. In activity modeling, we distinguish between 5 types of activities:

(1) **Sequential activities** where each activity is performed after another in a sequential fashion without any interleaving (e.g. make a phone call, washing hands and then cooking).
(2) **Interleaving activities** where a single occupant switches between many activities (e.g. switching between chopping vegetables and stirring soup in the kitchen) as shown in figure 2.

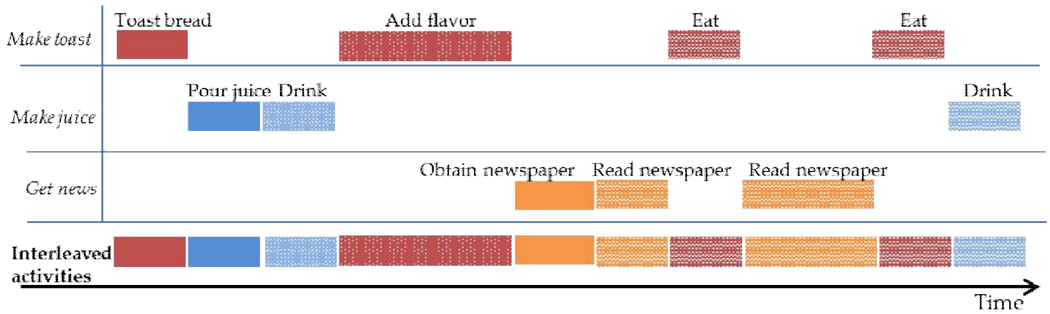

Figure 2: Interleaved activities (a single occupant)

(3) **Concurrent activities** where a single occupant carries out more than one activity at a time (e.g. talking on phone call, while cooking).

(4) **Parallel activities** where many occupants perform many activities at the same time (e.g. one occupant is watching (Television) TV in the living room, while the other is cooking in the kitchen).

(5) **Cooperative activities** where many occupants work together in a cooperative manner such that each occupant performs certain actions of the same activity, either together (e.g. two persons moving a table by holding it by the ends) or in parallel (e.g. one person is chopping vegetables, while the other is boiling broth to make soup) as shown in figure 3.

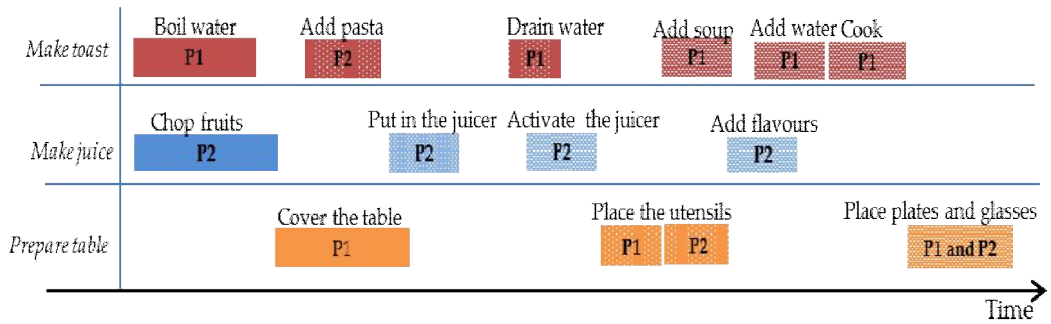

Figure 3: Cooperative activities (two occupants: P1 and P2)

While the first three activity types are concerned with a single person (also termed as individual activities), but done in the presence of multiple occupants, the latter two are relevant for multi-occupancy. Obviously, the complexity of ADLs increases as the number of occupants in the living environment increases and the activities tend to be cooperative (e.g. watching TV or play a board game). Also the cooperative activities tend to be generally instrumental.

The existing state-of-the-art literature on multi-occupant SHs indicates that these types of activities are not yet fully addressed. Much of the studies is done on simple scenarios like: elementary activities (Wilson and Atkeson, 2005) (e.g. whether a person moves or not) and sequential activities (Cook et al., 2010) (Singla et al., 2010), although, parallel individual and cooperative activities are the most frequent in nature. Almost, no work has addressed all types of activities. A more mature research in this area has been conducted by the computer vision community using normal cameras (Nguyen et al., 2006) (McCowan et al., 2005) (Du et al., 2006) (Natarajan and Nevatia, 2007) (Du et al., 2007). Vision-based studies are nevertheless out of the scope of this thesis.

I.3 Multi-occupancy problem

The challenge of multi-occupant SHs is to design a computational model to deal with the problem of data association (i.e. the identification of the occupant) and to efficiently capture the interactions between the occupants.

I.3.1 Data Association

In a SH environment shared by multiple occupants, the identification of the occupant is crucial. Recognizing who triggered the sensors' events allows efficient and accurate tracking the occupants' activities. The problem of data association consists of mapping the sensed data to the occupant causing its generation. Failing to do so, that data will not be useful and could even endanger the life of occupants in telehealth/telecare context, if important actions are to be taken based on the assessment of such activity data. The data association problem is encountered either when using non-intrusive sensors which cannot directly identify occupants in a SH (Crandall and Cook, 2008a) (Crandall and Cook, 2008b) (Crandall and Cook, 2010) (Hsu et al., 2010) (Wilson and Atkeson, 2005) (Cook et al., 2010) (Alemdar et al., 2013) (Chen and Tong, 2014) or when using unlabeled data (Alemdar et al., 2013). All the studies in the literature show that data association is a fundamental problem when modeling activities in a multiple-occupant environment (Hsu et al., 2010).

I.3.2 Cooperative activities (interaction)

The main difference between a single-occupant environment which is characterized by individual activities and a multi-occupant environment is the interaction between individuals to complete cooperative activities. Cooperative activities are usually inter-dependent activities. For example, a occupant cannot do "toileting" because the bathroom is busy. Instead, he decides to do "dressing" (Hsu et al., 2010). Clearly "dressing" takes place because "toileting" did not happen. But, "dressing" does not always occur when the bathroom is busy. Hence, these two activities are unrelated in the general case. Collective activities that involve many persons are usual in the daily life, such as watching TV, eating, gardening, etc.

People perform certain activities collectively because such activities require cooperation. In terms of interaction, we can distinguish two distinctive types of interdependence (Smith and Mackie, 1999): social interdependence and task interdependence. A task is socially interdependent if people rely on one another for its full completion, like playing monopoly. It would be more enjoyable to play monopoly in a group than alone. A task is interdependent if more than one person is required to accomplish the activity like moving a table into another room of the house requires at least two persons.

Studies report that old people tend to isolate themselves (Anon, 2014) which may lead to dementia (Fratiglioni et al., 2000) or simply to cause damages to their health (Cornwell and Waite, 2009). Recently, researchers have relied on wearable devices to study social behavior (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b) (Gross, 2007) (Olguin et al., 2009) (Eagle, 2008) (T. Choudhury, 2004) in the context of multi-occupant activity recognition. However, they have recognized the need to use non-intrusive sensors to monitor occupants' behavior and develop real-world applications for older adults.

I.4 Research Focus and Questions

In the last section we presented the most important issues encountered in a multi-occupant activity recognition that is the data association problem and modeling interaction. These issues also represent the major differences between single occupant activity recognition and multi-occupant activity recognition. In our thesis, we consider the data association problem to be solved and only focus on recognizing complex multi-occupant activities.

Recent developments in pervasive sensing technology make it possible to easily equip existing homes with low cost miniaturized sensor networks. Healthcare professionals emphasize that the activities must be monitored (Katz, 1983). Real world situations exhibit the need for SH systems to consider the coexistence of multiple occupants at home. What is missing are the pattern recognition methods to recognize multi-occupant activities automatically from sensor data. In this thesis, we answer the following questions with respect to that issue:

While there have been several review papers published over the recent years devoted to activity recognition and to smart environments in general (Acampora et al., 2013) (Chan et al., 2009) (Sadri, 2011) (Amiribesheli et al., 2015), does there exist a devoted survey to the area of multi-occupant activity recognition? there is no survey paper on multi-occupancy that draws the picture of the current advances in this area, hence, the importance of our survey paper (Benmansour et al., 2015). We provide full coverage of techniques, methods, and open issues related to multi-occupancy.

On which types of activities must we focus while performing multi-occupant activity recognition? Occupants can either perform the actions independently in parallel (e.g. one occupant is chopping vegetables while the other is boiling broth to make soup) or work together in a cooperative manner to accomplish an activity, such that each occupant performs certain actions of that activity (e.g. two occupants moving a table). In a multi-occupant activity recognition system, the recognition of cooperative activities and parallel individual activities is equally important. Some studies addressed the problem of multi-occupant activity recognition using non-intrusive sensors and few of them modeled cooperative activities (Alemdar et al., 2013) (Chen and Tong, 2014) (Chiang et al., 2010).

Which pattern recognition method is appropriate for modeling multi-occupant activities? Diverse computational models have been applied in the context of single occupant activity recognition ranging from probabilistic models to standard data mining and machine learning models like neural networks, decision trees, ontologies, etc. In the case of multi-occupancy, however, no such diversity of models exists. Almost all of the proposed models are essentially probabilistic based on graphical models. Our approaches are based on probabilistic graphical models as they are robust to the noise in activity data (i.e. sensor readings are usually noisy) and the uncertainty while performing activities (i.e. activities are typically performed in a nondeterministic fashion).

Which temporal probabilistic model is able to accurately recognize multi-occupant activities from sensor network data? Several probabilistic models have been applied in the context of single-occupant activity recognition. We distinguish two classes of probabilistic models that is generative and discriminative probabilistic models. Kasteren et al. (Kasteren et al., 2008) compared Hidden Markov Model (HMM) and Conditional Random Field (CRF), their study reports that generative models are more appropriate than discriminative models for imbalanced datasets. For example, in their "Ubicomp dataset", we can encounter more events related to the activity "going to bed" than those related to the activity "toileting". Since the same problem arises in multi-occupant activity datasets, our approaches investigate the use of generative models such as HMMs and variants of this model.

In (Chiang et al., 2010), Coupled Hidden Markov Model (CHMM) was applied. In (Chen and Tong, 2014) both HMM and CRF were applied and compared. In (Alemdar et al., 2013) HMM was used to model with multi-occupant activities. Our approach investigates this direction further. The goal is to accurately recognize both parallel and cooperative activities from non-intrusive sensors. We do not focus on only one of the two types of activities as done in the literature related to multi-occupant activity recognition, but on both types. Specifically, this work makes the following contributions:

- We propose a variant of the combined label approach based on HMM applied in (Chen and Tong, 2014) (Alemdar et al., 2013), we call it Hidden Markov Model-based combined label (CL-HMM). In addition to the use of combined labels for the pair of activities labels (i.e. Occupant 1 activity label and Occupant 2 activity label), our approach also suggests the use of combined labels for the pair of observations (i.e. the observation of Occupant 1 and the observation of Occupant 2).
- We also propose a linked version of HMMs called Linked Hidden Markov Model (LHMM) to model multi-occupant activities and describe the corresponding version of the Viterbi algorithm. To the best of our knowledge this model has never been applied for pervasive multi-occupant HAR. A baseline model that consists of parallel Hidden Markov Models (PHMMs) for the occupants is developed, where each occupant is modeled as one separate HMM. This model does not explicitly represent any inter-occupant interaction.

- We compare the performance of all of the proposed models, CL-HMM, LHMM and PHMM against the state-of-the-art model CHMM used in (Chiang et al., 2010).

I.5 Structure of the Thesis

This thesis is composed of four chapters. Chapter II gives an overview of pervasive multi-occupant activity recognition systems in the literature. Specifically, Section II.2 discusses sensor technology used in recent research related to pervasive multi-occupant activity recognition. Section II.3 presents a sample of publicly available datasets. Section II.4 presents the computational models used for modeling multi-occupant activities. Section II.5 presents some of the issues encountered in multi-occupancy such as identification and interaction. Section II.6 provides examples of international research groups in the pervasive computing area for both single-occupant and multi-occupant settings.

Chapter III presents our proposed approaches for multi-occupant activity recognition. Section III.2 describes the conventional HMM as all multi-occupant models applied in this chapter represent a variation of the latter model. In Section III.3 we present the details of the proposed models for Multi-occupant activity recognition that is the CL-HMM and the LHMM. We give the definition as well as details about parameter estimation and inference algorithms in Section III.3.1 and Section III.3.2 respectively for the two models. Section III.4 discusses our experiments. In Section III.4.1 we give a description of the experimental dataset as well as the pre-processing associated with. Section III.4.2 describes baseline models PHMM and CHMM against which our proposed models are compared. Two main experiments are studied in Section III.4.3 and Section III.4.4. In the first we present the results of the individual occupants using all models; while in the second experiment joint results after preprocessing are discussed. Special attention is given to the performance of the models on cooperative and parallel activities. Section III.5 discusses the comparison of our proposed models against existing studies which relied on the same dataset (i.e. "Multiresident ADLs" of Center for Advanced Studies in Adaptive Systems (CASAS)).

Finally, Chapter IV concludes the thesis. Specifically, we draw our conclusions and propose future works in Section IV.1 and Section IV.2 respectively. Section IV.3 goes through a sample of open questions in the area of multi-occupant activity recognition.

I.6 Publications

The work presented in this thesis has been published in the following papers:

- **Conference paper:** Asma Benmansour, Abdelhamid Bouchachia, Mohammed Feham, *"Human activity recognition in pervasive single resident smart homes: State of art"*, 12 th International Symposium on Programming and Systems (ISPS), April 2015.
  DOI: 10.1109/ISPS.2015.7244997

URL:
http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7244997&url=http%3A%2F%2Fieeexplore.ieee.org%2Fstamp%2Fstamp.jsp%3Ftp%3D%26arnumber%3D7244997

- **Article-1-:** Mohsen Amiribesheli, Asma Benmansour, Abdelhamid Bouchachia, *"A review of smart homes in healthcare"*, Journal of Ambient Intelligence and Humanized Computing, Volume 6, Issue 4, Pages: 495-517, August 2015.
DOI:10.1007/s12652-015-0270-2
URL: http://link.springer.com/article/10.1007%2Fs12652-015-0270-2

- **Article-2-:** Asma Benmansour, Abdelhamid Bouchachia, Mohammed Feham, *"Multioccupant Activity Recognition in Pervasive Smart Home Environments"*, ACM Computing Surveys (CSUR), ISSN: 0360-0300, Volume 48, Issue 3, Pages: 34:1-34:36, December 2015.
DOI: 10.1145/2835372
URL: http://dl.acm.org/citation.cfm?id=2835372

- **Article-3-:** Asma Benmansour, Abdelhamid Bouchachia, Mohammed Feham, *"Modeling Interaction in Multi-Resident Activities",* Neurocomputing-journal-Elseiver, ISSN: 0925-2312, EISSN: 1872-8286. Manuscript number: NEUCOM-D-15-01722R2

CHAPTER II

Multi-Occupant Activity Recognition: Related Work

II. Multi-Occupant Activity Recognition: Related Work

II.1 Introduction

In this chapter, we focus on state of the art studies relying on pervasive sensors. Pervasive sensors are used to collect data related to the human physiology, the human activity as well as the environment. Such data is processed in order to extract cues and patterns about various aspects such as the occupant's profile, health status of the occupant, the living environment and the occupant-environment interaction. Because of the very complex nature of human activities, the task of recognition in a pervasive context is very difficult, especially when pervasive data generated by sensors is noisy. Multi-occupancy comes with specific scientific and technological challenges (Chen and Tong, 2014) related to occupant identification (Crandall and Cook, 2008a) (Crandall and Cook, 2008b) (Crandall and Cook, 2010) (Hsu et al., 2010) (Wilson and Atkeson, 2005) (Cook et al., 2010) (Alemdar et al., 2013) (Chen and Tong, 2014), activity tracking (Prossegger and Bouchachia, 2014) (Crandall and Cook, 2009), behavior patterns of occupants (Gu et al., 2009b), and conflict management (Hsu and Wang, 2008).

The remainder of this chapter is organized as follows. Section II.2 discusses sensor technology used in recent research related to pervasive multi-occupant activity recognition. Section II.3 presents publicly available multi-occupancy datasets and outlines main features of these ones such as type of sensors used to register the data, activities concerned and type of annotation applied. Section II.4 presents the computational models used by state-of-the-art studies for modeling multi-occupant activities. Section II.5 presents main issues encountered in multi-occupancy such as occupant identification, interaction and scalability of activity models. Section II.6 provides examples of international research groups in the pervasive computing area for both single-occupant and multi-occupant activity recognition.

II.2 Sensing for Multi-Occupancy Activity Recognition

An SH system consists of two types of components: hardware components and software ones. The former integrates sensors and associated equipments like controllers and gateway equipments into a single network. Sensors are devices for detecting changes in the environment including the occupants. There is a large variety of sensors used to monitor SHs and the occupants. Sensors are used to collect various types of data related to: activities of the occupants, states of the objects and states of the environment (Orwat et al., 2008). In particular sensors capture the following data (Ding et al., 2011) (Ye et al., 2012):

- Strain and pressure
- Position, direction, distance and motion
- Light, radiation, temperature and humidity
- Type of material (e.g. solid, liquid and gas)
- Sound

- Image and video
- State of the object (e.g. present, absent)
- Physiological measurements (e.g. blood sugar, blood pressure)

Sensors can be classified according to different characteristics such as the type of data they produce that is discrete state sensors (e.g. Passive Infrared sensors (PIR) which produce binary values) and continuous state sensors (e.g. temperature sensors). They can also be categorized according to their mobility for example the location of infrastructure sensors and cameras is defined in advance, hence they are definitely maintained in the same place (e.g. PIRs and cameras are placed on the ceiling) of the environment in contrast to wearable sensors which usually move according to the occupant who carry them.

In terms of sensor deployment and selection, table 1 presents different types of sensors used for multiple-occupant activity recognition. We can clearly distinguish three major classes of approaches, the first one rely on the use of wearable sensors (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b) (e.g. Radio Frequency Identification (RFID)), the second one is based on infrastructure sensors (e.g. PIRs) while the third one lies between the use of wearable sensors (i.e. embedded sensors in each occupant's Smartphone) and the use of infrastructure sensors (i.e. motion sensors) (Roy et al., 2013) (Roy et al., 2016). In the following we give a short definition as well as some examples of both wearable and infrastructure sensors, we also highlight the advantages and disadvantages while using each sensor class:

Table 1: Summary of selected studies

Ref	Dataset	Sensors	Data association	Activities covered[1]	Approach	Test & validation [2]	Evaluation metric[3]	Results
(Crandall and Cook, 2008a)	CASAS Student lab	Motion, Light, Door switch	yes	-	NB	Hold-out method	-Accuracy -False positive	- 92% - 7%
(Crandall and Cook, 2008b)	Same as (Crandall and Cook, 2008a)	Same as (Crandall and Cook, 2008a)	yes	-	NB and HMM	Same as in (Crandall and Cook, 2008a)	- Accuracy - False positive	- NB: same as in (Crandall and Cook, 2008a). -HMM: 84%; NB: 76%
(Crandall and Cook, 2010)	CASAS B&B and TwoR	Motion, Door, Cabinet, Water flow, Power, light, power	yes	-	NB and HMM on datasets B&B and TwoR	3-fold CV	- Accuracy rate - Average lag - Error rate	-NB:93.3% (B&B),89.3% (TwoR) -HMM: 94% (B&B) 90.2% (TwoR)
(Hsu et al., 2010)	CASAS Multiresident ADLs	Motion, Item sensors, Cabinet, Water, Burner, Phone, Temperature	yes	Sequential	2 CRFs (data association, activity recognition)	Leave-one-out CV	Average accuracy	50% (for raw representation of data= best representation)
			no	Sequential Parallel	CRF with decomposition inference			59% (O1), 64% (O2)
(Wilson and Atkeson, 2005)	Simulated data	Motion, Contact switches, Pressure mat, Beam	yes	-	HMMs (one HMM for each occupant)	One day of data	Time slice accuracy	-Off-line learning: 100% (1O) to 82% for 4Os -On-line learning: 100% (1O) to 67% (4 Os) -Combined learning: 100% (1 O) to 74% (4Os)
	Real-world data	RFID, Motion,						-85.3% for 1O

[1] "-" means that the authors didn't perform activity recognition, only data association was considered
[2] CV stands for "cross validation"
[3] O stands for "occupant", O1 indicates Occupant1

		Contact						-82.1% for 2Os -86.4% for 3Os
(Chiang et al., 2010)	Same as (Hsu et al., 2010)	Same as (Hsu et al., 2010)	no	Sequential Parallel Cooperative	- PHMM - CHMM - DBNs extended	Leave-one-out CV	- Accuracy (O1) - Accuracy (O2) - Joint accuracy	-78.85%, 81.62%, 84.72% -75.92%, 84.03% and 86.44% -61.78%, 74.9%, 78.28%
(Cook et al., 2010)	Same as (Hsu et al., 2010)	Same as (Hsu et al., 2010)	yes	Sequential	- Bayesian update - HMMs (1 for data association, 1 HMM for recognition)	3-fold CV	-Average accuracy -Average precision -Average recall -Average f-score	- HMM: 90%, Bayes: 57% - HMM: 93% - HMM: 96% - HMM: 94%
(Singla et al., 2010)	Same as (Hsu et al., 2010)	Same as (Hsu et al., 2010)	no	Sequential	1 HMM for all activities of both occupants	3-fold CV	Average accuracy	60.60%
				Sequential Parallel	1 HMM for each occupant			73.15%
(Alemdar et al., 2013)	ARAS	Photocell, Infrared receiver, Force, Proximity, Sonar distance, Temp., Contact, Pressure mat	yes	Sequential Parallel Interleaved Cooperative	1 HMM (combined label for multi-occupant activities)	Leave-one-out CV	Average accuracy	- 61.5% (House A) - 76.2% (House B)
(Wang et al., 2009)	Real-world Data	Audio recorder, iMote2 with ITS400, RFID	no	Sequential Parallel Interleaved Cooperative	CHMM	10-fold CV	-Time slice overall accuracy -Time slice single-occupant	-82.22% (O1) and 88.71% (O2). -74.79% and 85.11. -96.91% and 95.91%.

		wristband reader					ADLs -Time slice multi-occupant ADLs	
(Wang et al., 2011)	Same as (Wang et al., 2009)	Same as (Wang et al., 2009)	no	Sequential Parallel Interleaved Cooperative	CHMM	10-fold CV	Same as in (Wang et al., 2009)	Same results as in (Wang et al., 2009)
					FCRF			-86.7% and 86.37% -85.75% and 82.56% -87.02% and 88.84%
(Gu et al., 2009b)	Same as (Wang et al., 2009)	Same as (Wang et al., 2009)	no	Sequential Parallel Interleaved Cooperative	EPs (EPs for single user ADLs and EPs for multi-occupant ADLs)	10-fold CV	-Time slice single-occupant ADLs -Time slice multi-occupant ADLs	-86.69% (O1) and 85.57% (O2) -95.06% (O1) and 95.71% (O2)
(Lin and Fu, 2007)	Real-world data	Motion, Thermometers, Humidity, light sensors, Smoke, RFIDs, Cameras	no	Sequential Parallel Interleaved Cooperative	Three-layer model Ontology + DBNs + BN	Leave-one-out CV	- Accuracy rate	-88.89%
(Chen and Tong, 2014)	Same as (Hsu et al., 2010)	Same as (Hsu et al., 2010)	yes	Sequential Parallel Cooperative	HMM and CRF on multi-occupant activity recognition	3-fold CV	Average accuracy	-HMM: 75,77%, CRF: 75,38% (recognition) -HMM: 84,19%, CRF: 82,88% (data association)
					Same models but in the context of multi-label classification measurement	Hold-out method	-Average accuracy -Average error rate -Average precision	-HMM:97.40%, CRF:97.25% -HMM:2.6%, CRF:2.75% -HMM:80.03%, CRF:80.05% -HMM:81.92%, CRF:79.91% -HMM:40.48%, CRF:39.99%

							-Average recall -Average f score	
(Prossegger and Bouchachia, 2014)	ARAS	Same as (Alemdar et al., 2013)	no	Sequential Parallel	Incremental decision trees (E-ID5R)	-Not mentioned	- Accuracy rate	-40% for House A -82% for House B
(Tunca et al., 2014)	ARAS	Same as (Alemdar et al., 2013)	yes	Sequential Parallel Cooperative	1 HMM, 1 MLP, 1 TDNN 1 KNN and 1 DT, (combined label for multi-occupant activities)	Leave-one-out CV	-Accuracy rate/Average f score (both metrics' results are given for each model respectively)	-House A Occupant 1: 67.5/ 67.4, 72, 72.3/ 71.8, 78.3/ 77.4, 68.3/ 67.1 and 72/ 71.8 Occupant 2: 56.8/ 58.9, 56.4/ 51.9, 65.3/ 60.7, 58.8/ 45.7 and 58.4/ 53.4 -House B Occupant 1: 81.3/ 83.3, 82.4 / 78.9, 87.0/ 85.3, 83.1/ 76.7 and 85.3/ 83.6 Occupant 2: 84.6/ 84.2, 81.7/ 79.5, 87.2/ 83.3, 79.2/ 73.9 and 87.4/ 81.8
(Afrin Emi and Stankovic, 2015)	-ARAS	-Same as (Alemdar et al., 2013)	yes	Sequential Parallel	Active learning + Domain knowledge about activities	-Not mentioned	-Average accuracy	-87% for House A and 95.3% for House B.
	- twor.2009 and two.summer.2009 of CASAS	- Motion, door, light, item, water flow, burner and motion, door, light,						-97.34% and 98.15% respectively.

		item, temperature, electricity respectively						
(Roy et al., 2013)(Roy et al., 2016)	PUCK dataset[4]	-Embedded sensors in each occupant's Smartphone (i.e. gyroscope, accelerometer) -Motion -Object	no	Sequential Parallel Cooperative	Layered approach (individual HMM to infer location of each occupant+ CHMM to infer multi-occupant activities)	10-fold CV	- Accuracy rate	-85%

[4] http://ailab.wsu.edu/casas/datasets/puck.zip

II.2.1 Wearable sensors

Wearable sensors refer to a kind of sensors that the occupants wear on the body or in their clothes. These sensors are either stored compactly in a single device that occupants can carry with them (e.g. embedded sensors in a Smartphone as shown in figure 4) (Roy et al., 2013) (Roy et al., 2016) or distributed over the body (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b). Figure 5 shows the wearable sensor platform (a) presented in (Wang et al., 2009) (Wang et al., 2011). In these studies sensors' layout allows specific measurements as voice interaction among occupants (i.e. audio recorder on each occupant's shoulder (b)), temperature, humidity, light and movement of hands (i.e. extracted from an iMote2 with`ITS400 (c) placed on each occupant's forearm), touched object by each occupant's hand (i.e. by the use of an RFID reader on each occupant's hand (d) and RFID tags on objects) and occupant's location at time.

Generally, researchers working on multi-occupancy problems tend to use wearable sensors to reduce the problem complexity as these types of sensors can address the data association problem. However, it is often the case that SH systems ignore ergonomic requirements. The data obtained from wearable sensors mainly provides information about the posture and movement of a person. This data is typically used to recognize types of movement, such as running or, walking and sitting (Roy et al., 2013) (Roy et al., 2016).

Although these sensors offer the possibility of capturing fine-grained observations they cause inconvenience and are not appropriate for SHs requiring privacy and comfort. Furthermore, this type of sensors may limit the body movements of an occupant and is inappropriate for people, especially for elderly people, not willing to wear them, tend to forget to wear them, or let the device’s power source die.

Figure 4: Example of embedded sensors in a Smartphone

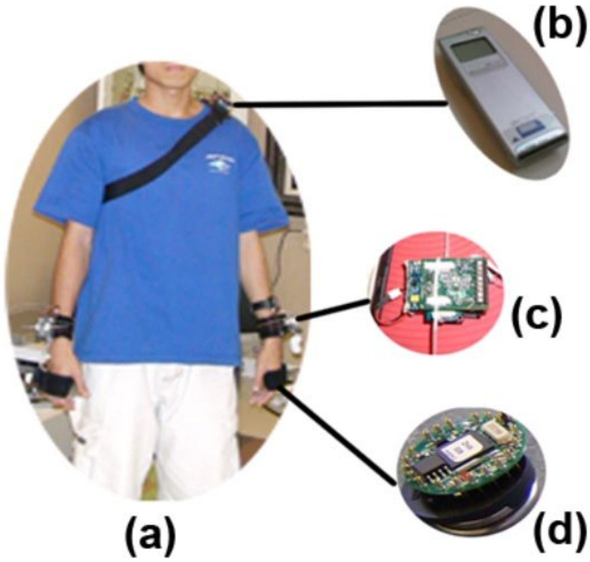

Figure 5: Example of a wearable sensor platform

II.2.2 Infrastructure sensors

Pervasive infrastructure sensors are connected though a wireless sensor network which consists of a collection of small network nodes as shown in (Alemdar et al., 2013) (Cook et al., 2009). Each node communicates wirelessly with the other nodes in the network. In a SH, a large variety of infrastructure sensors can be incorporated in the network nodes such as Contact Switches

Sensors (CSS) on doors, cupboards, item sensors on objects, light switch sensors, temperature sensors, burner sensor, water flow sensors, phone sensor, pressure mats to measure sitting on a couch or lying in bed, mercury contacts for the movement of objects such as drawers and PIR sensors to detect motion in a specific area. Figure 6 shows an example of infrastructure sensors deployed in a kitchen (red dots represent PIRs, yellow star indicate a water flow sensor, orange triangles indicate burner sensors and blue rectangles state for CSS). Given the variety of sensor types and activities, selecting the most suitable set of sensors in the deployment is an important task as shown in (Tunca et al., 2014). In this context, researchers face challenges while selecting the correct type and the required number of sensors targeted to the recognition of specific ADLs in order to have a consistent activity data (Tunca et al., 2014). The latter problem is closely related to the sensor characteristics and the occupant preferences (Tunca et al., 2014). Deploying embedding sensors on a multitude of daily living objects (e.g. microwave, drawers..), on the walls and on the ceilings is very challenging due to operational costs and battery life issues. Moreover, an increase in the number of sensors beyond the designated requirements could affect daily activities and consequently, intrude into the occupant' daily lives (Tunca et al., 2014).

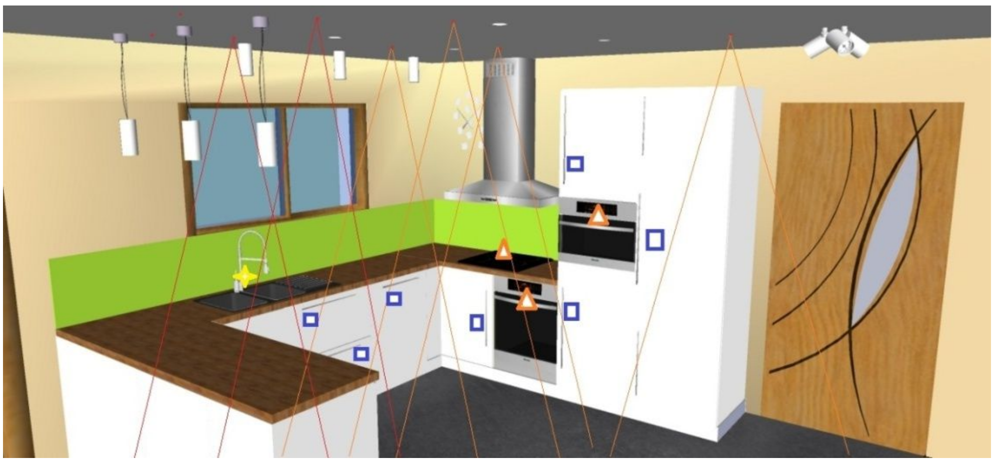

Figure 6: Example of infrastructure sensors deployment in a kitchen

The use of this type of sensors implies designing specific solutions for data association as shown in (Crandall and Cook, 2008a) (Crandall and Cook, 2008b) (Crandall and Cook, 2010) (Hsu et al., 2010) (Wilson and Atkeson, 2005) (Cook et al., 2010) (Alemdar et al., 2013) (Chen and Tong, 2014). For instance, Wilson and Atkeson (Wilson and Atkeson, 2005) used infrastructure and non-intrusive sensors to monitor the occupants at home. They studied the effect of sensor settings on the accuracy of occupant identification. Three types of configurations were defined: *normal*, *extra* and *fewer* configurations. The *normal configuration* contains one motion detector, one contact switch and one pressure mat for each room. The *extra* configuration contains three motion detectors, three contact switches, and three pressure mats per room. The *fewer* configurations contain only one motion detector per room. They found that the extra configuration achieves better accuracy regardless of the number of occupants. In contrast, with the fewer configurations, sensor observations do not provide enough information to the model to clearly identify the individuals and the model can be confused for a long period of time before it becomes able to distinguish between occupants.

Lu et al. (Lu et al., 2008) classify sensors into seamless and seamed ones following ergonomic criteria. They suggest taking advantage of many seamless sensors in the living space. In fact, by decreasing the number of seamed sensors, behavior of people will be not much impacted. The authors also claim that developing a passive solution secures a clean design that separates technology from the smart space and consequently makes the space as natural as possible.

II.3 Datasets for Multi-Occupant Activity Recognition

Usually HAR systems are developed and evaluated using datasets. Publicly available datasets are important for the research community to create standardized test beds which could be used for evaluating the performance of activity recognition algorithms and for comparison purposes. Among the benchmark datasets that are freely available, there exist many single-occupant ones used in (Riboni et al., 2011) (Sarkar et al., 2010) (Kasteren et al., 2008) (Kasteren et al., 2010) (Kasteren et al., 2011).

However, there is a real need for datasets collected from houses with multiple occupants. The CASAS group has collected several multi-occupant activity datasets: "twor.2009"[5], "twor.summer.2009"[6], "twor.2010"[7], "tulum"[8], "tulum2"[9],

[5]http://ailab.wsu.edu/casas/datasets/twor.2009.zip
[6]http://ailab.wsu.edu/casas/datasets/twor.summer.2009.zip
[7]http://ailab.wsu.edu/casas/datasets/twor.2010.zip
[8]http://ailab.wsu.edu/casas/datasets/tulum.zip
[9]http://ailab.wsu.edu/casas/datasets/tulum2.zip

"cairo"[10] and "Multiresident ADLs"[11]. Likewise, the (Activity Recognition with Ambient Sensing) ARAS[12] group has collected a multi-occupant dataset, named ARAS, which includes two datasets, House A and House B. To the best of our knowledge, these datasets are the only ones publicly available recorded from multiple occupants using pervasive sensors.

Table 2 summarizes the characteristics of the multi-occupant datasets which will be described further in the following sections.

Table 2: Characteristics of ARAS and CASAS Multi-Occupant Datasets

Dataset	# of occupants	Duration	# of Sensors	# of ADL	# of sensor events	Environment	Scripted	Annotation medium
House A of ARAS	1 pair	1 month (continuous)	20	27	2 592 000	Real house	No	GUI
House B of ARAS	1 pair	1 month (continuous)	30	27	2 592 000	Real house	No	GUI
"Multiresident ADLs"	26 pairs	Spread over 2months	37	15	17 258	Lab.	Yes	diaries
"twor.2009"	1 pair	Continuous period of 2 months	71	9	137 789	Lab.	No	diaries
"twor.summer.2009"	1 pair	Continuous period of 2 months	86	8	772 544	Lab.	No	diaries
"twor.2010"	1 pair	2009-2010 academic year	87	13	2 804 813	Lab.	No	diaries
"tulum"	1 pair	4 months (Several days are missing)	20	9	486 912	Lab.	No	diaries
"tulum2"	1 pair	2009-2010 academic year	36	15	1 085 902	Lab.	No	diaries

[10]http://ailab.wsu.edu/casas/datasets/cairo.zip
[11]http://ailab.wsu.edu/casas/datasetdlmr.zip
[12]http://www.cmpe.boun.edu.tr/aras/

"cairo"	1 pair +1 pet	Continuous period of 2 months	32	11	726 534	Lab.	No	diaries

II.3.1 CASAS multi-occupant datasets

Multi-occupant datasets of CASAS were collected in the Washington State University (WSU) smart apartment test bed. Multi-occupant activities were obtained using clinical questionnaires (Reisberg et al., 2001). Activities were annotated by recording the start and end time of the activities via a handwritten diary. We can distinguish two types of datasets:(i) unscripted activity datasets like "twor.2009", "twor.summer.2009", "twor.2010", "tulum", "tulum2", and "cairo" and (ii) scripted activity dataset like "Multiresident ADLs". Activities considered in the unscripted multi-occupant datasets and the scripted multi-occupant one are listed in table 3 and table 4 respectively.

Table 3: Activities of the Unscripted CASAS Multi-occupant Datasets

"twor.2009"	"twor.summer.2009"	"twor.2010"	"tulum"	"tulum2"	"cairo"
-Clean -Meal preparation -Bed to toilet -Personal hygiene -Sleep -Work -Study -Wash bathtub -Watch TV	-Bed to toilet -Cleaning -Cooking -Grooming -Shower -Sleep -Wake up -Work.	-Bathing -Bed to toilet -Eating -Enter home -Housekeeping -Leave home -Meal preparation -Personal hygiene -Sleep -Not sleeping in bed -Wandering in room -Watch TV -Work	-Cook breakfast -Cook lunch -Enter home -Group meeting -Leave home -Eat breakfast -Snack -Wash dishes -Watch TV	-Bathing -Bed to toilet -Eating -Enter home -Leave home -Meal preparation -Personal hygiene -Sleeping in bed -Wash dishes -Watch TV -Work bedroom 1 -Work bedroom 2 -Work living room -Work table -Yoga	-Bed to toilet -Breakfast -Sleep -Wake -Work in office -Dinner -Laundry -Leave home -Lunch -Night wandering -Take medicine

Table 4: Activities covered by the scripted "Multiresident ADLs" of CASAS

Individual	***Cooperative***
-Filling medication dispenser -Hanging up clothes -Reading magazine -Sweeping floor -Setting the table -Watering plants -Preparing dinner	-Moving furniture -Playing checkers -Paying bills -Gathering and packing -picnic food

The WSU smart apartment test bed is equipped with many types of sensors: motion sensors, door sensors, temperature sensors, light switch sensors, water flow sensors, burner sensor, phone sensor, and item sensors. A summary of the type and the number of sensors used for recording each CASAS dataset is shown in table 5.

Table 5: Sensors used in CASAS multi-occupant datasets

	"Multiresident ADLs"	"twor. 2009"	"twor. summer. 2009"	"twor. 2010"	"tulum"	"tulum2"	"cairo"
Motion sensors	27	51	51	51	18	31	27
Door sensors	8	9	15	15			
Light sensors		7	10	11			
Item sensors	2	1	4	4			
Temperature sensors			5	5	2	5	5
Electricity sensors			1	1			
Water flow sensors		2					
Burner sensors		1					

In the following, we give some details of the scripted and unscripted datasets:

II.3.1.1 Unscripted Multi-occupant datasets

To the best of our knowledge, "twor.2009", "twor.summer.2009", "twor.2010", "tulum", "tulum2" and "cairo" have not been much used in multi-occupant activity recognition research. Each of these datasets was collected through a pair of occupants who performed unscripted activities. Specifically, "tulum","tulum2" represent activity data of a married couple, where as the "cairo" dataset consists of three types of data: the activity data of a volunteer adult couple, the motion data related to their dog and the data related to their children who come sometimes to visit. All these datasets account for intra-subject variability.

Although, these datasets stemmed from a laboratory on a voluntary basis, they were recorded continuously in time. The recording time for "twor.2009", "twor.summer.2009", and "cairo" was approximately 2 months, for "tulum" 4 months and for both"twor.2010" and"tulum2" approximately one year. Multi-occupant activities in these datasets are described by records of the form (Date, Time, SensorID, Value). Each activity is delimited by specific markers: (OccupantID_ActivityName Begin) and (OccupantID_ActivityName End).

Regarding the annotation of all unscripted datasets the OccupantID performing the activity is missing in the annotation of many activities. For example considering the "two.2009" dataset, the annotation is complete for the activities bed to toilet, personal hygiene, sleep and work while in the annotation of the remaining ones the OccupantID performing the activity is not specified (e.g. clean begin). Moreover, among all the activities registered in the different datasets, some are performed in parallel by the occupants (e.g. considering "two.2009" dataset both Occupant 1 and Occupant 2 are sleeping/working or while occupant1 is sleeping, Occupant 2 start performing his/her personal hygiene) while the others are performed in a sequential manner by the occupants and are consecutively represented in the dataset.

II.3.1.2 Scripted Multi-occupant dataset

"Multiresident ADLs" collection has been used in many studies (Hsu et al., 2010) (Chiang et al., 2010) (Cook et al., 2010) (Singla et al., 2010) (Chen and Tong, 2014). It was generated in a laboratory setting and therefore it does not fully reflect on real-world scenarios. " Multiresident ADLs" was collected through 26 pairs of volunteers who performed scripted activities. Such activities are pre-determined and were repeatedly performed. This collection accounts for inter-subject variability, yet it is not sufficient for explaining real-world situations. This dataset was not recorded continuously in the time, and instead it was spread over two months. Multi-occupant activities in "Multiresident ADLs" come in the format of (Date, Time, SensorID, Value, OccupantID, TaskID). A full description of this dataset can be found in (Singla et al., 2010).

II.3.2 ARAS collection

ARAS data was collected from two pairs of occupants performing a large variety of activities (Alemdar et al., 2013). The first pair consists of two males while the second is a couple. This collection of two-home dataset offers a better opportunity to study and compare activity recognition algorithms more realistically. ARAS data accounts for intra-subject variability and do not account for the inter-subject one. It reflects on the natural behavior of the occupants during 2 months. An important feature of ARAS data is that it contains a large variety of human activities and a large number of activity occurrences. Activities and sensors considered by ARAS are presented in table 6 and table 7 respectively. Annotation of the activities was achieved by the occupants themselves using a simple graphical user interface (GUI). Several instances of GUI applications were placed in the most convenient places in the houses. This way of doing is more accurate than using a diary.

Table 6: Activities simulated by ARAS

House A & House B	
-Other -Going Out -Preparing Breakfast -Having Breakfast -Preparing Lunch -Having Lunch -Preparing Dinner -Having Dinner -Washing Dishes -Having Snack	-Having Shower -Toileting -Napping -Using Internet -Reading Book -Shaving -Brushing Teeth -Talking on the phone -Listening to Music -Cleaning

-Sleeping -Watching TV -Studying	-Having Conversation -Having Guest -Changing Clothes -Laundry

Table 7: Sensor infrastructure used by ARAS

House A		*House B*	
1	Wardrobe photocell	2	Kitchen cupboards CSs
1	Convertible Couch photocell (Occupant 2's bed)	1	House DCS
		2	Wardrobe DCSs
1	TV infrared receiver	1	Shower Cabinet DCS
2	Couch force sensors	1	Tap distance sensor
2	Chair proximity sensors	3	Chair force sensors
1	Fridge photocell	11	Fridge photocell
1	Kitchen Drawer photocell	2	Kitchen Drawer photocell
1	Wardrobe photocell	2	Couch pressure mat
1	Bathroom Cabinet photocell	1	Bed pressure mat
1	House DCS[13]	1	Armchair pressure mat
1	Bathroom DCS	1	Bathroom Door sonar distance
1	Shower Cabinet DCS	1	Kitchen sonar distance
1	Hall sonar distance	1	Closet sonar distance
1	Kitchen sonar distance		
1	Tap proximity sensor		
1	Water Closet proximity sensor		
1	Kitchen temperature sensor		
1	Bed force sensor		

ARAS collection offers the advantage of being ready to use. Each day of recordings consists of a 22 x 86400 matrix which is stored in a file. In all, the collection consists of 30 files for each dataset. The first 20 columns are the sensor binary values, fired/not fired; column 21 and 22 contain the activity labels for Occupant 1 and Occupant 2 respectively. Using a constant time intervals to discretize the data allows representing the sensors readings in time slices. This representation leads to a better discrimination between activities as reported by the studies in (Kasteren et al., 2008) (Kasteren et al., 2010) (Kasteren et al., 2011) on

[13] DCS stands for Door Contact Sensor

other datasets. For instance, in (Kasteren et al., 2010) two experiments were run. In the first experiment different lengths of the time slice were tested including (i.e. 1s, 10s, 30s, 60s, 300s and 600s) finding that short time slices produced better recognition results. 30s and 60s produced the best results. In the second experiment, Naive Bayes (NB), HMM, Hidden-Semi Markov Model (HSMM) and CRF models were compared using a number of feature representations which consisted of raw, change point and last representation. The change point representation produced the best results. It was found that the recognition performance of the activity model is strongly influenced by the time slice length and the feature representation.

II.4 Computational Models for Multi-Occupancy Activity Recognition

As mentioned earlier, a large variety of computational models was investigated in the context of single-occupancy ADLs ranging from probabilistic models to standard data mining and machine learning models like neural networks, decision trees, ontologies, etc. In the case of multi-occupancy, however, no such diversity of models exists. Almost all of the proposed models are essentially probabilistic based on graphical models. This conclusion can easily be observed in table 1 above which illustrates also a summary of a set of representative research studies covering the sensors used, the type of activities covered, the models used, and the evaluation metrics as well the results obtained.

In the following sections we first describe the mainly applied classes of models in the context of multi-occupancy that is probabilistic models, and association rule mining:

II.4.1 Probabilistic Models

As sensor readings are usually noisy and activities are typically performed in a nondeterministic fashion, probabilistic learning-based methods are frequently used in activity recognition to model sensor readings.

Probabilistic models are advantageous for problems such as activity recognition, because they allow us to deal with the noise and uncertainty in a principled manner. We present in the following, probabilistic models applied by state-of-the-art studies in multi-occupancy. They can be categorized into generative and discriminative models.

II.4.1.1 Generative Models

Generative models define the joint probability distribution and can be used to generate (sample) data from such distribution or to perform inference given a novel sequence of observations (Kasteren, 2011). We focus here only on models that have been applied in the context of multi-occupancy. Some of the generative models used by related work studies on multi-occupant HAR are graphical models such as NB, Dynamic Bayesian Network (DBN), HMM and variants such as PHMM and CHMM.

Graphical models are the most popular computational models used in activity recognition in general. As their names indicate, graphical models are probabilistic models having the structure of graphs that represent conditional dependence between nodes which are random variables.

Graphical models are defined as probability distributions that factorize according to a graph (Sutton and McCallum, 2006). The goal is to infer a matching sequence of hidden states that maximizes the probability of the activities given some sensor readings. In standard graphical model representation, square nodes represent discrete variables, circle nodes represent continuous variables. Observed nodes are shown as shaded nodes while the hidden nodes are blank.

In the following we only present NB and DBN. HMM and variants (i.e. PHMM and CHMM) are described in Chapter III.

a. Naive Bayes (NB)

The NB model can be considered as one of the most simplistic probabilistic models. It is a restricted version of the Bayesian network (BN). In fact, the NB model assumes all data points (e.g. the events of sensors in the case of activity recognition) are independently and identically distributed. The class nodes have no parents and the attribute nodes are not connected. Moreover, NB does not take into account any temporal relations between data points. The joint probability of observations and labels can be factorized as:

First, we note:

$$X_{k:l} = (X_k, \dots, X_l) \quad for \;\; k \le l$$

And

$$(X_{k:l} = x_{k:l}) \equiv (X_k = x_k, \dots, X_l = x_l) \quad for \; k \le l$$

And to simplify the ratings we will write:

$$P(X,Y) = P(X_{1:t} = x_{1:t}, Y_{1:t} = y_{1:t}) = \prod_{t=1}^{T} P(X_{t=}x_t \,|\, Y_{t=}y_t) P(Y_{t=}y_t)$$

$$= \prod_{t=1}^{T} P(x_t \,|\, y_t) P(y_t) \qquad \text{(II.1)}$$

where $P(Y_t = y_t) = P(y_t)$, for t=1…*T*, is a prior probability over activities. To compute the conditional probability of labeled data *(X,Y)* in a straightforward way, we assume independence between input features given the input labels. The probability can then be written as follows:

$$P\left(X_t^i = x_t^i \middle| Y_t = y_t\right) = \prod_{i=1}^{N} P(x_t^i | y_t) \qquad \text{(II.2)}$$

In our setting, the set *X* represents the sensor data. X = (X_1, X_2,...,X_t) where X_i is a vector of length N for i=1,....,t. *Y* represents the set of activities as shown in figure 7. *N* is the total number of sensors. The activity y_t is independent of previous activities $y_{1:t-1}$.

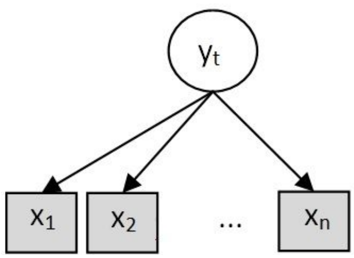

Figure 7: A NB representation

Activity recognition can be considered as a classification problem where activities as regarded as classes (Van Laerhoven et al., 2003) (Liao and Ji, 2009).

NB and in general conventional BNs are not suitable for modeling temporal processes because directed arcs of the network do not give any information about the time. In order to overcome this limitation, DBNs was proposed as an upgrade of BNs.

b. Dynamic Bayesian Network (DBN)

DBNs are designed to deal with temporal processes (time series). A DBN results from extending BN by sequencing interlinked time-sliced instances of the BN as shown in figure 8 (Where *N* is the total number of sensors, y_t is dependent of the activity at y_{t-1} and each activity is represented by all sensor values at time). When DBN is applied to activity modeling, the observables, X_t , correspond to the sensor readings, while the unobservable variables, Y_t , correspond to the activities. A state at a specific time t depends on the previous states.

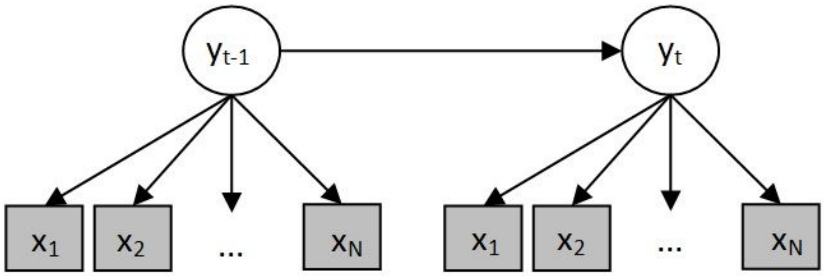

Figure 8: A DBN representation

Formally, a DBN (Sanghai et al., 2005) is defined as a pair of BNs *(B1,B→)*, where B_1 is prior which defines the initial distribution $p(Z_1)$ and $B\rightarrow$ is a 2-time-slice BN defining the transition distribution $p(Z_t|Z_{t-1})$ via a directed acyclic graph:

$$P(Z_t|Z_{t-1}) = \prod_{i=1}^{N} P(Z_t^i|Parents(Z_t^i)) \tag{II.3}$$

$$P(Z_t) = \prod_{t=1}^{T}\prod_{i=1}^{N} P(Z_t^i|Parents(Z_t^i)) \tag{II.4}$$

Z^i_t is a node at time slice t, it can be a hidden node, an observation node, or a control node (optional), while *Parents*(Z^i_t) are parent nodes of Z^i_t and can be at either time slice t or t-1.

II.4.1.2 Discriminative Models

In contrast to generative models where we attempt to model the joint probability distribution of paired observations and activity sequences *P(Y,X)*, in discriminative we rather attempt to directly model the conditional probabilities of the activities given the sequence of observations *P(Y|X)*. Moreover, generative models assume that the observations are independent which is not always satisfied.

In the following we will present some of the discriminative models used in multi-occupant activity recognition. Note that CRF and Factorial Conditional Random Field (FCRF) represent graphical models hence the same notation holds for their graphical representation:

a. Decision Trees (DTs)

A Decision Tree (DT) is used to model the relation between input data and the corresponding output. A DT can be used for either classification if the output is discrete indicating class labels or regression if the output is continuous. A classification tree consists of nodes that represent features and branches that represent the values of the features. The leaf nodes represent the class labels. When a DT is applied for activity recognition the leaf nodes represent the activities while the features represent the set of sensors as shown in figure 9.

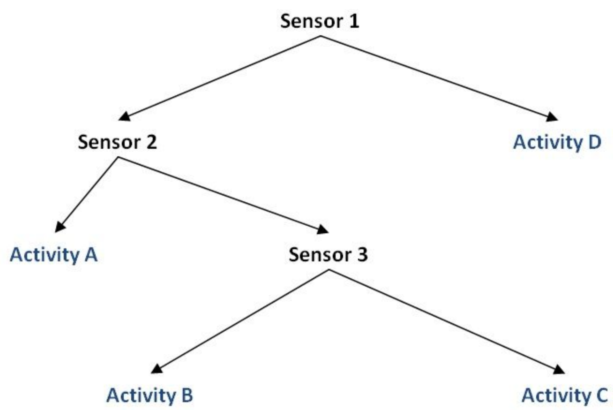

Figure 9: Example of a decision tree. Leaf nodes represent activities

DTs are built through an induction process using a training dataset. Many induction algorithms have been devised such as TDIDT/ID3, C4.5 and CART, MARS and CHAID. Some algorithms like C4.5 and CART execute two phases: growing and pruning of the tree, while others only grow the tree (Maimon and Rokach, 2005).

DTs have been widely used to recognize activities in a single-occupant setting as shown in (Isoda et al., 2004) (Ravi et al., 2005). In both studies, C4.5 was used to generate a DT for classifying occupant's activities. Specifically, in (Isoda et al., 2004), the activity data was collected from pressure sensors on the floor for locating the occupant and RFID tags on the objects (e.g. gas hobs, cupboard). The evaluation of the classifier on kitchen activities achieved an accuracy of 90-100% depending on the size of the learning data used. The activities considered in (Ravi et al., 2005) were: standing, walking, running, climbing up stairs, climbing down stairs, sit ups, vacuuming and brushing teeth. The results showed that C4.5 achieved 97.29% when trained and tested on data from the same user over many days. An accuracy of 98.53% was achieved when C4.5 was trained and tested on data stemming from many users and over many days and 77.95% when trained and tested on data stemming from the same day.

DTs were not much investigated in multi-occupant activity recognition. To best of our knowledge (Prossegger and Bouchachia, 2014) (Tunca et al., 2014) represent the only studies which applied DTs to model ADLs in a multi-occupant context. In (Prossegger and Bouchachia, 2014), an extension of ID5R called E-ID5R induces a DT incrementally to accommodate new instances and new activities as they become available over time. Their E-ID5R extends the leaf nodes to represent single or multiple activities (i.e. parallel activities are recognized). To evaluate the proposed algorithms, the ARAS dataset which is a real world multi-occupant dataset stemming from two houses was used. E-ID5R performs differently on activities of both houses: for house A whose data is quite challenging, the classification rate was model (40%), while for house B the rate approached 82%.

To recognize parallel activities that is both the activity performed by the Occupant 1 and the activity performed by the Occupant 2 at time, (Tunca et al., 2014) applied a combined activity label based DT approach (see details in II.5.5).

b. Artificial Neural Network (ANN) and variants

Artificial Neural Network (ANN) is a computing model made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs (Caudill, 1987). The fundamental processing elements of an ANN are artificial neurons (or nodes) which are interconnected by weighted links forming layers as shown in figure 10. Typically in an ANN there is one input layer that varies depending on the complexity of the problem at hand (Murata et al., 1994). Neurons transform the weighted input into output using an activation function which can take different forms (linear and non-linear). The process by which the weights are adjusted is called learning. A number of non-linear ANNs are known to perform as function approximators. There are various parameters that define the architecture of a neural network: the connection type (e.g. feed-forward networks, recurrent neural networks etc.), and activation functions (e.g. sigmoidal, hyperbolic tangent, etc.). Because of these shaping parameters, there are different types of ANNs (e.g. Multi-Layer Perceptron (MLP), Echo State Network (ESN), Radial basis function network (RBFN), Bolzmann machine, etc.).

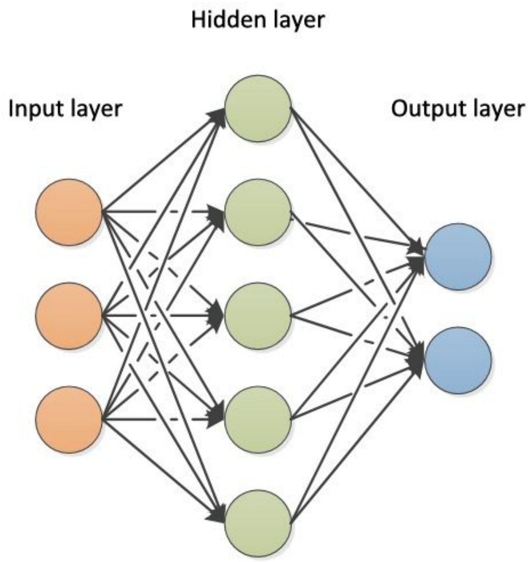

Figure 10: Architecture of a 3-layer feed forward network

ANNs can be applied to a number of SH problems such as activity classification, control of appliances, novelty and anomaly recognition and prediction of activities. In relation to health monitoring in SH, ANNs were used to diagnose and monitor chronic diseases as well to built medical decision support systems (Khan et al., 2001) (Lisboa and Taktak, 2006) (Er et al., 2010). Although ANNs have been widely used in single-occupant HAR, only variants of this model have been used in multi-occupant HAR (i.e. MLP and Time-Delay Neural Network (TDNN)) as shown below:

MLP is a feed forward ANN model that maps sets of input data onto a set of appropriate outputs. In HAR the set of inputs consists of sensory data while the outputs represent the activities. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. The two main activation functions used in current applications are both sigmoid, and are described by:

$$y(v_i) = tanh\ (v_i) \tag{II.5}$$

$$y(v_i) = (1 + e^{-v_i})^{-1} \tag{II.6}$$

In which the former function is a hyperbolic tangent which ranges from -1 to 1, and the latter, the logistic function, is similar in shape but ranges from 0 to 1. Here y_i is the output of the ith node (neuron) and v_i is the weighted sum of the input synapses.

TDNN is an ANN architecture whose primary purpose is to work on sequential data (i.e. capable of considering the sequential nature of a time series) . The TDNN units recognize features independent of time-shift (i.e. sequence position) and usually form part of a larger pattern recognition system. Converting sensory data into a stream of classified activity labels for HAR. An input signal is augmented with delayed copies as other inputs, the neural network is time-shift invariant since it has no internal state.

MLP and TDNN were applied to recognize multi-occupant ADLs in (Tunca et al., 2014). Their performances were also compared to other classifiers that is DT, KNN and HMM. These models were applied on ARAS datasets. Experiments' results reported that HMM and TDNN, which are sequential methods, perform better than the other three classifiers for House A. For House B, HMM provides fairly good results, in recognizing most of the activities, while the other classifiers confuse some key activities.

c. Conditional Random Field (CRF)

The linear-chain CRF model is one of the most popular discriminative model for dealing with sequential data. It is more flexible compared to HMM, because it does not assume any independence among the observation sequences. Like HMM, CRF is applied to determine the most likely sequence of states given the sequence of observations.

As shown in figure 11, a linear-chain CRF is an undirected acyclic graph where the hidden sate y_t depends only on the previous state y_{t-1} and the observation x_t depends only on the hidden state y_t.

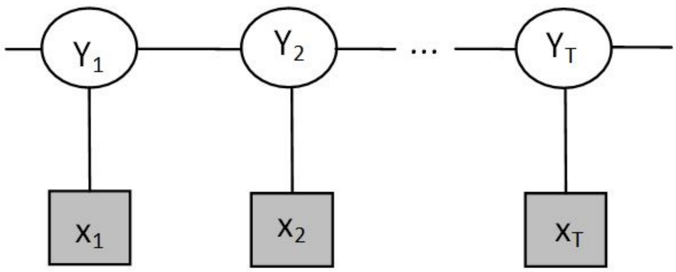

Figure 11: The linear-chain CRF model

The conditional probability distribution is defined as a multiplication of feature functions exponents:

$$P(Y|X) = \frac{1}{Z(X)} \prod_{t=1}^{T} exp \sum_{k=1}^{K} \lambda_k f_k(y_t, y_{t-1}, x_t) \tag{II.7}$$

Here T is the number of observations and K is the number of feature functions used to approximate the probability distribution, and λ_k $(k=1...K)$ are learning weights associated with the feature functions $f_k(y_t, y_{t-1}, x_t)$ which are estimated by training. The expression $\lambda_k f_k(y_t, y_{t-1}, x_t)$ is known as the energy function, while the exponential of the energy function is known as the potential function (Bishop, 2006).The quantity $Z(X)$ is a normalization term so that the probability distribution adds up to 1 resulting in a proper conditional probability as follows:

$$Z(X) = \sum_{y} \left\{ \prod_{t=1}^{T} exp \sum_{k=1}^{K} \lambda_k f_k(y_t, y_{t-1}, x_t) \right\} \tag{II.8}$$

Figure 12 shows how CRF is applied for activity recognition in the context of "adlnormal" activity data of CASAS. Activities are represented as hidden states and the sensor readings correspond to the observables.

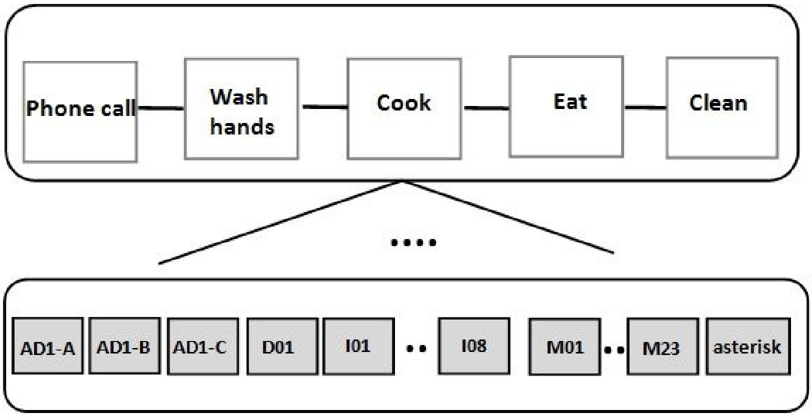

Figure 12: Representation of a global linear-chain CRF

d. Factorial Conditional Random Field (FCRF)

FCRF combines many linear chain CRFs (called chains) by linking not only the hidden states of each chain to input, but also linking the hidden states of the chains to result in co-temporal connections (Sutton et al., 2007). The co-temporal connections allow an efficient representation of the interactions between the chains.

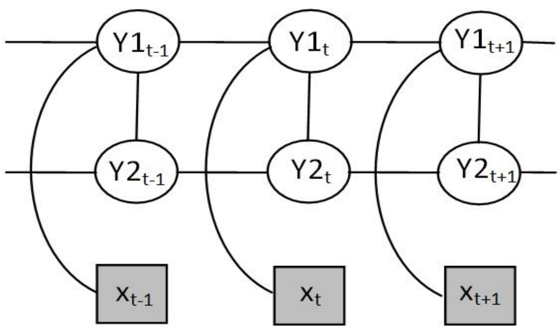

Figure 13: Representation of FCRF consisting of two CRF chains

The application of FCRF for modeling multi-occupant activities is straight forward. We can think of using one chain for each occupant, where the hidden states of the chains representing the activities are co-temporally connected to model the interaction (Wang et al., 2011). Figure 13 illustrates an FCRF model as a combination of two chains, each representing an occupant. The two sequences $\{y1_{t-1}, y1_t, y1_{t+1}\}$, $\{y2_{t-1}, y2_t, y2_{t+1}\}$ are the activities at time t-1, t, t+1 of Occupant 1 and Occupant 2 respectively. The corresponding sequence of observations $\{x_{t-1}, x_t, x_{t+1}\}$ represent sensor readings at the time steps t-1, t, t+1. FCRF is given by the following posterior probability:

$$P(Y|X) =$$

$$\frac{1}{Z(X)}\left[\prod_{t=1}^{T}\prod_{i,j}^{A} exp\left[\sum_{k=1}^{K}\lambda_k f_k(y_{it}, y_{jt-1}, X)\right]\right]$$

$$\left[\prod_{t=1}^{T-1}\prod_{i}^{A} exp\left[\sum_{k=1}^{K}\lambda_k f_k(y_{it}, y_{it+1}, X)\right]\right]\left[\prod_{t=1}^{T-1}\prod_{i}^{A} exp\left[\sum_{k=1}^{K}\lambda_k f_k(y_{it}, y_{it}, X)\right]\right] \quad \text{(II.9)}$$

where A is the set of activities and Y and X indicate the set of hidden and observable state sequences. That is, $Y=\{Y_1, Y_2, ..., Y_T\}$ where $Y_t=\{y_{1t}, y_{2t}, ..., y_{At}\}$ and Y it represents the state of the ith activity at time t. The observable state sequence X is defined in a similar way. $Z(X)$ is a normalization factor obtained over X.

Given an observation sequence X, to find the most likely sequence of activities states, the Maximum-A-Priori (MAP) algorithm is applied once the marginal probability of all node pairs is computed. Actually there exist many inference algorithms like the forward-backward algorithm, loopy belief propagation, mean field free energy, and junction tree (Wang et al., 2011) (Sutton et al., 2007).

II.4.2 Emerging Patterns (EPs)

Association rule mining is about finding interesting relations between features in data. Such relations are rules whose right-hand-side and left-hand-side are frequent itemsets (i.e. set of features). Itemsets can also be used to distinguish between datasets and in such case they are called Emerging Patterns (EPs). Thus, EPs can be considered as itemsets with support that changes significantly between datasets.

In the context of activity recognition, EPs are applied to model the activities using the discriminating features. An EP of an activity is the set of features which are the most discriminating for that activity. The set of EPs of an activity form the corresponding *activity model*. The set of features of an activity is selected as EP if the frequency count of such features changes from that activity's instances to the rest of other activities' instances.

A data instance refers to all observations that are part of an activity during a continuous period of time. The support of an itemset *V*, in a dataset *X*, is given as:

$$S_X(V) = \sigma_X(V)/|X| \tag{II.10}$$

where σ_X is the number of instances in *X* that include *V*. |*X*| is the total number of instances in *X*. Using the notion of support, we can compute the growth measure to identify EPs as follows. Given two activities *A* and *B*, the Growth Rate (GR) of an itemset *V* from *A* to *B*, denoted as *GR(V,A,B)*, is given as follows:

$$GR\ (V, A, B) = \begin{cases} 0\ if\ S_A(V) = 0\ and\ S_B(V) = 0 \\ \infty\ if\ S_A(V) = 0\ and\ S_B(V) > 0 \\ \dfrac{S_B(V)}{S_A(V)}\ Otherwise \end{cases} \tag{II.11}$$

An itemset *V* is *EP* of an activity *B* if and only if its GR exceeds a given threshold ρ (i.e. *GR(V,A,B)*>ρ), that is the change from *A* to *B* is significant.

EPs are thoroughly discussed in (Gu et al., 2009a). An example presented therein assumes the activity "cleaning a dining table" and the following itemset "object@cleanser, object@plate, object@wash_cloth, and location@kitchen" is an EP. The authors in (Gu et al., 2009a) (Gu et al., 2009b) apply an efficient algorithm described in (Li et al., 2007) to discover EPs from sequential activity data. Such EPs are used to construct the activity model in a single occupant setting. Using the epSICAR algorithm described in (Gu et al., 2009a), not only sequential activities but also concurrent and interleaved activities can be identified. Going a step further, the authors of (Gu et al., 2009b) apply EPs in a multi-occupant setting by mining EPs for each activity and for each occupant.

II.5 Facets of Multi-Occupant Activity Recognition

As shown in table 1, some studies have focused on solving the data association problem (Crandall and Cook, 2008a) (Crandall and Cook, 2008b) (Crandall and Cook, 2010) (Hsu et al., 2010) (Wilson and Atkeson, 2005) (Cook et al., 2010) (Alemdar et al., 2013) (Chen and Tong, 2014). Some other studies have considered that the data association problem had been already resolved and consequently focused on modeling the activities (Hsu et al., 2010) (Cook et al., 2010) (Chiang et al., 2010) (Singla et al., 2010) (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b) (Lin and Fu, 2007). As the two issues of data association and activity recognition tend to be treated separately, we will discuss in Section II.5.1 the data pre-processing approaches designed for the two problems separately. Section II.5.2 presents the studies dealing with data association. For activity recognition, two methodologies will be discussed. According to the first methodology, activity recognition models for occupants are independent ignoring the interactions between the occupants (Section II.5.3). In the second methodology, the models take the interaction into account (Section II.5.4). Section II.5.5 discusses the application of knowledge-based approaches for multi-occupant activity recognition. Scalability of all the studies is discussed in Section II.5.6. Limitations of the different approaches are summarized in Section II.5.7 Evaluation metrics for assessing the performance of algorithms for both problems are discussed in Section II.5.8 respectively.

II.5.1 Pre-processing methods

All multi-occupant approaches presented in this chapter are data-driven approaches which rely on data to construct the activity model. As a result, these approaches may be sensible to the representation of the activity data. This later is often incomplete, inconsistent, and prone to errors. Hence, generally it is pre-processed with the intention of making the activity recognition problem easier to solve so that: (1) data meets the computational model applied for developing the activity recognition algorithms and (2) efficient use of the raw data is guaranteed through a new representation.

Data pre-processing is an important step in the data mining process. It can consist of several tasks such as cleaning, transformation, normalization, feature extraction and selection. For instance, feature extraction are used in (Crandall and

Cook, 2008a) (Crandall and Cook, 2008b) to generate new features from the raw data. In particular, the date and the time information stamps are used to extract different features like “hour of day”, “part of day”, “day of week”, “hour of day” which are applied to handle the problem of data association. The studies in (Crandall and Cook, 2008a) (Crandall and Cook, 2008b) discuss the impact of the best temporal feature in capturing the differences in behavior between individuals showing that “hour-of-day” enhances significantly the classifier performance in occupant identification. They also show that depending on the facets of the dataset (e.g. the habits of occupants, type of environment, student laboratory or real home), different kinds of features can lead to different classification results. For example, hour-of-the-day is the most discriminating feature, because the dataset was collected from a student laboratory. Furthermore, in comparing the performance of NB and HMM for data association, the studies conclude that feature extraction is not valuable for all types of classifiers and in this case, it is valuable for NB but not for HMM. As a result, in (Crandall and Cook, 2010), the authors applied HMM but without feature extraction to deal with data association.

To check the effect of pre-processing, Hsu et al. (Hsu et al., 2010) investigate three configurations: raw data, environment data and room-level data. The raw data is obtained by removing the date and time from the observations and is represented using the sensor ID combined with its reading value. The environment data consist of all data captured in the house. For the room-level data, a pre-processing method is applied to represent each room by a feature. However, the environment data does not help in discriminating the occupants. This later is “on” if and only if one of the motion sensors in the room is “on”. This feature also does not help in discriminating between the occupants either. The experiments show that the raw data allows obtaining the best recognition results compared to the other two datasets.

In terms of pre-processing methods for activity recognition models, Chiang et al. (Chiang et al., 2010) apply three data pre-processing methods to obtain raw feature, loc-obj feature and loc-obj with locoff feature vectors. The three types of vector are represented as a tuple (event, interaction), where "event" in raw data is an integer indicating a sensor and its state. "Event" in loc-obj and loc-obj with locoff indicates whether it was captured by object sensors (e.g. item and cabinet sensors) or by location sensors (e.g. motion sensors). "Interaction" is only used in loc-obj with locoff to indicate whether the occupants where in the same room or not. The results show that better results are obtained with raw features which is consistent with the previous work in (Hsu et al., 2010). The low performance in the two pre-processing methods may be attributed to only the issue of representing a location sensor by the corresponding index of the room. The model confuses in the case of many activities sharing the same room. Adding all historical data of triggering of the events would better discriminate between activities as reported by the raw feature vector.

On the other hand, Chen and Tong (Chen and Tong, 2014) use the same dataset as in (Hsu et al., 2010) (Chiang et al., 2010), but pre-process it in a different way. An observation is represented as a binary vector whose length corresponds to the number of sensors. At time t, a position in the vector is set to 1 if the i^{th} sensor changed state. However such a representation ignores the date and time as features.

The authors in (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b) investigated a dataset obtained by means of wearable sensors (e.g. 3-axis acceleration, audio, location, and tagged objects). New features are extracted from the raw data like the mean, variance, energy, frequency-domain entropy, correlation, location name, and object name. Likewise, in (Lin and Fu, 2007) the light and motion data serve to derive new features like bright, dim, dark, no light, triggered and non-triggered.

However, some authors like in (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b) (Lin and Fu, 2007) do not compare their pre-processing methods against raw representation to show the effectiveness.

II.5.2 Data Association

To avoid the problem of data association in multi-occupant activity recognition, some studies relied on wearable sensors (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b). Because of the inconvenience of the wearable sensors in some situations, the use of infrastructure sensors has also been investigated. The challenge there is that infrastructure sensors cannot directly identify individuals. In the following, we will discuss the computational approaches used in the context of data association.

Crandall and Cook (Crandall and Cook, 2008a) apply NB on raw data for data association, but obtain low performance. In fact, NB tends to assign activities to the occupant who produced most of the sensor events in the training data, presumably due to imbalance of the training data. But after adding some feature like the temporal feature “hour-of-the-day”, NB shows better discrimination between the occupants. In another study by the same authors (Crandall and Cook, 2008b), HMM is found to outperform NB when using only raw data. Feature extraction is valuable for NB, but does not affect HMM. In a third study by these authors (Crandall and Cook, 2010), the results show that both algorithms perform well on other real-world datasets B&B and TwoR using the same experimental setting described in (Crandall and Cook, 2010) with a slightly better performance for HMM.

To investigate the correlation between data association and activity recognition, CRF is applied in (Hsu et al., 2010). As expected, the quality of data association impacts activity recognition if both are integrated in one system. In this study a two-layer cascade is proposed. Each layer consists of a CRF model. The first layer is designed for data association such that the CRF’s hidden states represented the occupants, while the observables correspond to the sensor events and activity labels. The second CRF in the second layer is dedicated to activity recognition. Thus, the hidden states in the CRF correspond to the activities, while the observables are sensor reading and occupant labels resulting from the previous layer. Likewise, Cook et al. (Cook et al., 2010) construct one HMM model to recognize the occupants followed by another HMM to recognize the activities. The disadvantage of these cascades is that the recognition accuracy depends on the performance of the data associator.

On the other hand, Wilson and Atkeson (Wilson and Atkeson, 2005) propose one motion model for each occupant using a particle filter based approach in order to identify the optimal assignment of sensors to the occupants. They study the impact of varying both the number of occupants and the number of particle filters to accurately identify an occupant on simulated data. The number of occupants varies between 1 and 5 and the number of particle filters varies between 1 and 20 (Wilson and Atkeson, 2005). An insignificant improvement of accuracy is observed after 20 particle filters. Also, more particles are usually required to recognize multiple occupants. Moreover, the accuracy decreases as the number of occupants increases.

II.5.3 Independent Models for Occupants

Many studies address the problem of multi-occupant activity recognition but they neither model interaction among occupants (Hsu et al., 2010) (Wilson and Atkeson, 2005) (Cook et al., 2010) (Singla et al., 2010), nor do they consider real situations where occupants perform separate, interleaved, parallel, or cooperative activities (Cook et al., 2010) (Singla et al., 2010). Often interaction is modeled only in a non-complex setting. The authors in (Hsu et al., 2010) (Wilson and Atkeson, 2005) (Singla et al., 2010) (Lin and Fu, 2007) claimed that multi-occupant activities can be better recognized if individual models for the occupants are learned. In the following we will discuss the approaches used to create independent models for occupants.

Specifically some studies (Wilson and Atkeson, 2005) (Singla et al., 2010) show that motion models can be useful for disambiguation of activities, because people usually tend to follow regular habits. One HMM is used to model each occupant. Likewise, one CRF per occupant is proposed in (Hsu et al., 2010). In this latter study, the accuracy reported is greater than the accuracy in (Singla et al., 2010) using the same benchmark. This seems to confirm that CRF performs better than HMM in handling complex situations. The modeling of activities separately is good when there is less collaboration among the occupants. Thus, if the data contains cooperative activities, the accuracy will be low.

In (Lin and Fu, 2007) separated models for occupants are applied using a layered Bayes network-based architecture which models the interaction between the occupants. Each layer in the model received the results from the previous layer. In the first layer, the input consists of the sensor readings along with the location data related to each occupant. In the second layer, one DBN for each occupant is used to model the activities. In the third layer a BN is used to model the interaction between the occupants.

Kasteren et al.(Kasteren et al., 2011) investigate the use of Hierarchical Hidden Markov Model (HHMM) in a single occupancy setting showing higher accuracy compared to HMM and HSMM. Furthermore, Nguyen et al.(Nguyen et al., 2006) apply this model in a multi-occupant setting by constructing a separate HHMM for each occupant and reported a high accuracy of the model when tested on a video data.

To the best of our knowledge, hierarchical models have not yet been investigated for multi-occupant activity recognition in the context of pervasive sensing. It would be interesting to apply HHMMs for multi-occupancy to check their ability to infer high level behavior and to deal with parallel and cooperative activities.

II.5.4 Interaction Modeling

In contrast to the pervasive setting, much work on interaction modeling has been done in computer vision (McCowan et al., 2005) (Du et al., 2006) (Du et al., 2007) (Natarajan and Nevatia, 2007). In the following, we summarize the approaches discussed in the literature.

Recently, a number of studies on modeling occupant's interaction in pervasive environment have been conducted (Chiang et al., 2010) (Alemdar et al., 2013) (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b) (Lin and Fu, 2007) (Chen and Tong, 2014). Existing approaches include both supervised (Chiang et al., 2010) (Alemdar et al., 2013) (Wang et al., 2009) (Wang et al., 2011) (Lin and Fu, 2007) (Chen and Tong, 2014) and unsupervised approaches (Gu et al., 2009b). In supervised approaches, we can enumerate HMM (Alemdar et al., 2013) (Chen and Tong, 2014), CHMM (Chiang et al., 2010) (Wang et al., 2009) (Wang et al., 2011), PHMM (Chiang et al., 2010), BN (Lin and Fu, 2007) and FCRF (Wang et al., 2011).

Chiang et al. (Chiang et al., 2010) investigate close-proximity interaction using an *interaction feature*, that is, a binary feature which is set to 1 if the two occupants are in the same region of the environment and to 0 otherwise. The study shows that the presence of occupants in the same room does not imply that the occupants are involved simultaneously in cooperative activities. Although the contribution of this interaction feature is not significant, the model is more accurate than without it. Using the same dataset as (Chiang et al., 2010), Cook et al. (Cook et al., 2010) investigate the detection of close-proximity interaction using a Bayesian approach. It is found that the number of events generated during interaction is more important compared to the number of interactions detected. More interestingly, it is found that the physical proximity does not imply interaction.

In (Gu et al., 2009b), the authors apply EPs which describe important changes from once activity to the other. A confidence measure is proposed to determine if the occupants had interacted. EPs are mined for individual activities and for cooperative activities. However, EPs tend to recognize the activities as cooperative activities even when they are not.

To study the effect of interaction modeling on the efficiency of multi-occupant activity recognition, Wang et al. (Wang et al., 2009) used CHMM to recognize multi-occupant activities. The same authors applied later CHMM along with FCRF in (Wang et al., 2011). In particular, they proposed one HMM is constructed for each occupant to form CHMM. Hidden states in each HMM represent the set of activities performed by the occupant. Likewise, one CRF is constructed for each occupant in the FCRF. However, co-temporal dependencies between activities of occupants are represented differently in CHMM and FCRF. Considering, these co-temporal dependencies in CHMM, each activity in HMM does not depends on only the previous activity at time t-1 of the same HMM, but also on the previous activities at time t-1 from the other HMMs. In FCRF, activities of all CRFs corresponding to the occupants are joined at each time step and the same observation sequence is fed to all CRFs. In this study, it was found CHMM performs better than FCRF in the case of cooperative activities (e.g. the accuracy of CHMM on the cooperative activity “watching TV” is 100%, while that of FCRF is 70.5%). Considering the same context, Chiang et al. (Chiang et al., 2010) compare the performance of three models: PHMM, CHMM and CHMM extended with auxiliary nodes. The results show that the extended CHMM performs the best, while CHMM outperform PHMM. Finally, as the models in (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b) use the same dataset, comparing their recognition performance show that the EPs approach performs the best for activity recognition in terms of accuracy, scalability and robustness.

Two interesting studies lies between the use of wearable sensors (i.e. a Smartphone for each occupant) and the use of infrastructure sensors (i.e. motion sensors) to recognize parallel activities in a multi-occupant environment (Roy et al., 2013) (Roy et al., 2016). Authors in these studies applied a layered approach where each layer in the model received the results from the previous layer. They first extracted from each occupant's Smartphone accelerometer data the corresponding postures (e.g. walking, sitting..) over time. Then, an HMM is constructed for each occupant in the environment. In each occupant's HMM the observation at time represent both the posture of the occupant and motion sensor data generated at time in the environment while the hidden states are the corresponding locations (e.g. kitchen, bathroom..) of the occupant. Finally, they applied a CHMM to recognize complex activities of all occupants. The observation at time for an occupant's HMM in the CHMM represent both the posture (i.e. inferred in the first layer) and the location (i.e. inferred in the second layer) of that occupant while the hidden state represent the activity. Their experiments' results applied on the activity data of five occupants showed an accuracy of 70% in the inference of occupant's room level location which leads to an activity recognition accuracy improvement of 30% in comparison to CHMM based Smartphone solutions (i.e. observations in the CHMM represent only Smartphone data). Note although their model is able to recognize interaction between occupant (thanks to the use of CHMM in the third layer), their experiment data does not include cooperative activities and only complex parallel ADLs were recognized.

In (Natarajan and Nevatia, 2007) Coupled Hidden Semi-Markov Model (CHSMM) and CHMM are used to model multi-occupant activities using a dataset related to simultaneous hand gesture obtained by camera in the context of sign language. CHSMM outperforms CHMM by a difference of 20-30% accuracy rate. Considering multi-occupancy, transfer learning can be applied by substituting the two hands by two occupants and test the ability of the model to deal with parallel and cooperative activities in a pervasive setting.

II.5.5 Knowledge-driven vs Data-driven Approaches

We point out that all work presented above is data-driven using mainly probabilistic algorithms to build activity models. Knowledge-driven approaches, on the other hand, use ontology and symbolic representation (i.e. logic) to specify the semantic relations of activities as in the ontology *snapshot* approach used in (Riboni et al., 2011). Although, data-driven techniques exploit temporal information which is a very important aspect in activity recognition, the knowledge-driven techniques have been proven to be effective in single-occupant activity recognition (Riboni et al., 2011). When extended with simple forms of temporal reasoning, knowledge-based methods are comparable to the state-of-the-art techniques based on HMMs (Riboni et al., 2011).

Interestingly enough and to the best of our knowledge, ontology modeling has not yet been fully investigated in the context of multi-occupant activity recognition. An exception to this is the work described in (Lin and Fu, 2007) (already mentioned in Section II.5.3) where a combination of data-driven and knowledge-driven methods is proposed resulting in a layered approach. In the first layer, an ontology is used to interpret raw data from sensors by exploiting knowledge about the occupants and their relationships. In the second layer a DBN is applied to learn single-occupant preferences and in the third layer a BN is used to learn multi-occupant preferences.

This study however focuses on learning user preferences, not on recognizing the activities. Its merit lies in the fact that it provides a unified framework for recognizing both individual preferences and cooperative preferences. Considering multi-occupancy such a layered model can be applied by substituting the preferences by the activities and test the ability of the model to deal with parallel and cooperative activities in the pervasive setting.

Another interesting study was presented in (Alemdar et al., 2013) (Chen and Tong, 2014) (Tunca et al., 2014) where knowledge-driven and data-driven approaches for multi-occupant activity recognition are combined. The same approach was applied in the three studies but using different models and different datasets. In fact the authors of that studies exploited some simple knowledge of multi-occupant activities by defining "*combined labels*". Specifically, each observation in the dataset is represented by a label pair (activity label of Occupant 1, activity label of Occupant 2). The pair is then converted into a scalar to result in a combined label which represents the two activities of the two occupants. Using the dataset, all possible combinations are collected. After mapping each pair of multi-occupant activities labels in the training dataset to their combined label, an HMM is applied to construct the activity model in (Alemdar et al., 2013).

In (Chen and Tong, 2014), HMM and CRF were applied to construct the activity model. In both HMM and CRF models, hidden states represent the combined labels and the observations represent the sensor readings. The authors apply a two-stage method in the inference step. In the first stage of the method, CRF and HMM are applied to infer the combined label state. In the second stage, the combined label states are inversely mapped onto the corresponding occupants' activity labels. The results show that this approach increases the average accuracy by approximately 10% in comparison with the approaches described in (Hsu et al.,

2010) (Singla et al., 2010) using the same dataset. The results are slightly better for HMM in comparison with CRF.

Tunca et al. (Tunca et al., 2014), applied the same approach as in (Alemdar et al., 2013) (Chen and Tong, 2014). They compared five different classifiers namely, K-Nearest Neighbor (KNN), DT, HMM, MLP and TDNN. The number of activities for each occupant is 27 which would conduct to 27^2 possible combinations of occupants' activities (i.e. combined labels). In order to reduce the number of combined labels the authors grouped similar activities into more general activities (e.g. the activities Preparing a meal and washing dishes were grouped to the "KIT" activity and Having a meal or snack were grouped to "EAT" activity) which leads to having 7 activities instead of 27 activities for each occupant. Results suggest that TDNN and HMM performed slightly better than the other classifiers. To the best of our knowledge the approach used by the authors in (Alemdar et al., 2013) (Chen and Tong, 2014) (Tunca et al., 2014) is the only one which allowed to solve both data association and activity recognition at the same time.

Afrin Emi and Stankovic (Afrin Emi and Stankovic, 2015) present an activity recognition platform based active learning techniques and knowledge-driven about activities called SARRIMA. The latter extends AALO (Hoque and Stankovic, 2012) by considering the coexistence of multiple occupant at home and operates based on domain knowledge about activities (e.g. locations, objects involved, most likely time of the day) and specific assumptions on these ones (e.g. preparing dinner is usually performed in the kitchen). SARRIMA solves both the data association problem by the use of the person identification module and the activity recognition problem (i.e. parallel and sequential ones) by the use of the ADL recognition module. These modules can operate independently and separately of each other when only one of the latter problems is posed. They can also exchange information in order to identify a person and to check the choice of a recognized ADL. This approach detects about 97% of activity instances. Regarding the data association accuracy, the authors concluded that the latter is very dependent to the similarity of occupants' behavior and to the type of passive sensors installed.

II.5.6 Applicability, Adaptability and Scalability of Multi-Occupancy Models

The applicability and the adaptability of all models used in the context of multi-occupancy have not yet been investigated. Existing multi-occupant activity recognition systems are trained on private datasets (Wilson and Atkeson, 2005) (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b) (Lin and Fu, 2007) or on publicly available datasets (Prossegger and Bouchachia, 2014) (Crandall and Cook, 2008a) (Crandall and Cook, 2008b) (Hsu et al., 2010) (Chiang et al., 2010) (Cook et al., 2010) (Singla et al., 2010) (Alemdar et al., 2013) (Chen and Tong, 2014). Thus, the models are closely adjusted to the living space and the person, to the training data and to the types of activities monitored in the home. Then, such recognition models would only be applicable to that environment be it a single-occupant setting or multi-occupant setting.

To overcome the above limitations, Sarkar et al. (Sarkar et al., 2010) suggest the use of an alternative source of activity data that is a web data. Although their approach would work for almost any environment, the web data is clean and therefore cannot be used for real-world systems. The proposed activity model is developed for a single occupant setting and thus the authors only discuss its scalability in terms of adding new activities. However, dealing with scalability in multi-occupancy setting should not only consider new activities, but also new occupants. Clearly, the scalability of the models in terms of the number of occupants is the most important issue.

Referring to the scalability of data association algorithms, all studies in the literature have considered only a two-occupant situation (Crandall and Cook, 2008a) (Crandall and Cook, 2008b) (Crandall and Cook, 2010) (Hsu et al., 2010) (Cook et al., 2010), except (Wilson and Atkeson, 2005). In fact, the authors in (Wilson and Atkeson, 2005) study the impact of varying the number of occupants from 1 to 5 on simulated data and from 1 to 3 on real-world data using HMM. It is found that there is no difference in accuracy when varying the number of occupants on real-world data set in comparison with simulated data. This is a good sign that the model can be applied in real-world environments. However, the accuracy drops, because the complexity, that depends on the number of occupants, increases (Wilson and Atkeson, 2005). On simulated data, an accuracy of 100% is obtained for one occupant and only 67% for 4 occupants.

To the best of our knowledge, the scalability of activity recognition algorithms has never been considered. Nevertheless, as the authors in (Gu et al., 2009b) point out, EP-based models are scalable to additional occupants. Adding a new occupant would only imply mining the set of EPs for each activity monitored in the environment for this occupant. The scalability of the model presented in (Alemdar et al., 2013) (Chen and Tong, 2014) seems to be feasible too. In fact, the authors use a combined label to represent the two activities of two occupants at the same time. Adding another occupant implies combining three labels and hence, implies only increasing the number of label combinations. Chiang et al. (Chiang et al., 2010) noted that the scalability of the activity model would be more difficult to achieve and that training and inference will be computationally highly demanding, especially for CHMM and FCRF. In a nutshell the scalability problem is a challenging research avenue that is about the general issue of learning more generalized multi-occupant activity models.

II.5.7 Limitations of multi-occupant activity recognition systems

As we mentioned earlier, a number of studies on multi-occupant activity recognition have been carried out using the pervasive computing technology. Some of them investigate activity recognition ignoring data association. In this context, the studies in second approach of (Hsu et al., 2010) and second approach of (Singla et al., 2010) applied an individual model for each occupant (see Section II.5.3 for additional details). The advantage of this approach lies in the fact that it is easily scalable to new occupants. It only requires learning a new chain for the new occupant. Furthermore, this approach is suitable in case occupants follow their regular routines and do not much interact with each other. Thus, in case of more interactions taking place between occupants, this approach may not be suitable.

CHMM and FCRF used in (Chiang et al., 2010) (Wang et al., 2009) (Wang et al., 2011) and (Wang et al., 2011) respectively offer the advantage of modeling both parallel and cooperative activities. In contrast to CHMM, FCRF does not require using a data association variable in order to construct the activity model, since all occupant's activities at time t depend on all occupants' data. Another, major limitation of these models lies in their scalability to new occupants in the setting. In contrast, to the independent models for occupants, adding a new occupant in the environment imply to relearn the model again on all occupants' sensory data that is, both old occupant' sensory data and new occupant's sensory data.

Studies in the first approach of (Hsu et al., 2010) and (Cook et al., 2010) (Alemdar et al., 2013) (Chen and Tong, 2014) present the advantage of solving both data association and multi-occupant activity recognition. However, the methodology differs from (Hsu et al., 2010) (Cook et al., 2010) (see Section II.5.2) to (Alemdar et al., 2013)(Chen and Tong, 2014) (see Section II.5.5). The solution suggested in (Cook et al., 2010) requires solving the data association problem before activity recognition. The disadvantage of this approach is that the misclassification of the occupant by the data associator strongly impacts the recognition of the activity. Moreover, adding a new occupant in the environment implies retraining both the data association model (adding a hidden state representing the new occupant) and the activity recognition model (adding the hidden states corresponding to the activities of the new occupant).

Furthermore, the activity recognizer in the first approach of (Hsu et al., 2010) (Cook et al., 2010) and the first approach of (Singla et al., 2010) consists of a single-chain CRF and a single-chain HMM respectively. In the inference step, one activity label is inferred representing either the activity label of Occupant 1 or Occupant 2. Tracking the activity of each occupant requires recognizing the activities of all occupants at each time step. Using a single chain HMM or CRF in which the hidden states represent the activities of all occupants is not suitable for activity recognition.

A similar approach presented in (Alemdar et al., 2013) (Chen and Tong, 2014) (see Section II.5.5) also use a single chain HMM and a single chain CRF, but the hidden states refer to combinations of activity labels that are obtained by aggregating pairs of activities from the occupants. In the inference step, the combined label is converted back into the individual activity labels. By doing so, conventional graphical models, like HMM and CRF, can be applied to multi-occupancy. This approach, presents the advantage of solving both data association and recognizing both parallel and cooperative activities simultaneously. Moreover, although the concept is applicable regardless of the number of occupants, the process needs to be repeated again on all occupants' data if new occupants are added.

Comparing HMM and CRF, Kasteren et al. (Kasteren et al., 2008) reports that HMM is more appropriate than CRF for imbalanced activity data which contains dominant activities. In the "Ubicomp dataset" described in (Kasteren et al., 2008), we can encounter more events related to the activity "going to bed" than those related to the activity "toileting".

II.5.8 Evaluation Issues

To evaluate the performance of computational models for both data association and activity recognition models, the evaluation method should describe how the data is to be used, how training is to be carried out, and how validation and testing are to be conducted. The performance metrics, used for evaluating the model, are very important in the validation of any model. Selecting the adequate metrics strongly depends on the specific problem (e.g. classification, regression) at hand. In the following, we will display the different criteria used in the literature references mentioned in this chapter to assess the performance of multi-occupant activity recognizers.

$$Accuracy = \frac{1}{E}\sum_{i=1}^{E}[inferred(i) = true(i)] \tag{II.12}$$

$$False\ Positive\ Rate = \sum_{i=1}^{l} fp_i \tag{II.13}$$

$$Error\ Rate = \sum_{i=1}^{l} \frac{fn_i + fp_i}{tp_i + fn_i + fp_i + tn_i} \tag{II.14}$$

$$Time\ Slice\ Accuracy = \frac{1}{N}\sum_{i=1}^{N}[inferred(i) = true(i)] \tag{II.15}$$

$$Accuracy_i = \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i} \quad i = 1, \dots l \tag{II.16}$$

$$Average\ Accuracy = \frac{1}{l}\sum_{i=1}^{l} Accuracy_i \tag{II.17}$$

$$Average\ Precision = \frac{1}{l}\sum_{i=1}^{l}\frac{tp_i}{tp_i + fp_i} \quad \text{(II.18)}$$

$$Average\ Recall = \frac{1}{l}\sum_{i=1}^{l}\frac{tp_i}{tp_i + fn_i} \quad \text{(II.19)}$$

$$Average\ F - score = \frac{Precision \times Recall}{Precision + Recall} \quad \text{(II.20)}$$

$$Average\ Error\ Rate = \frac{1}{l}\sum_{i=1}^{l}\frac{fn_i + fp_i}{tp_i + fn_i + fp_i + tn_i} \quad \text{(II.21)}$$

where l is the total number of classes/activities and N is the total number of time slices when sensory data is discretized using a constant length. E is the total number of sensor events. $[a = b]$ is a binary indicator having the value 1 if true and 0 otherwise. tp_i is the number of true positives (instances from the i^{th} class that are correctly recognized being from the i^{th} class), tn_i is the number of instances recognized as not part of class i and indeed they are not (true negatives), fp_i *is* the number of instances that are incorrectly recognized as part of the i^{th} class (false positives) and fn_i is the number of instances recognized as part of the i^{th} class, while they are not (false negatives). The quantities tp_i, tn_i, fp_i and fn_i for each class are computed from the confusion matrix.

In the context of data association, classes are the occupants, whereas in activity recognition classes represent the activities. Authors in (Crandall and Cook, 2008a) (Crandall and Cook, 2008b) study the impact of adding new features (hour of day, day of week, part of day, part of week) to the data and used NB and HMM in order to deal with data association. Two measures are used: the *accuracy rate* (equation II.12) and the *false positive rate* (equation II.13). They compute each of the two metrics for each feature type to select the feature which reports the best results and to evaluate the effect of features on the efficiency of occupant identification. For instance, in a two-occupant home, a person would spend much more time at home than the other one and hence, the probability that an event would be generated by a person will be attributed to the person that caused most of the events resulting a high false positive rate. A good occupant classification would result in high accuracy and low false positive rate.

When comparing NB and HMM for occupant identification, the authors in (Crandall and Cook, 2010) use the *average lag* to assess the performance of their HMM model. The average lag is defined as the average number of events after a transition in sensor data before HMM correctly classifies the occupant. An average lag of 1 indicates that HMM improperly classified one event in each transition from one occupant to the other one in activity data before correctly recognizing the occupant causing the events. The authors also use the *error rate* (equation II.14) which represents the ratio of errors made when classifying a number of instances. Wilson and Atkeson (Wilson and Atkeson, 2005) use *time-slice accuracy* (equation II.15) to evaluate the effectiveness of their HMM based occupant identification problem. Time slice accuracy is the ratio of correctly classified time slices when data is discretized using a time length. On the other hand for evaluating activity recognition, the authors in (Chiang et al., 2010) compute the accuracy for each occupant separately. They also compute the joint accuracy which is counted when the activity recognized for both Occupant 1 and Occupant 2 are correct.

Existing datasets in activity recognition such as ARAS (Alemdar et al., 2013) and CASAS" Multiresident ADLs" (Singla et al., 2010) are imbalanced which means that some classes have more instances in the dataset than do other classes. Hence, because of the class imbalance the correct classification of each class is equally important for activity recognition; many studies in the field tend to apply the *average accuracy* measure (equation II.17). As a result, the authors in (Hsu et al., 2010) (Cook et al., 2010) (Singla et al., 2010) (Alemdar et al., 2013) (Chen and Tong, 2014) (Afrin Emi and Stankovic, 2015) use the *average accuracy* to assess the performance of their activity models. In (Cook et al., 2010) (Chen and Tong, 2014) (Tunca et al., 2014) many measures are applied: the average accuracy, the *average precision* (equation II.18), the *average recall* (equation II.19), and the *average f-score* (equation II.20). In addition to these metrics the *average error rate* (equation II.21) was used in (Chen and Tong, 2014). Although Tunca et al. (Tunca et al., 2014) computed many measures: , the *average precision* (equation II.18), the *average recall* (equation II.19), the *average f-score* (equation II.20) and the accuracy rate they only reported the two latter metrics' results for each occupant and for each house of the ARAS dataset.

While many metrics are usually used to assess multi-occupant activity recognition models some studies only applied the accuracy rate as shown in (Lin and Fu, 2007) (Prossegger and Bouchachia, 2014) (Roy et al., 2013) and (Roy et al., 2016).

II.6 International Research Groups

Several research groups have equipped experimental living spaces with pervasive sensors for HAR research like GeorgiaTech Aware Home Research Initiative (AHRI)[14], Intel research laboratory at Seattle[15], Domotics and Mobile computing Research (DOMUS)[16]at Sherbrooke University (Canada) and the Place Lab at Massachusetts Institute of Technology (MIT)[17].

However, only few research groups have been working on multi-occupant activity recognition. The NTU Wisdom Family (Attentive Home)[18] targets the family environment as shown in recent work (Hsu et al., 2010) (Chiang et al., 2010); whereas others (Lin and Fu, 2007) looked at the problem of multi-user preference modeling. Members of the Institute of Computer Software (ICS) at Nanjing University worked on multi-occupant activity recognition from wearable sensors (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b).

The ARAS group project mainly focuses on the physical layer and communication layer of an SH system to collect multi-occupant activity data for HAR. They published the details of a multi-occupant dataset (see Section II.3.2), which includes a variety of pervasive sensors as well as a variety of activities (Alemdar et al., 2013). Later, the ARAS group project, presented in (Tunca et al., 2014) the details of the field study they conducted in (Alemdar et al., 2013) with the intention of contributing on the acceptance and commercialization of wireless sensor network based ambient assisted living systems compatible for homes with multiple occupants. They focus on the details of the system architecture, including the challenges of sensor selection, deployment (i.e. given the variety of sensor types and activities and considering the main issues which are privacy, unobtrusiveness and robustness, how to select the most suitable set of sensors in the deployment), networking and data collection. They also summarized their

[14]http://www.awarehome.gatech.edu/drupal/
[15]http://www.intel.com/research/network/seattle_human_activity_recognition.htm
[16]http://www.domus.usherbrooke.ca/
[17]http://web.mit.edu/cron/group/house_n/placelab.html
[18]http://www.attentivehome.org/index.html

experiences while meeting these challenges to guide future studies in the deployment of similar systems.

The CASAS group treats environments as intelligent agents, where the status of the occupants and their physical surroundings are perceived using sensors and the environment is acted upon using controllers in a way that improves the comfort, safety, and/or productivity of the occupants. This research group seems to be a driving force in the area of HAR, be it for a single-occupant (Cook et al., 2013) or a multi-occupant setting (Cook et al., 2010) (Singla et al., 2010). Their main areas of study are: activity learning, discovery, recognition and prediction, gerontology, multiple occupant profiling, automated clinical diagnoses, energy-efficient home automation, activity-aware intervention, identification of trends and anomalies in SH sensor data. One of their major contributions in this area is making around 24 datasets for use in their own research and also made publicly available, creating a collaborate approach and improving technology evolution.

II.7 Conclusion

So far research related to multi-occupant SHs has devoted significant attention to the application of graphical probabilistic algorithms to model and recognize activities. This chapter emphasizes the importance of the various technology aspects to fully realize the multi-occupancy paradigm. While there has been much effort invested on the single occupancy paradigm, multi-occupancy has started recently to be the central focus of many studies. Clearly, there were and are still many outstanding scientific questions related to single occupancy to be dealt with before dealing with those specific to multi-occupancy.

In this chapter, we pointed out the major issues pertaining to activity recognition in the context of multi-occupancy taking data association and interaction into account. We discussed in detail the state-of-art computational models used for modeling collaborative activities, the existing benchmark datasets and the evaluation metrics.

CHAPTER III

Proposed Approaches for Multi-Occupant Activity Recognition

III. Proposed approaches for Multi-occupant Activity Recognition

III.1 Introduction

This chapter investigates the problem of HAR in a multi-occupant setting. In this context parallel activities and cooperative activities are considered. The goal is to accurately recognize both types of activities from non-intrusive sensors. We do not focus on only one of the two types of activities as done in the literature related to multi-occupant activity recognition, but on both types.

To deal with multi-occupant activities we investigate different approaches based on HMMs. Specifically, we propose an HMM-based method, called CL-HMM, where we combine occupants' activities labels as well as occupants' observation labels at time to generate the corresponding sequence of activities as well as the corresponding sequence of observations on which a conventional HMM is applied. We also propose a LHMM in which activities of all occupants are linked at each time step. We compare these two models to baseline models which are CHMM and PHMM.

This chapter is organized as follows. Because our proposed approaches are based on HMM, we first give details the latter model in Section III.2. Section III.3, describes the proposed graphical models CL-HMM and LHMM. Section III.4, discusses the experiments conducted while evaluating the proposed models. Details about the experimental dataset (i.e. "Multiresident ADLs" of CASAS) as well as the pre-processing applied on are given in Section III.4.1. Description of baseline models which are PHMM and CHMM is given in Section III.4.2. Results of experiment 1 and experiment 2 are discussed in Section III.4.3 and Section III.4.4 respectively. Section III.5 compares our models against existing studies which relied on the same dataset. Section III.6, concludes the chapter.

III.2 Markov Models

A Markov Model (MM) is a simplification of a DBN that models the temporal aspect of processes as shown in figure 14.

Figure 14: Representation of a Markov Model

The first order Markov assumption was proposed to simplify the dependence relationship between consecutive states. It stipulates that the present state at time t depends only on the previous one, that is:

$$P(y_t|y_1, y_2, y_3, \dots, y_{t-1}) = P(y_t|y_{t-1}) \tag{III.1}$$

Thus, the future state depends only on the current state, not on past states (Sutton and McCallum, 2006); that is, y_t depends only on y_{t-1}.

A specific case of MMs is called the Markov Chain (MC) and corresponds to the case where the states are all observable. A MC is a sequence of random variables $X_1, X_2, X_3, \dots, X_t$ with the Markov property. Formally:

$$P(X_{t+1} = x|X_1 = x_1, X_2 = x_2, \dots, X_t = x_t) = P(X_{t+1} = x|X_t = x_t) \tag{III.2}$$

The possible values of X_i form a countable set S called the state space of the chain.

MCs are not popular in human activity modeling since we cannot always directly recognize the activities from sensory data. In general, only simple activities can be modeled using MCs (Kim et al., 2010). Interestingly, MCs were applied in (Crandall and Cook, 2008b) to model the data association problem in order to identify the occupants.

III.2.1 Hidden Markov Model (HMM)

The most popular generative temporal probabilistic model is the HMM. In contrast to MC, HMM consists of hidden and observable states. The data (x_1, x_2,..., x_T), is therefore assumed to be generated by a temporal process whose states are hidden, (y_1, y_2,..., y_T) as shown in figure 15.

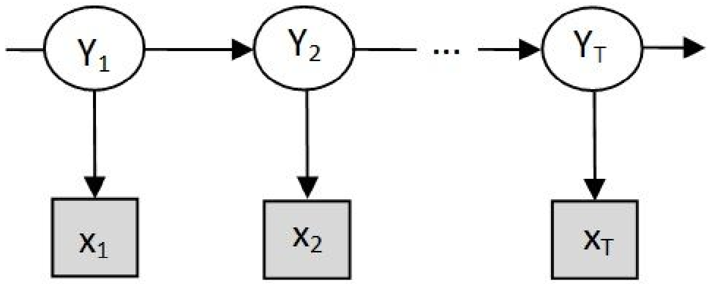

Figure 15: Representation of HMM

HMM relies on two assumptions: which are the 1st order Markov assumption in relation to the independence of hidden states and the conditional independence of observation parameters stipulating that:

$$P(x_t|y_t, x_1, x_2, \dots, x_{t-1};\ y_1, y_2, \dots, y_{t-1}) = P(x_t|y_t) \tag{III.3}$$

The observable state at time t, x_t depends only on the current hidden state y_t. That is, the probability of observing x_t while being at y_t is independent of all other observable and hidden variables (Sutton and McCallum, 2006).

The joint probability $P(y)$ of the observations and hidden states can be factorized as follows:

$$P(x, y) = \prod_{t=1}^{T} P(y_t|\, y_{t-1})\, P(x_t|y_t) \tag{III.4}$$

where $P(y_t|y_{t-1})$ and $P(x_t|y_t)$ indicate the probability of transition between the two consecutive hidden states y_{t-1} and y_t and the probability of observing x_t at state y_t respectively (Sutton and McCallum, 2006). Given the sequence of observables, the maximum of the joint probability corresponds to the highly probable sequence of hidden states.

In the context of activity recognition, similar modeling is adopted like with the previous computational models. That is, the hidden states are the activities and observations are the sensed data as shown in figure 16 (i.e. representation of an HMM in the context of *adlnormal* activity data). Both activities (i.e. hidden states) and sensors' readings (i.e. observable states) are represented using rectangles as they are discrete. The links between the hidden states are labeled with the transition probabilities and those between the hidden states and the observables are labeled with the emission probabilities.

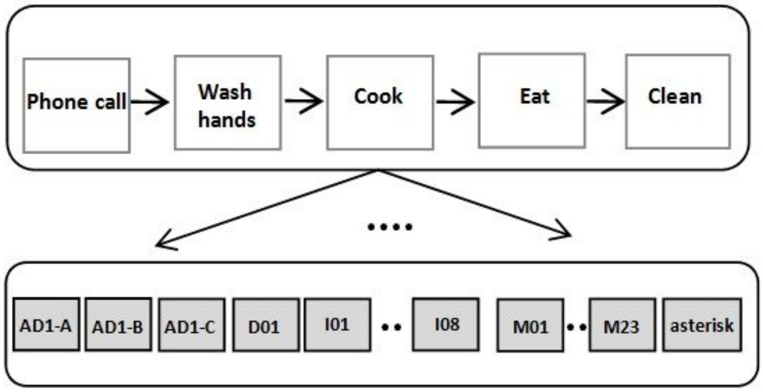

Figure 16: Representation of HMM in HAR

Usually when activities are sequential, it is possible to separate activity data and then create one HMM for each activity. However, if the activities are interleaved, this way of modeling is not suitable because of the interlacement of the activities will be disregarded. In addition, finding the optimum number of hidden states for each HMM corresponding to an activity is another issue. Creating an HMM for each activity would lead to have the same sensor model for each activity, but the number of hidden states for each activity is unknown. In fact, the authors in (Khan et al., 2012) used the accuracy to find the optimum number of hidden states and suggested to use techniques applied for Hierarchical Dirichlet Process-Hidden Markov Model (HDP-HMM) (Hu et al., 2009) and Infinite Hidden Markov Models (IHMMs) (Pruteanu-Malinici and Carin, 2008).

In some cases, even when the complex activity is decomposed, its sub-activities also form complex activities which are not directly observable (hidden). For instance, the activity "Prepare a dinner" can be decomposed into the activity "prepare a drink" and the activity "cook" and each of them also includes sub-activities. Then individually trained HMMs on the activities can be combined to build a global HMM. Thus, hierarchical graphical models (e.g. HHMM or Abstract Hidden Markov Model (AHMM)) look more suitable in this case.

III.3 Proposed Models for Multi-Occupant Activity Recognition

In the following both CL-HMM and LHMM are described. Without loss of generality, we assume that we have two occupants living in the same home. Let N and M be the number of activities performed by Occupant 1 and Occupant 2 respectively and let Q be the number of sensors present in the living space and trigged by the two occupants. The description presented below can be easily generalized to any number of occupants and is based on the following assumptions:

- The data association variables are given (i.e. we know who triggered which sensor).
- In each occupant's HMM of the LHMM, the hidden states represents the activities performed by the occupant while observable states correspond to the sensor events generated by the occupant.
- We are dealing with discrete data for both the activities and the sensory data.
- We have the ground truth activity corresponding to each occupant's observation in the dataset, we therefore apply a supervised learning which means that parameter estimation is achieved by frequency counting of occurrences of states, transitions and observations (Rabiner, 1989).
- The Viterbi algorithm is applied only in the test step to infer the sequence of hidden states (i.e. the activities) that best explains a new sequence of observations (i.e. sensor events).

We describe in the following parameter estimation as well as the Viterbi algorithm for each of the two models.

III.3.1 HMM-based Combined Label (CL-HMM)

a. Definition

Each occupant has own sequence of sensor events that the activities. Thus, each occupant has an observation at each time step as part of an activity. In an environment including k occupants, this would result into a vector of length k for the observations and a vector of length k for activity labels at each time step. The CL-HMM primarily consists of converting and combining the activities (L_{1j}, L_{2j}, L_{3j},…,L_{kj}) and the corresponding data (o_{1j}, o_{2j}, ..o_{kj}) of individual occupants into a single observation (o) and a single combined activity label (L). Once this pre-processing step is completed, the learning process can start.

b. Setup

Instead of considering that each occupant's activities A^1 and A^2 have their corresponding observations O^1 and O^2, in a two-occupant environment, we suppose that each pair of activities at each time step generates a pair of observations. That is, the pair of activities (a_t^1, a_t^2) generates the pair of observations (o_t^1, o_t^2) at time t. The CL-HMM approach can be divided into five main steps:

(1) Define the sets C*LA* and C*LO* of combined label of activity pairs and combined label of observation pairs respectively. If we consider all activities of the two occupants that can appear in parallel, the number of resulting combined labels is N×M. However, some activities cannot occur in parallel in real world situations. For instance, Occupant 1 cannot take a shower if the bathroom is occupied by the Occupant 2; thus the number of combinations can be reduced. Using the experimental data, we extract non redundant pairs of activity labels and attribute to each pair a scalar resulting then into the set of combined labels for activities *CLA*. For instance, *CLA* consists of = {1, 2, 3, 4} as the set of combined labels for activities pairs (1,2), (2,1), (1,3), (3,4).The same process is applied to extract non redundant pairs of observation labels from the data. This will result in Q×Q possible values for the *CLO* set. However, due to the fact that some sensors do not occur in parallel in real-world situations (e.g. the

phone sensor cannot be trigged by both occupants at the same time), the cardinality of *CLO* will be smaller than Q×Q.

(2) Using the CLA set, we then convert each pair of multi-occupant activities labels $(a^1, a^2)_{(1:T)}$ (i.e. $\{(a^1{}_{(1)}, a^2{}_{(1)}), (a^1{}_{(2)}, a^2{}_{(2)}),..., (a^1{}_{(T)}, a^2{}_{(T)})\}$) into combined activities $a_{(1:T)}{}^C$ as shown in figure 17 (blank squares indicate the hidden states, the shaded squares indicate the observed states). Likewise, using the CLO set we convert each pair of multi-occupant observations labels $(o^1, o^2)_{(1:T)}$ into combined observations $o_{(1:T)}{}^C$ respectively.

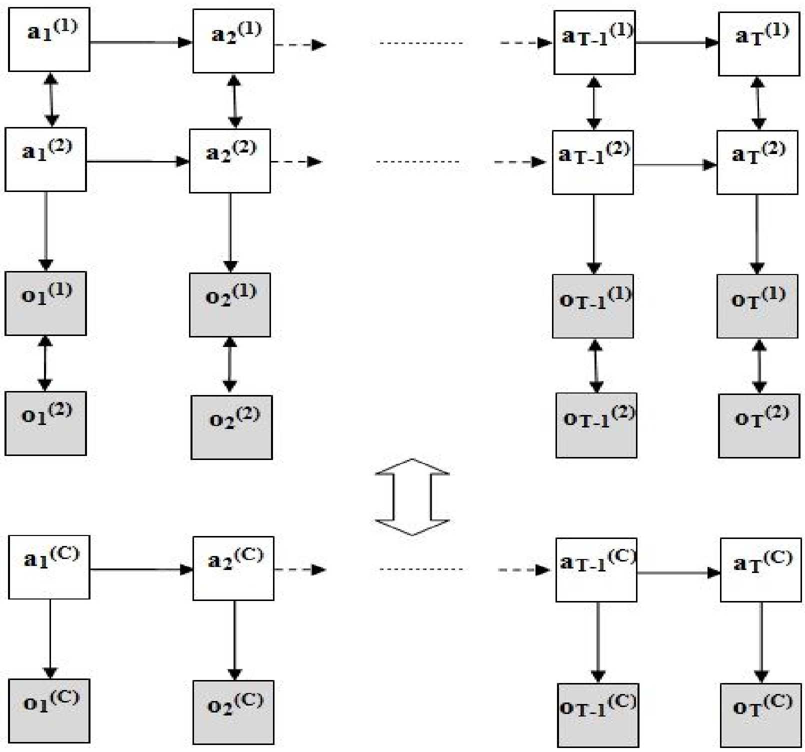

Figure 17: Topology of CL-HMM

(3) Estimate the HMM parameters - An HMM is then applied on the data resulting from the previous step to build an activity model. In this HMM, hidden states represent the combined activity label $a_{(1:T)}{}^C$, while the observed states represent the combined observations of both occupants

$o_{(1:T)}^{C}$. The HMM consists of the initial state vector, the transition matrix and the observation matrix as follow:

$$\pi^{C}(i) = P(a_1^C = i) \tag{III.5}$$

$$Q_{ij}^{C} = P(a_t^C = j | a_{t-1}^C = i) \tag{III.6}$$

$$B_t^C(i) = P(o_t^C | a_t^C = i) \tag{III.7}$$

where i=1... N×M, and $B^{C}{}_{t}(i)$ is an (N×M) by (QxQ) matrix
As mentioned earlier, parameter estimation is simply achieved by frequency counting of occurrences of initial states, transitions and observations (Rabiner, 1989). For example, $\pi^{C}(1)$ represents the number of training sequences in which the combined label state (1) appears at the beginning divided by the total number of training sequences. Likewise, for the transitions, $Q_{31}{}^{C}$ represents the number of transitions from the combined activity label 3 to the combined activity label 1 divided by the number of outgoing transitions from the combined activity label 3. Similar frequency count for the observations can be done for example $B^{C}(2)$ is a vector indicating the observation probability of all the combined observations (1...Q^2) from the combined activity label 2. To count the probability of a specific observation from the combined activity label 2 that is $P(o_t{}^{C}=1|a_t{}^{C}=2)$, we compute the number of occurrence of combined activity labels 2 in which the combined observation label 1 appears divided by the occurrence frequency of the combined label 2.

(4) Inference for the HMM - Given an observation sequence O^C we need to find a state sequence A^C which maximizes $P(A^C|O^C)$. The Viterbi algorithm for HMM (Rabiner, 1989) (see appendix A) outputs the best state sequence A^C which represents the best state sequence of the combined activities. This results in a computational complexity of $O(T^2N^2M^2)$ where T is the total number of events of the dataset. Considering, R occupants each having a number of corresponding activities N_i for i=1...R, the computational complexity would be $O(T^R\prod_{i=1}^{R} N_i^2)$.

(5) Extract each occupant activity sequence A^1 and A^2 - The obtained combined activity label, A^C, from the previous step is then converted back into the original individual activities of the two occupants A^1 and A^2.

III.3.2 Linked HMM (LHMM)

a. Definition

LHMM was introduced for the first time in (Brand, 1997). The latter represents a combination of multiple HMMs, where each HMM consists of a set of hidden states and a set of observed states. It is called LHMM because there are direct edges from hidden states of an HMM to the hidden states of the other HMMs.

When a LHMM is applied to HAR, an HMM is constructed for each occupant in the environment. For instance, in a two-occupant setting, $\{A^1, O^1\}$ and $\{A^2, O^2\}$ represent the sequence of activities and sensor events from Occupant 1 and Occupant 2 respectively. In each HMM chain, hidden states represent the activities of the corresponding occupant, whereas the observations represent sensor events. That is, $A^i=\{a_{(1:T)}{}^i\}$ (i=1,2) are the hidden states and $O^i=\{o_{(1:T)}{}^i\}$ are the corresponding observations for Occupant i. The pair of HMMs is combined to obtain LHMM (see figure 18, blank squares indicate the hidden states, the shaded squares indicate the observed states). When two occupants in a SH perform cooperative activities, the activity of one occupant at time t can affect not only the activity at time t+1 in the same model, $a_t{}^n$ to $a_{t+1}{}^n$, but also the activity of the other occupant, $a_t{}^n$ to $a_t{}^m$.

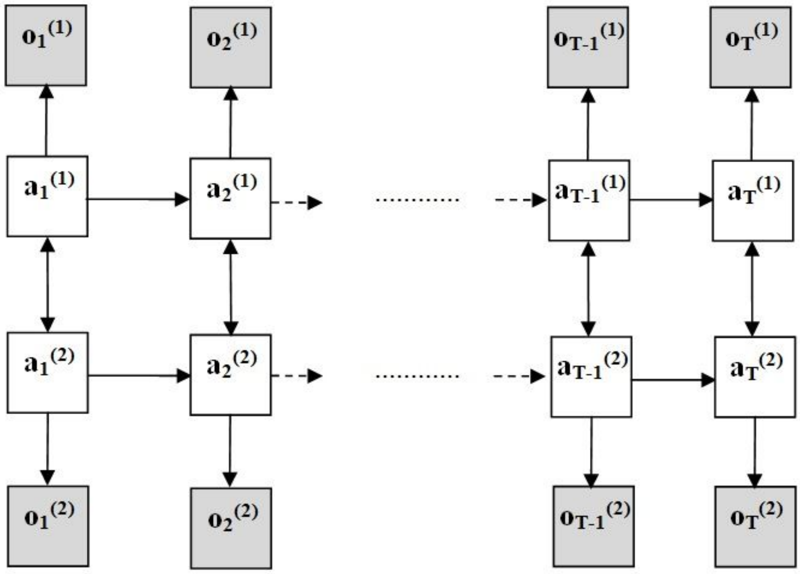

Figure 18: Topology of LHMM

Since both HMMs are not independent, the posterior of the activity sequences given all the observations can be expressed as:

$$P\left(A^{(1)}, A^{(2)} \middle| O^{(1)}, O^{(2)}\right) = \frac{P\left(O^{(1)}, O^{(2)} \middle| A^{(1)}, A^{(2)}\right) P\left(A^{(1)}, A^{(2)}\right)}{P(O^{(1)}, O^{(2)})}$$

$$\propto P\left(O^{(1)}, O^{(2)} \middle| A^{(1)}, A^{(2)}\right) P\left(A^{(1)}, A^{(2)}\right)$$

(III.8)

According to the condition of independence given in the structure (figure 18), we can factorize $P(O^1, O^2 | A^1, A^2)$ and $P(A^1, A^2)$ as follow:

$$P\left(O^{(1)}, O^{(2)} \middle| A^{(1)}, A^{(2)}\right) = P\left(o_1^{(1)}, o_1^{(2)} \middle| a_1^{(1)}, a_1^{(2)}\right) \dots P\left(o_T^{(1)}, o_T^{(2)} \middle| a_T^{(1)}, a_T^{(2)}\right)$$

(III.9)

where:

$$P\left(o_1^{(1)}, o_1^{(2)} \middle| a_1^{(1)}, a_1^{(2)}\right) \dots P\left(o_T^{(1)}, o_T^{(2)} \middle| a_T^{(1)}, a_T^{(2)}\right)$$
$$= \prod_{t=1}^{T} P(o_t^{(1)} \left| a_t^{(1)}, a_t^{(2)}) P(o_t^{(2)} \right| a_t^{(1)}, a_t^{(2)})$$

(III.10)

and

$$P\left(A^{(1)}, A^{(2)}\right) = P(a_1^{(1)}, a_1^{(2)}) \prod_{t=2}^{T} P(a_t^{(1)}, a_t^{(2)} \left| a_{t-1}^{(1)}, a_{t-1}^{(2)}\right.)$$

(III.11)

$P(a_t^{(1)}, a_t^{(2)})$ is a hidden state pair at time t as a combination of the hidden states of the chains A^1 and A^2 . $P(a_t^{(1)}, a_t^{(2)}| a_{t-1}^{(1)}, a_{t-1}^{(2)})$ is the transition probability of the hidden state pairs, $P(o_t^{(1)}|a_t^{(1)}, a_t(2))$ is the observation probability of $o_{(1:T)}^1$ from all possible hidden state pairs and $P(o_t^{(2)|}a_t^{(1)}, a_t^{(2)})$ is the observation probability of $o_{(1:T)}^2$ from all possible hidden state pairs.

We proceed like in the case of CL-HMM, we combine and convert the pair of activities corresponding to Occupant1 and Occupant 2 into a single combined activity label (L) to obtain the sequence A^C. To do that, we define the CLA set (see step one in section III.3.1 (setup)) and use it to convert each pair of multi-occupant activities labels $(a^1, a^2)_{(1:T)}$ into combined activities $a_{(1:T)}{}^{C}$.

b. Parameter estimation

Considering the pair of activities $(a_t^{(1)}, a_t^{(2)})$ as one activity $a_t^{(C)}$ (see equations (III.8)-(III.11)), the parameters of our LHMM are:

$$\pi^C(i) = P(a_1^C = i)$$

(III.12)

$$Q_{ij}^C = P\left(a_t^C = j \middle| a_{t-1}^C = i\right)$$

(III.13)

$$B_t^1(i) = P(o_t^1 \middle| a_t^C = i) \quad \text{(III.14)}$$

$$B_t^2(i) = P(o_t^2 | a_t^C = i)$$
(III.15)

where i=1...N×M and both $B_t^1(i)$ and $B_t^2(i)$ are N×M by Q matrix. $\pi^C(i)$ is the initial state vector of the combined activity label sequence A^C, Q_{ij}^C is the transition matrix of A^C. $B_t^1(i)$ and $B_t^2(i)$ are the probability of observing $O^1 = o_{(1:T)}^1$, $O^2 = o_{(1:T)}^2$ from A^C respectively. Like with CL-HMM, all parameters of LHMM are easily computed using frequency counting of occurrences of initial state vector, transitions and observations (Rabiner, 1989). It does not differ from CL-HMM to LHMM when computing the initial state vector and the transition matrix of the single combined activity label sequence A^C.

The main difference between our LHMM and our CL-HMM lies in the conditional independencies of the observations O^1 and O^2 over the single combined activity label A^C. Because, in CL-HMM, O^1 and O^2 are converted too into a single combined observation label O^C, they are considered to be relatively dependent over the single combined activity label sequence A^C.O^1, O^2 in LHMM are conditionally independent over A^C.

c. Inference

LHMMs inference is formulated as follows. Given an observation sequence O={O^1, O^2}, we need to find a state sequence A^C which maximizes $p(A^C| O^1, O^2)$.We apply the Viterbi algorithm for HMM (Rabiner, 1989). The latter outputs the best state sequence A^C which represents the best state sequence of the combined activities. According to equation III.10, the sequences of observations O^1, O^2 are conditionally independent over the single combined activity label A^C, that is, $P(O^1, O^2|A^C)= P(O^1|A^C)\times P(O^2|A^C)$(i.e. $B_t^C(i)= B_t^1(i)\times B_t^2(i)$). The description of the Viterbi is shown in the following:

Let $\delta_t(i)$ be the maximal probability of state sequences of the length t that end in state i and produce the t first observations for the given model.

$$\delta_t(i) = \max\left\{P((a_1^C)\ldots,(a_{t-1}^C);\ \left(o_1^{(1)},o_1^{(2)}\right),\ldots,\left(o_t^{(1)},o_t^{(2)}\right)\middle|(a_t^C)=i)\right\} \tag{III.16}$$

The matrix ψ is used to retrieve the optimal hidden states at the backtracking step.

- Initialization

$$\delta_1(i) = \pi^C(i)B_1^1(i)B_1^2(i) \tag{III.17}$$

$$\psi_1(i)=0 \tag{III.18}$$

- Recursion

$$\delta_t(j) = \max_i\{\delta_{t-1}(i)Q_{ij}^C\}B_t^1(j)B_t^2(j) \tag{III.19}$$

$$\psi_t(j) = \operatorname{argmax}_i\{\delta_{t-1}(i)Q_{ij}^C\} \tag{III.20}$$

- Termination

$$P^* = \max_i\{\delta_T(i)\} \tag{III.21}$$

$$\{a_T^C\} = \operatorname{argmax}_i\{\delta_T(i)\} \tag{III.22}$$

- Path backtracking

$$\{a_t^C\} = \psi_{t+1}(a_{t+1}^C), t = T-1, T-2, \ldots, 1 \tag{III.23}$$

where P^* is the maximum likelihood of $\delta_T(i)$ at time T and a_T^C is the most probable combined label for the activities at time T. The obtained combined activity label, A^C, from the previous step is then converted back into the original individual activities of the two occupants A^1 and A^2.

This Viterbi algorithm results in a computational complexity of $O(TN^2M^2)$ where T is the total number of events in the dataset. Considering, R occupants each having a number of corresponding activities N_i for i=1...R, the computational complexity would be $O(T\prod_{i=1}^{R} N_i^2)$.

III.4 Experiments

In the following we will describe the dataset as well as the preprocessing associated with before introducing two models PHMM and CHMM against which our models CL-HMM and LHMM are compared. Two main experiments are studied. In the first we present the results of the individual occupants using all models; while in the second experiment joint results after preprocessing are discussed. Special attention is given to the performance of the models on cooperative and parallel activities.

III.4.1 Experimental dataset description

To evaluate the models proposed in this study we use a publically available multi-occupant dataset which is the CASAS "Multi-occupant ADLs" dataset[19] (Singla et al., 2010). This dataset was collected through 26 volunteer pairs performing 15 scripted activities defined as follow:

(1) **Fill medication dispenser** in the kitchen using items obtained from the cabinet. Return items to the cabinet when done. (Occupant 1)
(2) **Hang up clothes in the hallway closet.** The clothes are laid out on the couch in the living room. (Occupant 2)
(3) **Move the couch and coffee table to the other side of the living room.** (Occupant 2) Request help from Occupant 1 (Occupant 1 will stop dispenser activity to help and finish dispenser activity when done with this activity)
(4) **Sit on the couch and read a magazine.** (Occupant 2)
(5) **Water plants located around the apartment.** Use the watering can located in the hallway closet. Return the watering can to the closet when finished. (Occupant 1)
(6) **Sweep the kitchen floor** using the broom and dust pan located in the kitchen closet. Return the tools to the closet when finished. (Occupant 2)
(7) **Play a game of checkers** for a maximum of five minutes. (Occupant 1 and Occupant 2)
(8) **Set out ingredients for dinner in the kitchen.** (Occupant 1)
(9) **Set dining room table for dinner.** (Occupant 2)
(10) **Read a magazine on the living room couch.** (Occupant 1)

[19] http://ailab.wsu.edu/casas/datasetdlmr.zip

(11) **Simulate paying an electric bill.** Retrieve a check, a pen, and an envelope from the cupboard underneath the television in the living room. Use the telephone book in the dining room to look up a number for a utility company to confirm the amount on the bill. Occupant 2 Request help from Occupant 1 to find number for utility company (Occupant 1 will stop current activity to help and finish activity when done helping).

(12) **Gather food for a picnic from the kitchen cupboard and pack them in a picnic basket.** (Occupant 1)

(13) **Retrieve dishes from a kitchen cabinet. (**Occupant 2) Request help from Occupant 1 to identify cabinet in which the dishes are located. (Occupant 1will stop current activity to help and finish activity when done helping)

(14) **Pack supplies in the picnic basket.** (Occupant 2)

(15) **Pack food in the picnic basket and bring the basket to the front door of the apartment.** (Occupant 1)

Note that the activities 3, 7, 11 and 13 are cooperative, while the 11 remaining ones are parallel individual ones.

In all, the dataset contains 17 232 events described by (Date, Time, SensorID, Value, OccupantID, ActivityID). If an event is trigged by the two occupants, it is then represented by (Date, Time, SensorID, Value, OccupantID, ActivityID, OccupantID, ActivityID).

Activities were manually annotated by recording their start and end time. Although, the SH (i.e. Kyoto testbed) includes a variety of sensor types (i.e. motion sensors, item sensors, burner sensor, water sensors, light controllers, phone sensor), only 27 motion sensors (i.e. from M01 ...M26 and M51), 2 item sensors (i.e. I04 and I06) and 8 door sensors (i.e. D07 and from D09 ..D15) were used to collect this dataset as shown by red rectangles in figure 19 (remaining sensors were not involved in the registration of this dataset). All these sensors produce binary values: ON/OFF for motion sensors, PRESENT/ABSENT for item sensors, OPEN/CLOSE for the cabinet sensor. The motion sensors provide the real-time location of the occupants. In their absence, the location of the person corresponds to the location of the last object used.

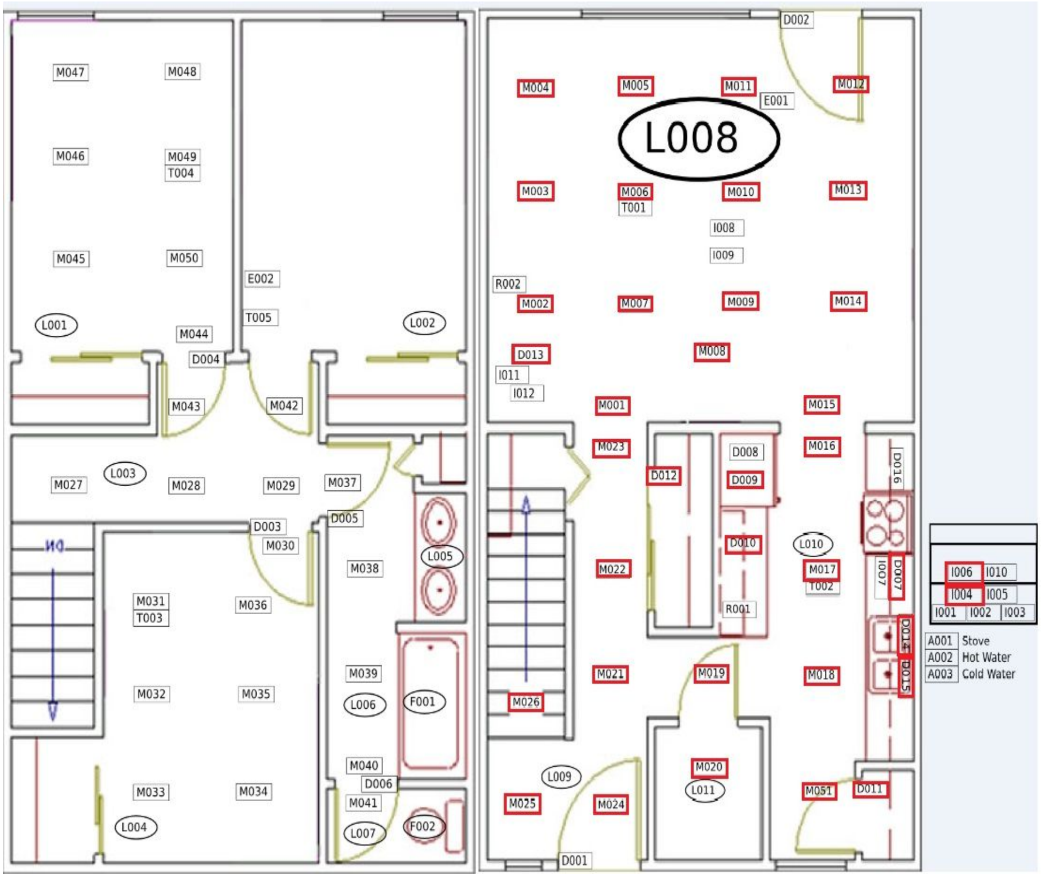

Figure 19: Floor plan of Kyoto of WSU CASAS

Because this dataset stemmed from a laboratory experiment on a voluntary basis, the recording of the activities is not continuous in time and is spread over two months. However, it is a very good benchmark for the community and was used in most of the studies in the context of multi-occupant HAR (Chen and Tong, 2014) (Chiang et al., 2010) (Cook et al., 2010) (Hsu et al., 2010) (Singla et al., 2010). Additional details on the dataset are given in (Singla et al., 2010).

III.4.1.1 Multi-occupant activity data sequences segmentation

In order to accurately learn the activity model parameters, we need to produce sequences from the raw data. The design of the training sequences does not affect much observations and transitions estimation, in contrast to initial state estimations. The initial states do get a better estimate when accurate activity data segmentation is applied. The design of sequences strongly depends on the type of activities monitored in the environment. Generally, studies consider each day of registration as a sequence when activity data reflects the natural human behavior of occupants as in (Kasteren et al., 2008). But in the case of CASAS "Multiresident ADLs", many days represent activity data of a single pair of volunteers, while others represent activity data of two pairs of volunteers as shown in table 8. Segmenting activity data on a daily basis would result into a lot less samples for initial and transition estimates. Therefore, each file serves to build one sequence. We run leave-one-out cross-validation on our 26 sequences of activity data.

Table 8: Summary of dates and number of event for volunteer files

Volunteer ID	date	# of events
1	10 NOV 2008	674
2	11 NOV 2008	788
3	11 NOV 2008	826
4	12 NOV 2008	765
5	12 NOV 2008	810
6	15 NOV 2008	699
7	17 NOV 2008	576
8	17 NOV 2008	757
9	18 NOV 2008	390 (missing)
10	18 NOV 2008	693
11	19 NOV 2008	623
12	19 NOV 2008	572
13	20 NOV 2008	616
14	21 NOV 2008	623
15	21 NOV 2008	612
16	2 DEC 2008	549
17	3 DEC 2008	650
18	3 DEC 2008	715
19	4 DEC 2008	630
20	5 DEC 2008	558
21	5 DEC 2008	500
22	8 DEC 2008	620
23	9 DEC 2008	683
24	9 DEC 2008	661
25	10 DEC 2008	778
26	10 DEC 2008	866

III.4.1.2 Multi-occupant activity data Pre-processing

In order to compare the performance of our multi-occupant models against other models like CHMM, we followed the same data pre-processing procedure applied in (Chiang et al., 2010). As mentioned earlier, annotation of sensor events and activities in "Multiresident ADLs" comes in the format of (Date, Time, SensorID, Value, OccupantID, ActivityID). The pre-processing procedure aims to represent at each time step the sensor event, its corresponding value as well as the given activity label for each occupant (i.e. Date, Time, Occupant 1 <SensorID, Value, Activity label>, Occupant2 <SensorID, Value, Activity label>). Then, we first separate the events according to their occupant IDs and generate the sequence of activities performed by each occupant using the following procedure, the latter is applied on each of the 26 sequences:

At the beginning of the sequence, a null activity as well as a null observation are assigned to each occupant (i.e. *at time t Occupant 1 (null, null, null), Occupant2 (null, null, null)*). Each time a sensor event is generated by Occupant 1 his/her observation value and his/her activity value are updated to the new ones and both observation and activity values of Occupant 2 remain unchanged (e.g. *at time t+1* the sensor M19 with the value ON is generated by the Occupant 1, the given activity label is 1 this would correspond to *Occupant 1 (M19, ON, 1), Occupant2 (null, null, null)*). The same procedure is applied when a sensor event is generated by Occupant 2 (e.g. *at time t+2* the sensor M23 with value ON is generated by the Occupant 2, the given activity label is 4, this would correspond to *Occupant 1 (M19, ON, 1), Occupant2 (M23, ON, 4)*). In case a sensor event is trigged by both occupants, observations of both occupants as well as their activity labels are updated to the new ones (e.g. *at time t+3* the sensor M19 with the value OFF is generated by Occupant 1 and Occupant 2 with the activity labels 5 and 6 respectively, that would correspond to *Occupant 1 (M19, OFF, 5), Occupant2 (M19, OFF, 6))*. Table 9 shows a sample of data after applying the above pre-processing procedure.

Table 9: Sample of first sequence activity data (i.e. p01.txt) before and after pre-processing procedure

Sensor ID	Value	Occupant ID	Activity ID	Occupant 1			Occupant 2		
				Sensor ID	Value	Activity Label	Sensor ID	Value	Activity Label
				NULL	NULL	NULL	NULL	NULL	NULL
M22	ON	2	2	NULL	NULL	NULL	M22	ON	2

M19	ON	1	1	M19	ON	1	M22	ON	2
M23	ON	2	2	M19	ON	1	M23	ON	2
M18	ON	1	1	M18	ON	1	M23	ON	2
M01	ON	2	2	M18	ON	1	M01	ON	2

The data contains 16 activities for each occupant, 15 known activities and one void (Null) activity that represents unknown activities. In all, there are 37 different binary sensors in the dataset resulting in 75 observation values for each occupant including the null value.

III.4.2 Comparison against PHMM and CHMM

To further illustrate the performance of our two HMM-based approaches, CL-HMM and LHMM, we will consider comparing them against PHMM and CHMM. The models PHMM and CHMM rely on the assumptions we made about the CL-HMM and the LHMM.

III.4.2.1 Parallel HMM (PHMM)

a. Definition

A PHMM consists of a set of independent HMMs. In other terms, PHMMs are standard HMMs that are used in parallel under the assumption that the corresponding individual processes being modeled evolve independently from one another with independent output. That is, this kind of model combines HMMs without considering any relationship between them. Therefore, when applied for activity recognition in a *k*-occupant environment (Chiang et al., 2010), PHMM will consist of *k* independent HMMs, one for each occupant, where the hidden states correspond to the activities and the observations correspond to the sensor values. While the application of PHMMs is easy and straightforward, their capabilities are limited in the context of multi-occupant activity recognition due to the lack of interaction between HMMs and data independence between the models. This is the main reason why PHMM have been mainly used for modeling parallel activities rather than modeling the interaction.

Figure 20 (blank squares indicate the hidden states, the shaded squares indicate the observed states) shows a PHMM obtained by modeling each occupant as a separate HMM that is $\{A^1, O^1\}$and $\{A^2, O^2\}$ for Occupant 1 and Occupant 2 respectively.

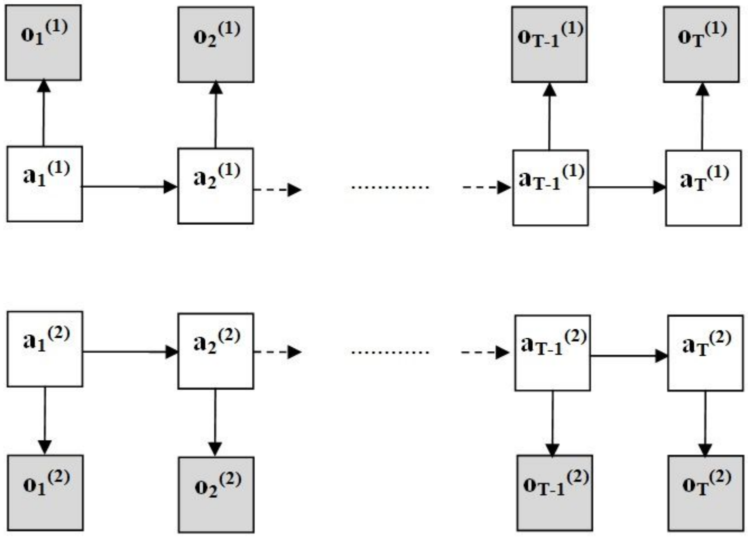

Figure 20: Topology of PHMM

b. Parameter estimation

There are many existing algorithms for training HMMs such as Baum-Welch. Since we have the activity label for each occupant sensor event in the dataset for both Occupant 1 and Occupant 2 (supervised learning), parameters estimation is straightforward and would be simply achieved by frequency counting of occurrences of initial states, transitions and observations for each occupant chain, resulting then into an initial state vector, a transition matrix and an observation matrix for each occupant's chain. The parameters of a PHMM are defined below:

$$\pi^R(i) = P(a_1^R = i) \tag{III.24}$$

$$Q_{ij}^R = P(a_t^R = j \middle| a_{t-1}^R = i) \tag{III.25}$$

$$B_t^R(i) = P(o_t^R | a_t^R = i) \tag{III.26}$$

where R={1,2}, R is the occupant index, i=1...N in case R=1 (i.e. N is the number of activities performed by Occupant1 while i=1...M in case R=2 (i.e. M is the number of activities performed by Occupant 2).

c. Inference

Since the two HMMs in the PHMM are independent, the posterior of activities given the observation is just the multiplication of the two HMMs, which is:

$$\begin{aligned} P\left(A^{(1)}, A^{(2)} \middle| O^{(1)}, O^{(2)}\right) \\ &= \prod_{R=\{1,2\}} P(A^R | O^R) = \prod_{R=\{1,2\}} \frac{P(A^R) P(O^R | A^R)}{P(O^R)} \\ &\propto \prod_{R=\{1,2\}} P(A^R) P(O^R | A^R) \end{aligned} \tag{III.27}$$

and

$$\begin{aligned} \prod_{R=\{1,2\}} P(A^R) P(O^R | A^R) \\ = \prod_{R=\{1,2\}} P(a_1^R) \left(\prod_{t=2}^{T} P(a_t^R | a_{t-1}^R) \right) \left(\prod_{t=1}^{T} P(o_t^R | a_t^R) \right) \end{aligned} \tag{III.28}$$

In PHMM, the inference for one chain is independent of the other chain. Specifically, given an observation sequence O^1 and the already learned parameters of the HMM corresponding to Occupant1 that is ($\pi^1(i)$, Q_{ij}^1, $B_t^1(i)$), we need to find a state sequence A^1 which maximizes $P(A^1|O^1)$. Thus, the conventional Viterbi algorithm for HMMs (see appendix A) (Rabiner, 1989) is applied on the observation sequence O^1 in order to compute the most probable state sequence A^1. Similarly, for inference on Occupant 2 chain, the Viterbi algorithm is applied on O^2 to compute the most probable state sequence A^2. The computational complexity of the algorithm is then $O(TN^2+TM^2)$, where T is the total number of sensor events in the dataset (i.e. length of the entire dataset). Considering R occupants, each having a number of corresponding activities N_i, i=1...R, the computational complexity would be $O(\sum_{i=1}^{R} T\, N_i^2)$.

III.4.2.2 Coupled HMM (CHMM)

a. Definition

CHMM, on the other hand, is a combination of HMMs that interact with each other (see figure 21). In each HMM, there is a directed edge from each hidden state at time t to the hidden state at time t+1. In addition, there are direct edges from each hidden state at time t of an HMM to all hidden states of the other HMMs at time t+1 to indicate the interaction between the occupants when performing cooperative activities.

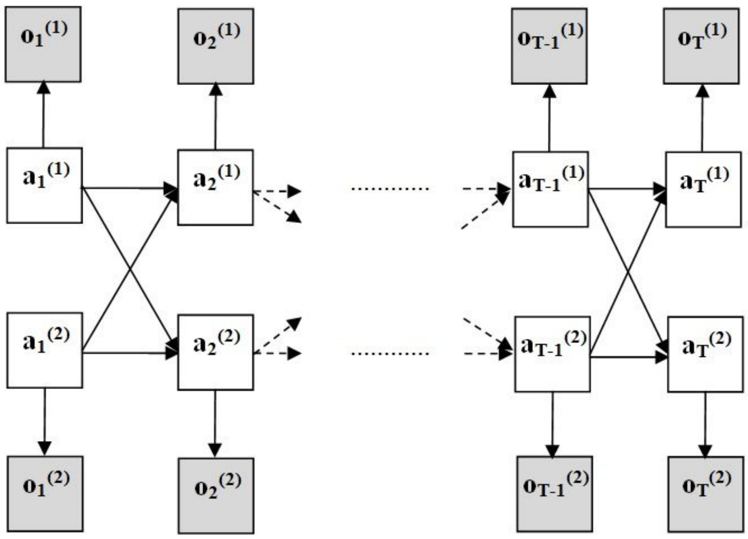

Figure 21: Topology of CHMM

b. Parameter estimation

CHMM parameters are obtained by frequency counting of occurrences of initial states, transitions and observations for each occupant chain, resulting then into an initial state vector, a transition matrix and an observation matrix for each occupant's chain. The only difference between the parameters of a PHMM and the ones of a CHMM lies in the transition matrices corresponding to Occupant 1 chain and Occupant 2 chain as shown below:

$$Q^R_{i|j,k} = P(a^R_t = i \middle| a^1_{t-1} = j, a^2_{t-1} = k)$$

(III.29)

$Q_{i|j,k}{}^{1}$ is an N×M by N matrix and $Q_{i|j,k}{}^{2}$ is N×M by M matrix.

This means that the process doesn't differ for computing the initial state vector and the observation matrix of each occupant chain. However, in addition to estimating inner-chain transition probabilities $P(a_t^{(1)}|a_{t-1}^{(1)})$ and $P(a_t^{(2)}|a_{t-1}^{(2)})$ in transition matrices of the individual chains, CHMM requires to estimate the inter-chain transition probabilities $P(a_t^{(1)}|a_{t-1}^{(2)})$ and $P(a_t^{(2)}|a_{t-1}^{(1)})$ which represent the incoming transitions from Occupant 2 chain to Occupant 1 chain and incoming transitions from Occupant 1 chain to Occupant 2 chain respectively. These are used to estimate the likelihoods $P(a_t^{(1)}|a_{t-1}^{(1)}, a_{t-1}^{(2)})$ and $P(a_t^{(2)}|a_{t-1}^{(1)}, a_{t-1}^{(2)})$ for Occupant 1 chain and Occupant 2 chain respectively. For instance, $Q_{3|1,4}{}^{1}$ represents the number of outgoing transitions from the pair of activities (1,4) and arriving in activity 3 in Occupant 1 chain divided by the total number of outgoing transitions from the pair of activities (1,4) to Occupant1's chain.

c. Inference

Since the two HMMs are no longer independent in a CHMM, the posterior of the activity sequences given all the observation becomes:

$$\begin{aligned} P\left(A^{(1)},A^{(2)}\middle|O^{(1)},O^{(2)}\right) &= \frac{P\left(O^{(1)},O^{(2)}\middle|A^{(1)},A^{(2)}\right)P(A^{(1)},A^{(2)})}{P(O^{(1)},O^{(2)})} \\ &\propto P\left(O^{(1)},O^{(2)}\middle|A^{(1)},A^{(2)}\right)P(A^{(1)},A^{(2)}) \end{aligned}$$

(III.30)

Given the condition independence in the graphical structure of figure 21, we can factorize $P\left(O^{(1)},O^{(2)}\middle|A^{(1)},A^{(2)}\right)$ and $P\left(A^{(1)},A^{(2)}\right)$ as follows:

$$\begin{aligned} &P\left(O^{(1)},O^{(2)}\middle|A^{(1)},A^{(2)}\right) \\ &\quad = P\left(o_1^{(1)},o_1^{(2)}\middle|a_1^{(1)},o_1^{(2)}\right)\dots P\left(o_T^{(1)},o_T^{(2)}\middle|a_T^{(1)},o_T^{(2)}\right) \\ &\quad = \prod_{t=1}^{T} P\left(o_t^{(1)}\middle|a_t^{(1)}\right)P\left(o_t^{(2)}\middle|a_t^{(2)}\right) \end{aligned}$$

(III.31)

And

$$P(A^{(1)}, A^{(2)}) = P\left(a_1^{(1)}, a_1^{(2)}\right) \prod_{t=2}^{T} P(a_t^{(1)}, a_t^{(2)} \left| a_{t-1}^{(1)}, a_{t-1}^{(2)}\right.)$$
$$= \prod_{R=\{1,2\}} P(a_1^R) \prod_{t=2}^{T} P(a_t^R | a_{t-1}^{(1)}, a_{t-1}^{(2)}) \tag{III.32}$$

In contrast, inference in CHMM is not independent from Occupant 1 chain to Occupant 2 chain. The inputs of the Viterbi algorithm (Nefian et al., 2002) consist of observations of both occupants O={O^1, O^2} and the algorithm outputs the best state sequence A={A^1, A^2} which maximizes *P(A|O)*. Details of the Viterbi algorithm are given in appendix B and its computational complexity is about $O(TN^2M^2)$, where T is the total number of sensor events in the dataset (i.e. length of the entire dataset). Considering R occupants, each having a number of corresponding activities N_i, i=1...R, the computational complexity would be $O(T\prod_{i=1}^{R} N_i^2)$.

III.4.3 Experiment 1: Results by individual occupants

a. Description

In this experiment results are computed for each occupant separately without any interference. That is the misclassification of the activity by Occupant 1 doesn't impact the correct classification of Occupant 2 and vice versa. In PHMM, CHMM a pair of labels is inferred representing Occupant 1 activity label and Occupant 2 activity label while in CL-HMM and LHMM a combined activity label is inferred. We convert the inferred combined activity label for the two latter models into the corresponding inferred Occupant 1 activity label and inferred Occupant 2 activity label. As the activities come with their true labels, each inferred occupant activity label is compared to its corresponding true label. We therefore compile for each occupant a confusion matrix and compute the overall accuracy, precision, recall and F-measure which are given as follows:

$$Overall\_Accuracy = \sum_{i=1}^{C} {}^{tp_i}/_{N} \tag{III.33}$$

$$Precision = \frac{1}{C}\sum_{i=1}^{C}(tp_i/tp_i + fp_i) \tag{III.34}$$

$$Recall = \frac{1}{C}\sum_{i=1}^{C}(tp_i/tp_i + fn_i) \tag{III.35}$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{III.36}$$

where N is the total number of events, C is the number activities, tp_i is the number of true positive, fn_i is the number of false negative, fp_i is the number of false positive for activity i.

These measures are chosen because they are the most popular ones (see Section II.5.8). The overall accuracy is defined as the percentage of correctly classified events. We compute the precision, recall and f-measure according to the definition given in (Kasteren et al., 2010). That is, we compute the precision and recall for each class (i.e. activity) separately and take the average over all classes. Precision is defined as the percentage of inferred activity labels which was correctly classified. Recall is defined as the percentage of true activity labels which was correctly classified. The f-measure is a combination of the two latter metrics. The recall is equivalent to the average accuracy (i.e. the average percentage of correctly classified events per activity). Most HAR datasets are imbalanced which means that some classes (i.e. activities) appear much more frequently than other classes. Considering the experimental dataset that is "Multiresident ADL of CASAS", figure 22 clearly demonstrates the imbalance in the percentage of sensor events between all activities (e.g. the percentage of sensor events differs greatly from one activity to another one: about 6% for the activity 1, 8% for the activity 2, 2% for the activity 5, 9% for the activity 6 and 16% for the activity 13).

If dominant classes yield a good recognition performance, the overall accuracy would be high even if all other classes are not well recognized, but, recall will be low. Because all activities are equally important, precision, recall and f-measure seem a better choice in order to demonstrate the HAR performance on each of the different activities.

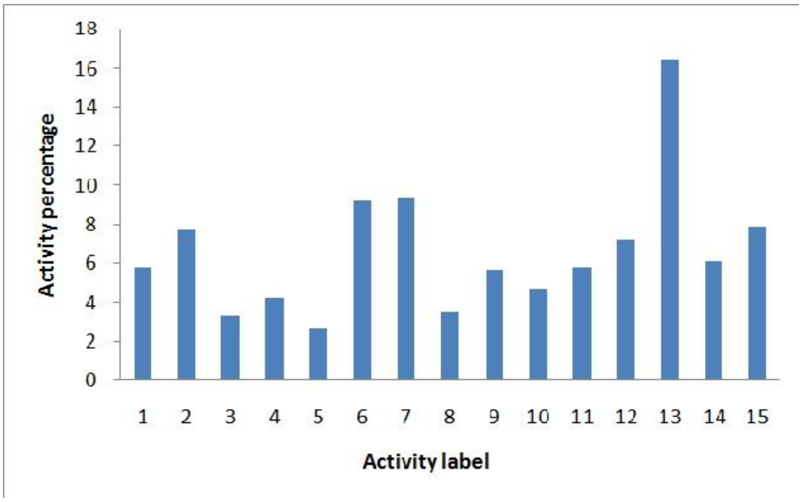

Figure 22: Sensor events' percentage for each activity of "Multiresident ADLs"

We separated the overall accuracy reported by each occupant for parallel individual activities and cooperative activities as done in (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b). For each occupant, cooperative activity labels are 3, 7, 11 and 13. The remaining activities are parallel ones. An occupant is supposed to perform a cooperative activity at time when its corresponding activity label is included in {3, 7, 11, 13}, even if the other occupant performs an individual activity in parallel. For each occupant, we extract cooperative activities accuracy which represents the overall accuracy over the activities 3, 7, 11 and 13 and parallel individual activities accuracy which represents the overall accuracy on the remaining ones {1, 2, 4, 5, 6, 8, 9, 10, 12, 14, 15, 16}. Cooperative activities accuracy for both occupants is then obtained by computing the average of these one over the two occupants. Likewise for parallel individual activities accuracy for both occupants, after computing this latter for each occupant, we give the average of them over both occupants. These ones provide more insight on the ability of each model to deal with the two types of activities.

Please note that the code used in this study was implemented in MATLAB. The parameter estimation of all the models is computed by frequency counting of occurrences of initial state vector, transition of hidden states and observations. Some part of the code is obtained from (Kasteren et al., 2010). We also used the

Viterbi decoder of the HMM toolbox of Kevin Murphy[20] for our PHMM, CL-HMM, and LHMM.

b. Results

Results, shown in Table 10, are given for each of the proposed models, CL-HMM and LHMM, as well as for the state-of-the-art models used for multi-occupant HAR namely PHMM and CHMM. We cycle over all the 26 training sequences using the leave-one-out cross validation and report the average performance measure for each evaluation metric.

Table 10 shows that each model produces similar results for the overall accuracy and recall which means that each model recognized all the activities with an equal performance (i.e. the recognition rate does not vary greatly from one activity to another one).In fact, using all of the metrics, it is easy to see in table 10 that LHMM outperforms all of the other models, while CL-HMM looks more accurate than CHMM and PHMM. The least accurate recognizer is PHMM, which is expected to a large extent. In particular, in comparison with CHMM, LHMM improves the overall accuracy, precision, recall and f-measure by roughly 5%, 4.5%, 5% and 5% respectively; whereas CL-HMM does by about 3%, 3%, 5% and 4% respectively. Moreover, when looking at the accuracies separated for parallel individual occupant activities and cooperative activities, it is clear that CHMM improves the recognition of parallel individual activities by 2.5% in comparison to PHMM. On the other hand, compared to CHMM, LHMM and CL-HMM improve the accuracy by approximately 7% and 6% respectively.

For cooperative activities, the experiments show that CHMM outperforms PHMM with a difference of 7%, while LHMM improves the recognition of cooperative activities by approximately 4% in comparison with the CHMM. However, CL-HMM seems to be the best with an improvement of 20.65%, 14% and 9.5% in comparison to PHMM, CHMM and LHMM. Moreover, for cooperative activities the breakdown by occupants show that the performance difference is about 30% from Occupant 2 to Occupant 1 for PHMM, CHMM and LHMM. CL-HMM performs similarly for both occupants (92.76% for Occupant 1 and 91.22% for Occupant2).

---

[20] http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html

Table 10: Results breakdown by occupant

Approach	Results	Overall over all activities	Separated for Individual and Cooperative activities: Individual activities (1)	Cooperative activities (2)	Average on 1&2	Precision	Recall	F-measure
PHMM	R1[21]	84.98 ± 9.9	87.08 ± 9.07	53.27 ± 47.12	72.39 ± 23.19	86.26 ± 9.66	84.87 ± 7.98	85.42 ± 8.23
	R2	83.16 ± 8.34	79.19 ± 12.26	85.55 ± 12.73	82.37 ± 8.11	85.56 ± 7.81	83.54 ± 7.97	84.46 ± 7.56
	Average on R1 &R2	**84.07 ± 7.09**	**83.13 ± 8.17**	**71.68 ± 24.04**	**77.4 ± 12.33**	**85.91 ± 6.85**	**84.2 ± 5.98**	**85 ± 6.06**
CHMM	R1	87.85 ± 9.15	89.29 ± 8.61	60.76 ± 44.82	77.05 ± 22.37	89.55 ± 8.54	87.62 ± 8.25	88.48 ± 7.96
	R2	88.37 ± 7.1	82.13 ± 10.82	92.26 ± 8.31	87.2 ± 7.39	88.52 ± 6.23	87.21 ± 7.98	87.75 ± 6.63
	Average on R1 &R2	**88.11 ± 7.48**	**85.71 ± 8**	**78.53 ± 23.09**	**82.12 ± 12.86**	**89.03 ± 6.21**	**87.41 ± 6.82**	**88.17 ± 6.22**
CL-HMM	R1	91.33 ± 8.15	91.11 ± 8.41	92.76 ± 21.87	91.78 ± 11.68	92.25 ± 6.99	92.54 ± 6.59	92.38 ± 6.71
	R2	91.61 ± 7.87	92.37 ± 6.64	91.22 ± 11.07	91.8 ± 6.96	91.12 ± 7.43	91.7 ± 7.99	91.35 ± 7.5
	Average on R1 &R2	**91.47 ± 7.5**	**91.74 ± 6.07**	**92.33 ± 11.24**	**91.91 ± 7.3**	**91.68 ± 6.1**	**92.12 ± 6.42**	**91.89 ± 6.17**
LHMM	R1	92.36 ± 8.48	93.86 ± 7.89	65.19 ± 43.57	81.4 ± 21.32	93.25 ± 7.46	91.93 ± 7.56	92.48 ± 6.98
	R2	94.17 ± 5.05	90.8 ± 7.52	96.42 ± 5.48	93.61 ± 5.12	93.9 ± 5.44	93.43 ± 6.4	93.61 ± 5.63
	Average on R1 &R2	**93.27 ± 6.21**	**92.33 ± 6.95**	**82.77 ± 21.3**	**87.53 ± 11.22**	**93.58 ± 5.41**	**92.68 ± 6.18**	**93.1 ± 5.62**

[21] R1 indicates Occupant 1 and R2 indicates Occupant 2

III.4.4 Experiment 2: Joint results for both occupants

a. Description

In this second experiment, we are interested to measure the recognition performance of the pair of activities by both occupants. The pair of activity labels is assumed to be correctly classified when both Occupant 1 activity and Occupant 2 activity are correctly classified, that is when the combined activity label is correctly classified. As a result, the misclassification of individual activities of both occupants impacts the overall classification outcome. In the inference step of CL-HMM and the LHMM, both models infer a combined activity label. PHMM and CHMM, on the other hand, infer individual activity labels which are then combined to form pairs of combined labels. In order to run this experiment, we first define the true label of the combined label for each pair of activities against which the inferred combined label is compared to compute the evaluation measures: overall accuracy, precision, recall and f-measure. The overall accuracy is equivalent to the joint accuracy measure computed in (Chiang et al., 2010).

We also separate the overall accuracy results for parallel individual activities and cooperative activities. Note that occupants are considered to perform parallel unrelated activities if Occupant 1 activity and Occupant 2 activity are different (e.g. Occupant1 activity label=1 and Occupant 2 activity label=2). In contrast, occupants perform cooperative activities when the activity labels are the same. Among all cooperative activities defined in the dataset which are 3, 7, 11, 13, only activity 13 appears for both occupants at the same time in the training data. The activities 3, 7 and 11 appear with other parallel individual activities.

b. Results

Table 11 shows the results obtained by all models using leave-one-out cross validation. Clearly the outcome of accuracy, precision, recall and f-measure indicate that the proposed LHMM outperforms the rest of the models. In particular, the overall accuracy results indicate that CL-HMM performs much better than PHMM and CHMM; while results related to precision, recall and

Table 11: Joint Accuracy results for occupants

Approach / Results	Overall accuracy	Overall accuracy separated for individual and cooperative activities			Precision	Recall	F-measure
		Individual (1)	Cooperative(2)	Average on 1& 2			
PHMM	72.8 ± 11.13	74.79 ± 10.59	40.23 ± 46.22	59.55 ± 22.63	65.95 ± 9.94	71.89 ± 11.26	68.64 ± 10
CHMM	81.65 ± 10.27	82.81 ± 10	60.72 ± 44.76	73.21 ± 22.08	73.77 ± 10.88	77.6 ± 10.74	75.5 ± 10.4
CL-HMM	86.04 ± 10.84	85.57 ± 11.23	88.03 ± 28.77	86.78 ± 17.02	72.26 ± 11.01	76.75 ± 9.38	74.32 ± 9.81
LHMM	88.23 ± 10.23	89.46 ± 9.58	65.21 ± 43.56	79.06 ± 21.63	81.46 ± 10.27	79.43 ± 10.35	80.3 ± 9.84

f-measure show that CL-HMM and CHMM perform similarly. Note also that although PHMM is known to report good results in the case of individual activities performed in parallel, results of overall recognition accuracy of individual activities show that PHMM is outperformed by CL-HMM, LHMM and CHMM by approximately 10.78%, 14.67% and 8.02% respectively. LHMM performs the best in the case of individual activities producing 3.89% and 6.65% better than CL-HMM and CHMM respectively.

For cooperative activities, PHMM achieves a very low performance and is less accurate than CHMM with a difference of 20.5% which is consistent with our expectations. The difference in performance between LHMM and CHMM, which is 4.49%, is much smaller than that between CL-HMM and CHMM (i.e. 27.31%).

This illustrates that CL-HMM is the best model for recognizing cooperative activities followed by LHMM.

III.5 Comparison against existing studies on "Multiresident ADLs" dataset of CASAS

Several studies used the same dataset (Hsu et al., 2010) (Cook et al., 2010) (Singla et al., 2010) (Chen and Tong, 2014) (Chiang et al., 2010). The common metrics used by these studies are the average accuracy (i.e. recall) and the overall accuracy. Hence, our decision to these measures for evaluating the various models.

(Cook et al., 2010) developed an integrated system for data association and activity recognition based on CRF and HMM respectively. This study reported an average accuracy of 50.67% and 90% respectively. While (Singla et al., 2010) reported an average accuracy of 60.60% using HMMs. Nevertheless, these studies do not consider cooperative and parallel activities.

(Hsu et al., 2010) reported an average accuracy of 64.16% for their independent CRFs for occupants. (Singla et al., 2010) reported an average accuracy of 73.15% for their independent HMMs for occupants (i.e. PHMM). The average accuracy of our PHMM is higher by 11% which is may be due to data pre-processing. Note that the authors did not provide details about the pre-processing of the activity data. Moreover, three-fold cross validation was used; while in our case leave-one-out cross validation was applied. Nevertheless, these approaches only recognize parallel activities.

The only studies that considered both parallel individual and cooperative activities are (Chen and Tong, 2014) and (Chiang et al., 2010). The authors in (Chen and Tong, 2014) applied a combined label approach on the pair of activity using HMM and CRF. They reported an average accuracy (i.e. computed from the average accuracy of Occupant 1 and average accuracy of Occupant 2) of 75.77% and 75.38% respectively. Although, the performance of our CL-HMM seems to be better than theirs, we do not have the same experimental setting.

In (Chiang et al., 2010), a leave-one-out cross validation was applied producing an accuracy of 77.38%, 82.82% and 85.58% and a joint accuracy of 61.78%, 74.78% and 78.26% for their PHMM, CHMM and CHMM with the interaction feature respectively. On the contrast, our PHMM and CHMM produce an overall accuracy of 84.07% and 88.11% and overall joint accuracy (First column in table 11) of 72.8% and 81.65% respectively. However, this comparison is subjective because the authors in (Chiang et al., 2010) did not provide details about the way they computed the accuracy. Note we used the same pre-processing as theirs.

III.6 Conclusion

In this chapter we proposed two HMM-based models, Combined-label HMM and Linked HMM to investigate the problem of multi-occupant activity recognition. These two models were compared against the state-of-the-art baseline methods which are Parallel HMM and Coupled HMM. Our first experiment shows that the proposed models outperform the baseline models for both cases of parallel individual activities and cooperative activities. In particular, CL-HMM not only significantly improves the recognition accuracy of cooperative activities but also performs equivalently in recognizing the individual activities of Occupant 1 and Occupant 2. That is, the recognition rate of the activities does not vary from Occupant 1 to Occupant 2 in contrast to the other models. This is an important aspect in a multi-occupant setting as the recognition of the activities from one occupant to another one is equally important.

Our second experiment shows another important aspect of multiple-occupant monitoring which is about the correct inference of the activities of all occupants at any time. CL-HMM performed the best in the case of cooperative activities; while LHMM performed the best in the case of parallel individual activities. This is to say that the proposed models are not only valuable for dealing with cooperative activities but also for individual parallel ones.

CHAPTER IV
General Conclusion

IV. General Conclusion

IV.1 Conclusion

Home automation or SH is a home, that is equipped with electronic devices (i.e. sensors, controllers). It enables occupants to control home appliances remotely or automatically (e.g. an occupant can use a Smartphone to arm the home security system, adjust the temperature, control lighting).

Recent advances in networking and electronic technologies such as miniaturization and low cost nature of infrastructure sensors (i.e. non intrusive sensors) lead to an acceptance of SHs systems and as a result conducted to a rapid emergence of these ones.

Because of the continued increase of the aging population, the lack of nursing homes and health care professionals designed to manage them, one of the main motivation of SHs is the monitoring of older adults wellbeing in order to maintain them at home (i.e. independent living) for longer durations. One of the important services that can be offered by such a system is remotely assessing the physical and cognitive capabilities of the people by monitoring their ADLs, such as sleeping, cooking, eating and going out. In this context HAR aims to recognize the sequence of ADLs performed by the occupants at home. The completion of these ADLs represent an indicator of occupant's health status and is used to characterize human behaviors. Moreover, the continuous monitoring of activities allows to detect anomalies in order to intervene quickly with the required assistance.

HAR in single occupant context has been widely invested and a diversity of HAR models exist ranging from probabilistic graphical models (e.g. HMM and CRF), conventional statistical machine learning techniques (e.g. DTs, ANNs) to ontology modeling, in contrast to multi-occupant HAR.

The nature of human activities are usually more complex in a multi-occupant environment in comparison with a single occupant environment. While single occupant environments imply designing a model to recognize sequential, interleaved and concurrent ADLs, multi-occupant environments must also manage and recognize parallel and cooperative activities. Details and main concepts about SHs, ADLs and multi-occupant problems as well as our research focus are discussed in Chapter I.

We focus in this thesis on HAR from infrastructure sensor readings in multi-occupant environment. The latter problem poses two main challenges: (i) data association that is identifying the occupant responsible for a sensor triggering at each time step (ii) recognizing multi-occupant activities which are parallel and cooperative activities. We only focus on the latter problem.

Regarding state-of-the-art studies on mutli-occupant HAR presented in Chapter II, we showed that a number of studies in the pervasive environment have been conducted. In term of sensor deployment there exist two main classes of approaches. Those relying on wearable sensors (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b). Those relying on infrastructure sensors. In this context, some studies focused on the data association problem which is about recognizing the occupants (Hsu et al., 2010) (Cook et al., 2010). Although, we are not dealing with the problem of data association in this thesis, we argue that developing solutions for this problem is crucial for deploying HAR systems. For instance, (Hsu et al., 2010) showed that the quality of data association results impacts the quality of activity recognition if both are integrated in one system.

Other studies not dealing with data association also proposed to recognize the activities without differentiating between the occupants. Singla et al. (Singla et al., 2010) used a single chain HMM to model the activities of two occupants (e.g. Occupant 1 performs activity 1, 2 and 3; while the Occupant 2 performs activity 3, 4 and 5, In the inference step, one activity label is inferred; it represents either activity of Occupant 1 or activity of Occupant 2). As in (Hsu et al., 2010) (Cook et al., 2010), these approaches do not recognize neither parallel activities nor cooperative ones and are therefore not suitable for HAR in multi-occupant setting. Still not considering the data association problem, some studies proposed a separate model for each occupant as shown in the second approach of (Hsu et al., 2010) and the second approach of (Singla et al., 2010). These studies, however, did not deal with cooperative activities and only recognized parallel ones. Few approaches did consider both parallel and cooperative activities between occupants (Alemdar et al., 2013) (Chen and Tong, 2014) (Chiang et al., 2010). We concluded that all proposed approaches were essentially probabilistic based on graphical models.

Chapter III investigates multi-occupant HAR direction further. Our aim is to accurately recognize both parallel and collaborative activities from non-intrusive sensors. We do not focus on only one of the two types of activities as done in the literature related to multi-occupant activity recognition, but on both types.

The HMM is a conventional graphical probabilistic model that has been studied for years and is very well understood. This model has been successfully applied in many sequential data modeling problems such as speech recognition, handwritten digit recognition, biological sequence analysis as well as HAR. Some activities' characteristics have a significant impact on the activity recognition performance. For example an activity can be performed in different ways that is in non-deterministic way and sometimes with a different sequential ordering regarding the actions composing the activity. HMMs are robust to the noise in sensor readings and to the uncertainty while performing activities. Moreover, the HMM is capable of considering the sequential nature of activities.

Hence we suggest two variants of HMM named CL-HMM and LHMM and compared our two approaches to conventional models (i.e. PHMM, CHMM). The CL-HMM represents a variant of the combined label approach based on HMM applied in (Chen and Tong, 2014) (Alemdar et al., 2013) and is based on the use of one HMM chain on both combined activities labels and combined observations labels. Applying the LHMM, we proposed one

HMM is constructed for each occupant. Hidden states in each HMM represent the set of activities performed by the occupant. Likewise for the CHMM and the PHMM. However, co-temporal dependencies between activities of occupants are represented differently in CHMM and LHMM while in PHMM no dependency exists between occupants' activities. Considering, these co-temporal dependencies in CHMM, each activity in HMM does not depends on only the previous activity at time t-1 of the same HMM, but also on the previous activities at time t-1 from the other HMMs. In LHMM, activities of all HMMs corresponding to the occupants are joined at each time step.

Two main experiments are studied with a special attention on to the performance of the models on cooperative and parallel activities. The first experiment presents the results of HAR for each occupant using all models. Compared to CHMM, LHMM and CL-HMM improved individual activities accuracy as well as cooperative accuracy by approximately 7%, 6% and 4%, 14% respectively. Although both proposed approaches performed equivalently while recognizing parallel individual activities we point out that the CL-HMM performed the best in the recognition of cooperative activities (i.e. difference in performance between the two models is significant 9.5%). Another important result while comparing cooperative activities results breakdown by occupants is that the performance difference is about 30% from Occupant 2 to Occupant 1 for PHMM, CHMM and LHMM while CL-HMM performs similarly for both occupants. The latter result emphasizes the importance of a balanced model which recognizes the activities equivalently from an occupant to another one in a multi-occupant environment.

An important aspect of multi-occupant activity recognition is advocated by experiment 2. Specifically, the correct inference of occupants' activities at each time step. The aim is to correctly infer the activity of each occupant present in the living space at each time step. Hence, joint results for both occupants using all the models are discussed in this experiment. If both occupants 'inferred activities labels are equal to occupants 'true activities labels at time t, occupants 'activities are then assumed to be correctly classified. As a result, the misclassification of the activity of one of the two occupants at time impacts the overall classification outcome in contrast to experiment 1. We concluded from this experiment that LHMM is the best model for the recognition of parallel individual activities (i.e. 3.89% and 6.65% better than CL-HMM and CHMM respectively) While the CL-HMM is more appropriate to recognize cooperative activities (i.e. improvement of 27.31% in comparison to CHMM) followed by the LHMM model.

Although state-of-the-art studies reported that PHMM performs well in the case of individual activities performed in parallel, our two experiments shows that the PHMM model is the least accurate model for both types of activities.

Most studies in the literature focused on developing a graphical probabilistic model-based solution to better recognize cooperative activities assuming that interactions between occupants 'individual chains improve the recognition rate of these activities. Our experiments prove that these added interactions between occupant's models are also valuable for the recognition of individual activities performed in parallel.

We conclude from our two experiments that both proposed approaches outperformed baseline ones and are not only appropriate for the recognition of cooperative activities but also for individual parallel ones.

IV.2 Future Research

There are a number of directions to move our research forward.

(1) **Due to the novelty of the field of multi-occupant activity recognition, there are many open issues with respect to sensory data pre-processing steps such as discretization and feature representation.**
Sensory data used in our experiments is represented as a sequence of events. It is a binary sensor data and every change in a sensor state (i.e. value) generates an event. We assigned two different symbols for each change in sensor value (e.g. symbol "1" and symbol "2" to "ON" and "OFF" values respectively for the motion sensor "M01"). Hence, the observation variable is multinomial with support {1,...N} while N is the number of encoding symbols. The use of events avoid us the need for discretization.
We plan to discretize the sensor data using a constant time slice after applying the pre-processing procedure (see Section III.4.1.2). Hence, observations are referred to as timeslices and each time slice is represented as a binary vector whose length corresponds to the number of sensors. At time t, a position in the vector is set to 1 if the i^{th} sensor changed state. We also plan to experiment different lengths of the time interval used for discretization (e.g. 10 seconds, 20 seconds,...60 seconds). This would allow us to suggest a feature representation or to experiment existing ones such as changepoint and last-fired representation (Kasteren et al., 2010). Transforming sensor data to a different feature representation sometimes improves the recognition performance of a model significantly as shown in (Kasteren et al., 2010).

(2) **How much training multi-occupant sensory data is needed to accurately learn the different models' parameters?** Both proposed HMM-based approaches and baseline ones presented in Chapter III rely on labeled data to learn the model parameters in a supervised context. Because collecting labeled data is expensive especially in a multi-occupant environment, we plan to experiment the required amount of labeled activity data for learning the different models' parameters. This experiment cannot be done without a pre-discretization of sensory data.

(3) **Evaluating our proposed models on real behaviors:** We will investigate real-world multi-occupants activities in a more complex scenario as the data used is rather scripted and does not reflect on the real-world setting. Specifically, we plan to evaluate the proposed models on another data collections like those of ARAS (Alemdar et al., 2013).
The dataset used in our experiments is labeled as the data association variables are given. Thus, the performance of activity recognizer is independent to that of data

associator. This allows comparing the performance of activity recognition models in an objective manner. In contrast, the data association variables are not given by ARAS dataset, hence, we first plan to develop a knowledge-based approach that is an ontological approach in order to identify by whom is trigged each active sensor at time based on context information (i.e. activity performed by each occupant, properties of the activity performed at time, locations of active sensors in the living space, ...). Then, we could apply our proposed models. However, the activity recognition performance of each model would be immediately affected by the performance of the data association recognizer (i.e. when the sensor data is incorrectly associated with the occupant).

IV.3 Open Research Questions

In order to bring the multi-occupant activity recognition systems to a more mature stage, some research avenues require to be further investigated. Next, a list of open research questions is discussed.

IV.3.1 Complex activity recognition in multi-occupant setting

In real world situations, human activities are often carried out in complex way. The exiting research literature dealing multi-occupancy has not fully addressed the problem of cooperative activities in a way to cope with different situations like these:

(1) **Interleaved or concurrent activities performed in parallel by multiple occupants:** Each occupant performs his/her activities in a concurrent or an interleaved manner and at the same time another occupant performs his/her activities in a concurrent or an interleaved manner.

(2) There exist more complex situations in which an occupant switches between an activity and a collaborative activity or perform both in a concurrent manner.

(3) **Ambiguity of interpretation:** The interpretation of similar activities may differ depending on the context, for example an activity turn the water tap can be part of many activities like cooking and drinking and the model should be able to handle these situations which appear in both single and multi-occupant settings.

IV.3.2 Scalability of the activity model

All the studies discussed in this thesis (both Chapter II and our proposed approaches in Chapter III) use datasets which are related to only two occupants and do not investigate the scalability of the models proposed therein. Evaluating such models with more than two occupants is an important aspect for real world situations.

In Section II.5.6 we discussed the scalability of data association models and activity recognition models separately, as researchers working on multi-occupant activity recognition tend to focus on one of the two latter problems. First, we pointed out that dealing with scalability in multi-occupancy setting should not only consider new activities, but also new occupants. Second, the scalability of the models in terms of the number of occupants is the most important issue. Although all studies presented in multi-occupant activity recognition have only considered a two-occupant situation (Chiang et al., 2010) (Cook et al., 2010) (Singla et al., 2010) (Alemdar et al., 2013) (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b) (Lin and Fu, 2007) (Chen and Tong, 2014), some of them (Alemdar et al., 2013) (Gu et al., 2009b) (Chen and Tong, 2014) would be easily scalable to additional occupants than others (Chiang et al., 2010) (Cook et al., 2010) (Singla et al., 2010) (Wang et al., 2009) (Wang et al., 2011).

IV.3.3 Resolution of conflicts

People using the same resources may have different preferences. For instance, one prefers watching TV while the light is on, and another prefers it to be off. This is valid for most of the shared but parallel activities.

In (Davidoff et al., 2006) a fieldwork with 12 families is presented. The study reports on 6 social characteristics of home life including multiple users' conflicts at home. The study points out that an understanding of these characteristics should be more tightly-coupled with what services should ultimately be developed for, and how these services should be implemented. It also concludes that SH systems need to participate in value decisions and in negotiating group goal setting.

In this context, the authors in (Hsu and Wang, 2008) propose a resource management system for a multi-occupant SH. The system relies on the strategy of agent conceding negotiation to manage the SH resources. The system consists of 3 components named as: home ontology, device controller and resource allocator. The home ontology describes the spatial organization of the SH as well as information on the devices equipping the living space. The device controller is responsible for collecting information about the occupants. It applies case-based reasoning to predict the resources an occupant may need. The controller finds matching cases in the case base to determine the potential resource conflicts. The resource allocator relies on Belief Desire Intention (BDI) agents, a communication blackboard and conceding negotiation mechanisms to manage conflicts over resources. To implement the SH system, a BDI agent is assigned to each occupant. The blackboard enables the BDI agents to communicate and facilitates the management of the resource conflicts. In terms of conceding negotiation, each agent is assigned a computed conceding risk and in case there is conflict the one with the lowest risk is chosen as conceder of the resource that looks for another resource. The negotiation cycle continues until a common resource use plan is obtained.

IV.3.4 Pervasive multi-occupant activity datasets

The quantitative comparison of multi-occupant activity recognition methods is not straightforward because studies use different datasets. The lack of standard benchmarks makes difficult to exhaustively and fairly evaluate multi-occupant activity recognition models. The studies presented in (Wang et al., 2009) (Wang et al., 2011) (Gu et al., 2009b) relied on private dataset as can be seen in table 1. Studies described in (Hsu et al., 2010) (Chiang et al., 2010) (Cook et al., 2010) (Singla et al., 2010) (Chen and Tong, 2014) and in (Prossegger and Bouchachia, 2014) (Alemdar et al., 2013) used CASAS "Multiresident ADLs" and ARAS datasets respectively. Selecting the best approach from all these studies is not possible as they do not rely on the same activity data.

As the process of collecting activity data requires financial resources which are not within the reach of all research laboratories, researchers tend to use publicly available datasets. Hence, the motivation for presenting eight datasets publicly available may serve as benchmarks for future research on multi-occupancy (see Section II.3). It must be emphasized that there is a real lack of pervasive multi-occupant activity recognition datasets, in particular, datasets that include activity data of more than two individuals and covering the various types of cooperative activities. This will allow researchers to experiment the scalability of their proposed activity models.

IV.3.5 Online learning and inference for real-time multi-occupancy

In comparison to offline activity recognition, online activity recognition has not been much investigated. Indeed most of the work presented in this thesis is based on offline supervised learning. Few work based on online learning has been, however, presented in (Kasteren et al., 2008) (Bouchachia and Vanaret, 2014). Online learning and online inference are required for some situations to adapt the models incrementally as new data becomes available or to make decision in pseudo-real time respectively.

Often a monitoring system needs to make inference instantly in situation that may render an elderly person at risk, (e.g. forgetting to take medication). In these situations we need to detect these unusual events at time in order to intervene. The latter problem is of more importance in a multi-occupant setting as it may put not only the person who caused the unusual events at risk and vulnerable but also all the occupants at home, (e.g. forgetting the stove on). Online inference is also important in situations in which the SH system would have temporary occupants, such as guests. In such cases, the system needs to recognize the new guests and to distinguish between them and the occupants. Because of the relevance of online learning and online inference in this context of activity monitoring, it is important that more effort should be devoted to it.

Bibliography

Acampora, G., Cook, D.J., Rashidi, P., Vasilakos, A.V., 2013. A Survey on Ambient Intelligence in Healthcare. Proc IEEE Inst Electr Electron Eng 101 (12), 2470–2494. doi:10.1109/JPROC.2013.2262913

Afrin Emi, I., Stankovic, J.A., 2015. SARRIMA: smart ADL recognizer and resident identifier in multi-resident accommodations, in: Proceedings of the Conference on Wireless Health WH ’15. New York, NY, USA. doi:10.1145/2811780.2811916

Alemdar, H., Ertan, H., Incel, O.D., Ersoy, C., 2013. ARAS human activity datasets in multiple homes with multiple residents, Presented at the 2013 7th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), pp. 232–235.

Amiribesheli, M., Benmansour, A., Bouchachia, A., 2015. A review of smart homes in healthcare. J. Ambient Intell. Humaniz. Comput. 6 (4), 495–517. doi:10.1007/s12652-015-0270-2

Anon, 2014. Centers for Disease Control and Prevention [WWW Document]. URL http://www.cdc.gov/ (accessed 10.13.14).

Benmansour, A., Bouchachia, A., Feham, M., 2015. Multioccupant Activity Recognition in Pervasive Smart Home Environments. ACM Comput Surv 48 (3), 34:1–34:36. doi:10.1145/2835372

Bishop, C.M., 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Bloom, D.E., Canning, D., Fink, G., 2010. Implications of population ageing for economic growth. Oxf Rev Econ Policy 26 (4), 583–612.

Bookman, A., Harrington, M., Pass, L., Reisner, E., 2007. Family Caregiver Handbook: finding elder care resources in Massachusetts [WWW Document]. URL http://web.mit.edu/workplacecenter/hndbk/ (accessed 10.15.14).

Bouchachia, A., Vanaret, C., 2014. GT2FC: An Online Growing Interval Type-2 Self-Learning Fuzzy Classifier. IEEE Trans. Fuzzy Syst. 22 (4), 999–1018. doi:10.1109/TFUZZ.2013.2279554

Brand, M., 1997. Coupled hidden Markov models for modeling interacting processes. Technical Report 405.1997

Caudill, M., 1987. Neural Networks Primer, Part I. AI Expert 2, 46–52.

Chan, M., Campo, E., Estève, D., Fourniols, J.-Y., 2009. Smart homes — Current features and future perspectives. Maturitas 64, 90–97. doi:10.1016/j.maturitas.2009.07.014

Chen, R., Tong, Y., 2014. A Two-stage Method for Solving Multi-resident Activity Recognition in Smart Environments. Entropy 16 (4), 2184–2203. doi:10.3390/e16042184

Chiang, Y., Hsu, J.Y.-J., Lu, C.-H., Fu, L.-C., Hsu, J.Y.-J., 2010. Interaction models for multiple-resident activity recognition in a smart home, Presented at the 2010 IEEE/RSJ International

Conference on Intelligent Robots and Systems (IROS), pp. 3753–3758. doi:10.1109/IROS.2010.5650340

Cook, D.J., Crandall, A., Singla, G., Thomas, B., 2010. Detection of Social Interaction in Smart Spaces. Cybern. Syst. 41 (2), 90–104. doi:10.1080/01969720903584183

Cook, D., Krishnan, N., Rashidi, P., 2013. Activity Discovery and Activity Recognition: A New Partnership. IEEE Trans. Cybern. 43 (3), 820–828. doi:10.1109/TSMCB.2012.2216873

Cook, D., Schmitter-Edgecombe, M., Crandall, A., Sanders, C., Thomas, B., 2009. Collecting and disseminating smart home sensor data in the CASAS project. Proc. CHI Workshop Dev. Shar. Home Behav. Datasets Adv. HCI Ubiquitous Comput. Res. 1–7.

Cornwell, E.Y., Waite, L.J., 2009. Social Disconnectedness, Perceived Isolation, and Health among Older Adults. J. Health Soc. Behav. 50, 31–48.

Crandall, A.S., Cook, D.J., 2010. Using a Hidden Markov Model for Resident Identification, Presented at the 2010 Sixth International Conference on Intelligent Environments (IE), pp. 74–79. doi:10.1109/IE.2010.21

Crandall, A.S., Cook, D.J., 2009. Coping with Multiple Residents in a Smart Environment. J Ambient Intell Smart Env. 1, 323–334.

Crandall, A.S., Cook, D.J., 2008a. Attributing events to individuals in multi-inhabitant environments. 4th International Conference on Intelligent Environments, 21-22 July 2008 IET pp. 4B4–4B4. doi:10.1049/cp:20081164

Crandall, A.S., Cook, D.J., 2008b. Resident and Caregiver: Handling Multiple People in a Smart Care Facility, in,In: Proceedings of AAAI Fall Symposium 2008 – AI in Eldercare: New Solutions to Old Problems, pp. 39–47. AAAI Press (2008)

Davidoff, S., Lee, M.K., Zimmerman, J., Dey, A., 2006. Socially-aware requirements for a smart home, in: In Proceedings of the International Symposium on Intelligent Environments. pp. 41–44.

Demiris, G., Rantz, M., Aud, M., Marek, K., Tyrer, H., Skubic, M., Hussam, A., 2004. Older adults' attitudes towards and perceptions of "smart home" technologies: a pilot study. Med. Inform. Internet Med. 29, 87–94. doi:10.1080/14639230410001684387

Ding, D., Cooper, R.A., Pasquina, P.F., Fici-Pasquina, L., 2011. Sensor technology for smart homes. Maturitas 69, 131–136.

Du, Y., Chen, F., Xu, W., 2007. Human Interaction Representation and Recognition Through Motion Decomposition. IEEE Signal Process. Lett. 14, 952–955. doi:10.1109/LSP.2007.908035

Du, Y., Chen, F., Xu, W., Li, Y., 2006. Recognizing Interaction Activities using Dynamic Bayesian Network, Presented at the 18th International Conference on Pattern Recognition, ICPR 2006, pp. 618–621. doi:10.1109/ICPR.2006.977

Eagle, N., 2008. Behavioral Inference across Cultures: Using Telephones as a Cultural Lens. IEEE Intell. Syst. 23, 62–64. doi:10.1109/MIS.2008.58

Er, O., Yumusak, N., Temurtas, F., 2010. Chest Diseases Diagnosis Using Artificial Neural Networks. Expert Syst Appl 37, 7648–7655. doi:10.1016/j.eswa.2010.04.078

Fratiglioni, L., Wang, H.X., Ericsson, K., Maytan, M., Winblad, B., 2000. Influence of social network on occurrence of dementia: a community-based longitudinal study. Lancet 355, 1315–1319. doi:10.1016/S0140-6736(00)02113-9

Gross, J., 2007. A Grass-Roots Effort to Grow Old at Home. N. Y. Times.

Gu, T., Wu, Z., Tao, X., Pung, H.K., Lu, J., 2009a. epSICAR: An Emerging Patterns based approach to sequential, interleaved and Concurrent Activity Recognition, Presented at the IEEE International Conference on Pervasive Computing and Communications, 2009. PerCom 2009, pp. 1–9. doi:10.1109/PERCOM.2009.4912776

Gu, T., Wu, Z., Wang, L., Tao, X., Lu, J., 2009b. Mining Emerging Patterns for recognizing activities of multiple users in pervasive computing, Presented at the Mobile and Ubiquitous Systems: Networking Services, MobiQuitous, 2009. MobiQuitous '09. 6th Annual International, pp. 1–10. doi:10.4108/ICST.MOBIQUITOUS2009.6818

Hoque, E., Stankovic, 2012. AALO: Activity recognition in smart homes using Active Learning in the presence of Overlapped activities, in: 2012 6th International Conference onPervasive Computing Technologies for Healthcare (PervasiveHealth). San Diego, CA, pp. 139 – 146. doi:10.4108/icst.pervasivehealth.2012.248600

Hsu, C.-C., Wang, L.-Z., 2008. A smart home resource management system for multiple inhabitants by agent conceding negotiation, Presented at the IEEE International Conference on Systems, Man and Cybernetics, 2008. SMC 2008, pp. 607–612. doi:10.1109/ICSMC.2008.4811344

Hsu, K.-C., Chiang, Y.-T., Lin, G.-Y., Lu, C.-H., Hsu, J.Y.-J., Fu, L.-C., 2010. Strategies for Inference Mechanism of Conditional Random Fields for Multiple-Resident Activity Recognition in a Smart Home, in: García-Pedrajas, N., Herrera, F., Fyfe, C., Benítez, J.M., Ali, M. (Eds.), Trends in Applied Intelligent Systems, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 417–426.

Hu, D.H., Zhang, X.-X., Yin, J., Zheng, V.W., Yang, Q., 2009. Abnormal Activity Recognition Based on HDP-HMM Models, in: Proceedings of the 21st International Jont Conference on Artifical Intelligence, IJCAI'09. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 1715–1720.

Isoda, Y., Kurakake, S., Nakano, H., 2004. Ubiquitous sensors based human behavior modeling and recognition using a spatio-temporal representation of user states. Presented at the 18th International Conference on Advanced Information Networking and Applications (AINA 2004), pp. 512 – 517. doi:10.1109/AINA.2004.1283961

Kasteren, T.L.M. van, Englebienne, G., Kröse, B.J.A., 2011. Hierarchical Activity Recognition Using Automatically Clustered Actions, in: Keyson, D.V., Maher, M.L., Streitz, N., Cheok, A., Augusto, J.C., Wichert, R., Englebienne, G., Aghajan, H., Kröse, B.J.A. (Eds.), Ambient Intelligence, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 82–91.

Kasteren, T.L.M. van, 2011. Activity recognition for health monitoring elderly using temporal probabilistic models. Amsterdam.

Kasteren, T.L.M. van, Englebienne, G., Kröse, B.J.A., 2010. Human Activity Recognition from Wireless Sensor Network Data: Benchmark and Software, in: Chen, L., Nugent, C.D., Biswas, J., Hoey, J. (Eds.), Activity Recognition in Pervasive Intelligent Environments, Atlantis Ambient and Pervasive Intelligence. Atlantis Press, pp. 165–186.

Kasteren, T.L.M. van, Noulas, A., Englebienne, G., Kröse, B., 2008. Accurate Activity Recognition in a Home Setting, in: Proceedings of the 10th International Conference on Ubiquitous Computing, UbiComp '08. ACM, Seoul, South Korea, pp. 1–9. doi:10.1145/1409635.1409637

Katz, S., 1983. Assessing self-maintenance: Activities of daily living, mobility and instrumental activities of daily living. J. Am. Geriatr. Soc. 721–726.

Khan, J., Wei, J.S., Ringnér, M., Saal, L.H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C.R., Peterson, C., Meltzer, P.S., 2001. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat. Med. 7, 673–679. doi:10.1038/89044

Khan, S.S., Karg, M.E., Hoey, J., Kulic, D., 2012. Towards the Detection of Unusual Temporal Events During Activities Using HMMs, in: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, UbiComp '12. ACM, New York, NY, USA, pp. 1075–1084. doi:10.1145/2370216.2370444

Kim, E., Helal, S., Cook, D., 2010. Human Activity Recognition and Pattern Discovery. IEEE Pervasive Comput. 9, 48–53. doi:10.1109/MPRV.2010.7

Liao, W., Ji, Q., 2009. Learning Bayesian Network Parameters Under Incomplete Data with Domain Knowledge. Pattern Recognit. 3046–3056.

Li, J., Liu, G., Wong, L., 2007. Mining Statistically Important Equivalence Classes and Delta-discriminative Emerging Patterns, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '07. ACM, New York, NY, USA, pp. 430–439. doi:10.1145/1281192.1281240

Lin, Z.-H., Fu, L.-C., 2007. Multi-user Preference Model and Service Provision in a Smart Home Environment, in: IEEE International Conference on Automation Science and Engineering, 2007. CASE 2007. Presented at the IEEE International Conference on Automation Science and Engineering, 2007. CASE 2007, pp. 759–764. doi:10.1109/COASE.2007.4341737

Lisboa, P.J., Taktak, A.F.G., 2006. The use of artificial neural networks in decision support in cancer: a systematic review. Neural Netw. Off. J. Int. Neural Netw. Soc. 19, 408–415. doi:10.1016/j.neunet.2005.10.007

Liu, Y., Nie, L., Han, L., Zhang, L., Rosenblum, D.S., 2015. Action2Activity: Recognizing Complex Activities from Sensor Data, in: Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15. AAAI Press, Buenos Aires, Argentina, pp. 1617–1623.

Lu, C.-H., Lin, Y.-C., Fu, L.-C., 2008. Hide and Not Easy to Seek: A Hybrid Weaving Strategy for Context-Aware Service Provision in a Smart Home, Presented at the IEEE Asia-Pacific Services Computing Conference, 2008. APSCC '08, pp. 595–600. doi:10.1109/APSCC.2008.77

Maimon, O., Rokach, L., 2005. Data Mining and Knowledge Discovery Handbook. Springer-Verlag, New York, USA.

McCowan, I., Gatica-Perez, D., Bengio, S., Lathoud, G., Barnard, M., Zhang, D., 2005. Automatic analysis of multimodal group actions in meetings. IEEE Trans. Pattern Anal. Mach. Intell. 27, 305–317. doi:10.1109/TPAMI.2005.49

Meditskos, G., Kontopoulos, E., Kompatsiaris, I., 2015. ReDef: Context-aware Recognition of Interleaved Activities using OWL 2 and Defeasible Reasoning., in: Kyzirakos, K., Henson, C.A., Perry, M., Varanka, D., Grütter, R., Calbimonte, J.-P., Celino, I., Valle, E.D., Dell'Aglio, D., Krötzsch, M., Schlobach, S. (Eds.), SSN-TC/OrdRing@ISWC, CEUR Workshop Proceedings. CEUR-WS.org, pp. 31–42.

Miskelly, F.G., 2001. Assistive technology in elderly care. Age Ageing 30, 455–458. doi:10.1093/ageing/30.6.455

Murata, N., Yoshizawa, S., Amari, S., 1994. Network information criterion-determining the number of hidden units for an artificial neural network model. IEEE Trans. Neural Netw. 5, 865–872. doi:10.1109/72.329683

Nait Aicha, A., Englebienne, G., Kröse, B., 2013. How Lonely is Your Grandma?: Detecting the Visits to Assisted Living Elderly from Wireless Sensor Network Data, in: Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, UbiComp '13 Adjunct. ACM, New York, NY, USA, pp. 1285–1294. doi:10.1145/2494091.2497283

Natarajan, P., Nevatia, R., 2007. Coupled Hidden Semi Markov Models for Activity Recognition, in: Proceedings of the IEEE Workshop on Motion and Video Computing, WMVC '07. IEEE Computer Society, Washington, DC, USA, p. 10–. doi:10.1109/WMVC.2007.12

Nefian, A.V., Liang, L., Pi, X., Xiaoxiang, L., Mao, C., Murphy, K., 2002. A coupled HMM for audio-visual speech recognition, Presented at the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. II–2013–II–2016. doi:10.1109/ICASSP.2002.5745027

Nguyen, N., Venkatesh, S., Bui, H., 2006. Recognising behaviours of multiple people with hierarchical probabilistic model and statistical data association. BMVC 2006 Proc. 17th Br. Mach. Vis. Conf. 1239–1248.

Ni, Q., García Hernando, A.B., de la Cruz, I.P., 2015. The Elderly's Independent Living in Smart Homes: A Characterization of Activities and Sensing Infrastructure Survey to Facilitate Services Development. Sensors 15, 11312–11362. doi:10.3390/s150511312

Olguin, D.O., Waber, B.N., Kim, T., Mohan, A., Ara, K., Pentland, A., 2009. Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior. IEEE Trans. Syst. Man Cybern. Part B Cybern. 39, 43–55. doi:10.1109/TSMCB.2008.2006638

Orwat, C., Graefe, A., Faulwasser, T., 2008. Towards pervasive computing in health care – A literature review. BMC Med. Inform. Decis. Mak. 8, 26. doi:10.1186/1472-6947-8-26

Prossegger, M., Bouchachia, A., 2014. Multi-resident Activity Recognition Using Incremental Decision Trees, in: Bouchachia, A. (Ed.), Adaptive and Intelligent Systems, Lecture Notes in Computer Science. Springer International Publishing, pp. 182–191.

Pruteanu-Malinici, I., Carin, L., 2008. Infinite Hidden Markov Models for Unusual-Event Detection in Video. IEEE Trans. Image Process. 17, 811–822. doi:10.1109/TIP.2008.919359

Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition, in: PROCEEDINGS OF THE IEEE. pp. 257–286.

Ravi, N., Dandekar, N., Mysore, P., Littman, M.L., 2005. Activity recognition from accelerometer data. Assoc. Adv. Artif. Intell. 5, 1541–1546.

Reisberg, B., Finkel, S., Overall, J., Schmidt-Gollas, N., Kanowski, S., Lehfeld, H., Hulla, F., Sclan, S.G., Wilms, H.U., Heininger, K., Hindmarch, I., Stemmler, M., Poon, L., Kluger, A., Cooler, C., Bergener, M., Hugonot-Diener, L., Robert, P.H., Antipolis, S., Erzigkeit, H., 2001. The Alzheimer's disease activities of daily living international scale (ADL-IS). Int. Psychogeriatr. IPA 13, 163–181.

Riboni, D., Pareschi, L., Radaelli, L., Bettini, C., 2011. Is ontology-based activity recognition really effective?, Presented at the 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 427–431. doi:10.1109/PERCOMW.2011.5766927

Roley, S.S., DeLany, J.V., Barrows, C.J., Brownrigg, S., Honaker, D., Sava, D.I., Talley, V., Voelkerding, K., Amini, D.A., Smith, E., Toto, P., King, S., Lieberman, D., Baum, M.C., Cohen, E.S., Cleveland, P.A.M., Youngstrom, M.J., American Occupational Therapy Association Commission on Practice, 2008. Occupational therapy practice framework: domain & practice, 2nd edition. Am. J. Occup. Ther. Off. Publ. Am. Occup. Ther. Assoc. 62, 625–683.

Roy, N., Archan Misra, Cook, D.J., 2016. Ambient and smartphone sensor assisted ADL recognition in multi-inhabitant smart environments. J. Ambient Intell. Humaniz. Comput. 7, 1–19.

Roy, N., Misra, A., Cook, D.J., 2013. Infrastructure-Assisted Smartphone-based ADL Recognition in Multi-Inhabitant Smart Environments, in: 2013 IEEE International Conference on Pervasive Computing and Communications (PerCom). San Diego, CA. doi:10.1109/PerCom.2013.6526712

Sadri, F., 2011. Ambient Intelligence: A Survey. ACM Comput Surv 43, 36:1–36:66. doi:10.1145/1978802.1978815

Sanghai, S., Domingos, P., Weld, D., 2005. Relational dynamic Bayesian networks. J. Artif. Intell. Res. 24, 759–797.

Sarkar, A.M.J., Lee, Y.-K., Lee, S., 2010. ARHMAM: An Activity Recognition System Based on Hidden Markov Minded Activity Model, in: Proceedings of the 4th International Conference on Uniquitous Information Management and Communication, ICUIMC '10. ACM, New York, NY, USA, pp. 71:1–71:9. doi:10.1145/2108616.2108702

Singla, G., Cook, D.J., Schmitter-Edgecombe, M., 2010. Recognizing independent and joint activities among multiple residents in smart environments. J. Ambient Intell. Humaniz. Comput. 1, 57–63. doi:10.1007/s12652-009-0007-1

Smith, E.R., Mackie, D.M., 1999. Social Psychology. Routledge, London.

Sutton, C., McCallum, A., 2006. An introduction to conditional random fields for relational learning, in: Introduction to Statistical Relational Learning. MIT press, pp. 93–128.

Sutton, C., McCallum, A., Rohanimanesh, K., 2007. Dynamic Conditional Random Fields: Factorized Probabilistic Models for Labeling and Segmenting Sequence Data. J Mach Learn Res 8, 693–723.

T. Choudhury, A.P., 2004. Characterizing Social Networks using the Sociometer.

Tunca, C., Alemdar, H., Ertan, H., Durmaz Lucel, O., Ersoy, C., 2014. Multimodal Wireless Sensor Network-Based Ambient Assisted Living in Real Homes with Multiple Residents. Sensors 14, 9692–9719. doi:10.3390/s140609692

Van Laerhoven, K., Kern, N., Gellersen, H.-W., Schiele, B., 2003. Towards a wearable inertial sensor network, in: IEE Eurowearable, 2003. Presented at the IEE Eurowearable, 2003, pp. 125–130. doi:10.1049/ic:20030158

Wang, L., Gu, T., Tao, X., Chen, H., Lu, J., 2011. Recognizing multi-user activities using wearable sensors in a smart home. Pervasive Mob. Comput., Knowledge-Driven Activity Recognition in Intelligent Environments 7, 287–298. doi:10.1016/j.pmcj.2010.11.008

Wang, L., Gu, T., Tao, X., Lu, J., 2009. Sensor-Based Human Activity Recognition in a Multi-user Scenario, in: Tscheligi, M., Ruyter, B. de, Markopoulus, P., Wichert, R., Mirlacher, T., Meschterjakov, A., Reitberger, W. (Eds.), Ambient Intelligence, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 78–87.

Wilson, D.H., Atkeson, C., 2005. Simultaneous Tracking and Activity Recognition (STAR) Using Many Anonymous, Binary Sensors, in: Gellersen, H.-W., Want, R., Schmidt, A. (Eds.), Pervasive Computing, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 62–79.

Ye, J., Dobson, S., McKeever, S., 2012. Review: Situation Identification Techniques in Pervasive Computing: A Review. Pervasive Mob Comput 8, 36–66. doi:10.1016/j.pmcj.2011.01.004

Appendix A: Viterbi algorithm for HMM

We follow the notation specified in Chapter III, the following algorithm represents the viterbi decoder for the hidden states (i.e. activities) of one occupant at home that is R=1.

Let $\delta_t(i)$ be the maximal probability of state sequences of the length t that end in state i and produce the t first observations for the given model.

$$\delta_t(i) = \max\{P\big((a_1^1),(a_2^1),\ldots,(a_{t-1}^1);\ (o_1^1),(o_2^1),\ldots,(o_t^1)\big|(a_t^1)=i\big)\}$$

The matrix ψ is used to retrieve the optimal hidden states at the backtracking step.

- Initialization

$$\delta_1(i) = \pi^1(i)B_1^1(i),\ i=1,\ldots,N$$

$$\psi_1(i) = 0$$

- Recursion

$$\delta_t(j) = max_i\{\delta_{t-1}(i)Q_{ij}^1\}\,B_t^1(j),\ \text{j=1,...,N}$$

$$\psi_t(j) = argmax_i\{\delta_{t-1}(i)Q_{ij}^1\}$$

- Termination

$$P^* = max_i\{\delta_T(i)\}$$

$$\{a_T^1\} = argmax_i\{\delta_T(i)\}$$

- Path backtracking

$$\{a_t^1\} = \psi_{t+1}(a_{t+1}^1),\ \text{t=T-1, T-2,...,1}$$

where P^* is the maximum likelihood of $\delta_T(i)$ at time T and a_T^1 is the most probable activity at time T.

This Viterbi algorithm results in a computational complexity of $O(T\ N^2)$ where T is the total number of events in the dataset. Note that N is the number of activities of the occupant (i.e. R=1).

Appendix B: Viterbi algorithm for CHMM

Let $\delta_t(i)$ be the maximal probability of state sequences of the length t that end in state i and produce the t first observations for the given model.

$$\delta_t(i) = max\,\{P\big((a_1^1, a_1^2), (a_2^1, a_2^2), \ldots, (a_{t-1}^1, a_{t-1}^2);\ (o_1^1, o_1^2), (o_2^1, o_2^2), \ldots, (o_t^1, o_t^2)\big|(a_t^1, a_t^2) = i, j\big)\}$$

The matrix ψ is used to retrieve the optimal hidden states pairs at the backtracking step.

- Initialization

$$\delta_1(i,j) = \pi^1(i)\pi^2(j)B_1^1(i)B_1^2(j),\ i=1,\ldots,N\ j=1,\ldots,M$$

$$\psi_1(i,j) = 0$$

- Recursion

$$\delta_t(i,j) = max_{k,l}\{\delta_{t-1}(k,l)Q_{i|k,l}^1 Q_{j|k,l}^2\}\,B_t^1(i)B_t^2(j),\ k=1,\ldots,N\ l=1,\ldots,M$$

$$\psi_t(i,j) = argmax_{k,l}\{\delta_{t-1}(k,l)Q_{i|k,l}^1 Q_{j|k,l}^2\}$$

- Termination

$$P^* = max_{i,j}\{\delta_T(i,j)\}$$
$$\{a_T^1, a_T^2\} = argmax_{i,j}\{\delta_T(i,j)\}$$

- Path backtracking

$$\{a_t^1, a_t^2\} = \psi_{t+1}(a_{t+1}^1, a_{t+1}^2),\ \text{t=T-1, T-2,...,1}$$

where P^* is the maximum likelihood of $\delta_T(i)$ at time T and a_T^1, a_T^2 are the most probable activities corresponding to Occupant 1 and Occupant 2 at time T.

This algorithm was extracted from (Nefian et al., 2002). It results in a computational complexity of $O(TN^2M^2)$, where T is the total number of events in the dataset. N and M represent the number of hidden states (i.e. activities) for Occupant 1, Occupant 2 respectively.

Abstract

In Smart Home environments, automatic health monitoring of elderly persons allows to assess their cognitive and physical wellbeing though Human Activity Recognition (HAR) that is the recognition of their Activities of Daily Living. Most of the research has been devoted to HAR of single occupants in the environment. However, living environments are usually inhabited by more than one person. We focus in this thesis on the problem of modeling multi-occupant activities. In particular parallel activities and cooperative activities are considered. To deal with multi-occupant activities we investigate different approaches based on Hidden Markov Models (HMMs). Specifically, we propose an HMM-based method, called HMM based Combined Label (CL-HMM), where activities labels as well as observation labels of different occupants are combined to generate the corresponding sequence of activities as well as the corresponding sequence of observations on which a conventional HMM is applied. We also propose a Linked HMM (LHMM) in which activities of all occupants are linked at each time step. We compare these two models to baseline models which are Coupled HMM (CHMM) and Parallel HMM (PHMM). The experimental results show that the proposed models outperform CHMM and PHMM when tested on parallel and cooperative activities.

Keywords: Human Activity Recognition, smart homes, activities of daily living, multiple occupants, pervasive sensing, probabilistic model, hidden markov model.

Résumé

Dans le contexte des maisons intelligentes, la surveillance automatique de la santé des personnes âgés permet d'évaluer leurs bien-être cognitif et physique à travers la Reconnaissance des Activités Humaines (RAH) à savoir la reconnaissance de leurs activités de la vie quotidienne. La plupart des recherches ont été consacrés à la RAH de personnes vivant seules dans l'environnement. Cependant, les environnements de vie sont généralement habités par plus d'une seule personne. Nous nous concentrons dans cette thèse sur le problème de la modélisation des activités de multiple occupants. En particulier les activités parallèles et les activités coopératives sont considérées.

Afin de traiter les activités de multiple occupants, nous étudions différentes approches basées sur les Modèles de Markov Cachés (MMCs). Plus précisément, nous proposons une méthode basée sur les MMCs, appelé MCC basé Etiquettes Combinées (MMC-EC), où les étiquettes des activités ainsi que les étiquettes des observations des différents occupants sont combinées afin de générer la séquence correspondante des activités ainsi que la séquence correspondante d'observations sur les quelles un MMC classique est appliqué. Nous proposons également le MMC Lié appelé MMCL dans lequel les activités de tous les occupants sont liées à tout instant. Nous comparons ces deux modèles aux modèles de références qui sont les MMC Couplés (MMCC) et MMC Parallèles (MMCP). Les résultats expérimentaux montrent que les modèles proposés surpassent le MMCC et le MMCP lorsqu'ils sont testés sur les activités parallèles et les activités coopératives.

Mots-clés: Reconnaissance des activités humaines, maison intelligente, activités de la vie quotidienne, occupation multiple, détection envahissante, modèle probabiliste, modèle de markov caché.

ملخّص

في البيئات المنزليّة الذكيّة المراقبة الأوتوماتكيّة لصحّة كبار السّن تكتسب اهتماما متزايدا. تسمح هذه الأنظمة لتقييم العافية المعرفية و الجسدية من خلال الاعتراف على الأنشطة الإنسانية (أأأ) و هو الاعتراف على أنشطة معيشتهم اليوميّة. معظم الأبحاث تخصّصت في أأأ لمقيمون مفردون في بيئتهم غير أنّ، بيئات العيش عادة يسكنها أكثر من شخص واحد. نركّز في هذه الأطروحة على مشكلة نمنجة أنشطة مقيمون متعدّدون. بشكل خاص نعتبر الأنشطة المتوازية و الأنشطة التعاونية. للتعامل مع أنشطة مقيمون متعدّدون نستخدم مقاربات مختلفة على أساس نمادج ماركوف المخفية (ن م م). بشكل محدّد، نقترح طريقة قائمة على ن م م، تدعى ن م م على أساس بطاقات مدمجة أين يتم الدّمج بين بطاقات الأنشطة بالإضافة إلى بطاقات الملاحظات من سكان مختلفون لتوليد السّلسلة المقابلة من الأنشطة إلى جانب السلسة المقابلة من الملاحظات و التي يتم تطبيق عليهن ن م م التقليدي. نقدم أيضا ن م م مرتبط، أين أنشطة جميع المقيمون مرتبطة في كلّ خطوة وقت. يتم مقارنة هذين النموذجين إلى نموذجين المرجعية هم ن م م المترابطة (ن م م م) و ن م م المتوازية (ن م م ز). تظهر النتائج التجريبية أنّ النماذج المقترحة تتفوّق على "ن م م م" و "ن م م ز" عند اختبارهم على الأنشطة الموازيّة و التعاونية.

الكلمات المفتاحية: الاعتراف على الأنشطة الإنسانية، المنزل الذّكي، الأنشطة المعيشيّة اليوميّة، مقيمون متعدّدون، الاستشعار المنتشر، نموذج احتمالي، نموذج ماركوف المخفي.