Machine Learning Methods for Mining Mobility Data

Negar Ghourchian

Doctor of Philosophy

Electrical and Computer Engineering Department

McGill University

Montreal,Quebec

2016-12-20

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

DEDICATION

This thesis is dedicated to my parents for their endless love and support.

To my best friend, Amin.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Professor Doina Precup, for her continuous support during my study. Doina's perpetual energy, stimulating suggestions and enthusiasm in research had motivated all her students, including me, and without her valuable knowledge and mentorship this thesis would not have been accomplished. She is one of the smartest and strongest people I know and she will always remain my role model and mentor.

I would also like to extend my deepest gratitude to Professor Mark Coates and Professor James Clark for their insightful comments, expertise and encouragement throughout my research.

I thank members of the Reasoning and Learning Lab for making my time at McGill so enjoyable. I consider myself fortunate to be a part of such a creative and smart group of people. The internship at Tandemlaunch was a unique opportunity to engage in solving real-word problems. My sincerest thanks go to my friends at Aerial team, Michel Allegue, Xi Chen, Francois Morel and Mathieu Rozon, who helped me with my data collection and research at Tandemlaunch, and made it so much fun.

I also thank Fonds Qubcois de la Recherche sur la Nature et les Technologies (FQRNT), Natural Sciences and Engineering Research Council of Canada (NSERC) and MITACS for their support of this work.

I have an incredibly supportive family, and I thank my parents, Sorour and Nader, and my brother Nima for their endless love. Last but not least, I would like to express my gratitude to my beloved husband, Amin, for always believing in me.

ABSTRACT

The growing popularity of mobile smart devices and advances in wireless technologies, coupled with state of the art machine learning techniques have created a new era in context aware computing. Context and location aware systems provide the opportunity of the collection of large amount of rich information about human behavior and activities. Therefore, automatic activity recognition, whose goal is to infer semantic patterns and routines from data gathered by such systems has become a great interest among researchers in the field. This thesis applies machine learning techniques to infer meaningful patterns about human activities and mobility from sensory data. We present a sensor selection strategy for activity recognition on smart mobile devices, which requires energy-efficient learning techniques due to the limited computational and energy resources of such devices. We propose an online approach that actively selects a smaller subset of sensors that are the most informative, yet energy-effective, for each time frame. The empirical results confirm that the proposed online method provides good power efficiency, while maintaining accuracy. Secondly, we introduce a novel methodology for analyzing human location and mobility data. Our approach aims to provide an understanding of human trajectories by modeling Places of Interests for individuals, and constructing behavioral signatures for different groups of users based on their interests and specific mobility patterns. We propose an unsupervised learning framework for clustering and labeling mobile data trajectories based on Hierarchical Dirichlet Processes. This hierarchical clustering model adapts the number of clusters identified in a dataset to the complexity of the

data. We evaluate our method on three different real data sets including both fine-grained indoor trajectories and coarse-grained outdoor mobility traces. The results demonstrate that the proposed model is capable of learning the underlying structure of human mobility behavior, even in the presence of noisy and complex data. Third, we present a device-free activity recognition system in the context of smart spaces, consisting of three main building blocks: entrance detection, user identification and localization. These recognition modules leverage a very recent sensing technology based one wifi network coverage from off-the-shelf wireless devices, in order to monitor the behavior and movements of users within an indoor space. The experimental results of the proposed wifi-based system demonstrate that device-free activity recognition is a promising line of research both for academia and industry.

ABRÉGÉ

La popularit toujours croissante des appareils mobiles et les avancements des technologies sans-fils combins aux nouveauts des techniques en apprentissage automatique ont cr une nouvelle re dans le domaine de linformatique sensible au contexte. Les systmes informatiques de type contextuel et de localisation accumulent une grande quantit dinformation riche en contenu sur les comportements et activits humaines. En consquence, la reconnaissance automatique des activits, qui a pour but dinfrer les modles smantiques et routines partir des donnes accumules par de tels systmes est devenu un sujet de grand intrt pour les chercheurs de ce domaine. Cette thse dmontre des techniques dapprentissage automatique servant dduire des modles significatifs sur les activits humaines et la mobilit partir de donnes de senseurs. Nous prsentons une stratgie de slection de senseur pour fin de reconnaissance dactivits sur des appareils mobiles qui requirent des techniques dapprentissage efficaces au niveau de lnergie et de puissance de traitement pouvant oprer avec les ressources limites de ces appareils. Nous avons propos une approche en ligne qui choisit de faon active un petit sous-ensemble de senseurs qui sont les plus informatifs et efficaces en nergie pour chaque plage de temps. Les rsultats empiriques confirment que la mthode propose dmontre une bonne efficacit nergtique tout en maintenant la prcision recherche. Nous avons introduit une nouvelle mthodologie qui offre un haut niveau de comprhension des donnes de localisation et de mobilit humaine base sur la modlisation des lieux d'intrt des individus et qui construit des signatures comportementales pour diffrents groupes dusagers en fonction de leurs intrts et leurs modles spcifiques de dplacements. Nous proposons un cadre dapprentissage sans supervision pour regrouper et tiqueter les trajectoires de donnes mobiles bas sur les processus hirarchiques de Dirichlet. Ce modle de regroupement hirarchique adapte le nombre de regroupements identifis dans un ensemble de donnes la complexit des donns. Nous avons valu notre mthode en utilisant trois diffrents ensembles de donnes qui comprennent des trajectoires dtailles intrieures ainsi que des traces de mobilit extrieures moins dfinies. Les rsultats soutiennent lide que le modle propos est capable dapprendre la structure sous-jacente des comportements de mobilit humaine et ce mme en prsence de donnes complexes et bruites. Nous avons prsent un systme de reconnaissance dactivit sans appareil dans un contexte despaces restreints qui comprend trois composantes principales : la dtection dentre, lidentification des usagers et la localisation. Ces modules de reconnaissance sappuient sur une technologie de dtection trs rcente base sur la couverture de rseaux WiFi gnre par des appareils sans-fils commerciaux et qui permet de reconnaitre les comportements et mouvements dusagers dans des espaces intrieurs. Les rsultats exprimentaux du systme bas WiFi dmontrent que la solution de reconnaissance sans appareil est un sujet de recherche prometteur tant pour le monde acadmique que pour une application en industrie.

TABLE OF CONTENTS

DEDICATION . . . ii

ACKNOWLEDGEMENTS . . . iii

ABSTRACT . . . iv

ABRÉGÉ . . . vi

LIST OF TABLES . . . xi

LIST OF FIGURES . . . xii

1 Introduction . . . 2

1.1 Motivation . . . 2
1.2 Contribution . . . 4
1.2.1 Sensor Selection for Efficient Activity Recognition . . . 5
1.2.2 Location-based Activity Recognition . . . 6
1.2.3 Wifi-based Activity Recognition . . . 7
1.3 Overview . . . 8

2 Background . . . 10

2.1 Human Data Analysis . . . 11
2.1.1 Common Issues in Human Data Analysis . . . 12
2.1.2 Sensing Technologies . . . 19
2.2 Taxonomy of Machine Learning . . . 24
2.2.1 Supervised Learning . . . 26
2.2.2 Unsupervised Learning . . . 28
2.2.3 Semi-supervised Learning . . . 29
2.3 Learning Human Activities . . . 31
2.3.1 Activity Recognition Challenges . . . 33
2.3.2 Machine Learning for Human Data Analysis . . . 38
2.3.3 Hierarchical Activity Recognition . . . 44

3 Sensor Selection for Efficient Activity Recognition 47

3.1 Background 49
3.1.1 Related work 49
3.1.2 Random Forest: A classifier with built-in certainty score . . 50
3.2 Active Learning for Sensor Selection 52
3.2.1 Dataset 53
3.2.2 Baseline Classifiers 54
3.2.3 Smart Classifier Selection 57
3.3 Evaluation 60
3.4 Conclusion 63

4 Location-based Activity Recognition 65

4.1 Background 66
4.1.1 Location Data Mining 66
4.1.2 Non-Parametric Bayesian Modeling 69
4.2 Clustering Location Traces with HDP 80
4.2.1 Inferring Places of Interest 83
4.2.2 Regularization 84
4.3 Task Inference from Location Data 85
4.3.1 Reality Commons - Badge Dataset 86
4.3.2 Trajectory Segmentation 89
4.3.3 Evaluation 91
4.3.4 Hierarchical Location Clustering 92
4.3.5 Evaluation 94
4.4 Learning Social Interaction from Location Data 97
4.4.1 Reality Commons - Reality Mining Dataset 97
4.4.2 Dataset Challenges 98
4.4.3 User-specific Modeling and Global Modeling 99
4.4.4 KL-Divergence based Regularization 100
4.4.5 Evaluation 105
4.5 Mobile Communication Data Analysis 111
4.5.1 D4D Challenge Dataset 112
4.5.2 Mining Call Patterns 113
4.5.3 Evaluation 116
4.6 Conclusion 122

5 Wifi-based Activity Recognition . . . 124
5.1 Background . . . 126
5.1.1 Radio-frequency Data Mining . . . 126
5.1.2 Channel State Information . . . 130
5.1.3 Latent Dirichlet Allocation . . . 133
5.2 Intelligent Indoor Spaces . . . 136
5.2.1 Interpreting from CSI Readings . . . 140
5.2.2 Entrance Detection . . . 147
5.2.3 User Identification . . . 150
5.2.4 Localization in Smart Spaces . . . 153
5.3 Evaluation . . . 156
5.3.1 Experimental Setup . . . 156
5.3.2 Entrance Detection Validation . . . 158
5.3.3 Person Identification . . . 160
5.3.4 Localization Identification using LDA . . . 163
5.4 Conclusion . . . 165
6 Conclusion . . . 167
6.1 Contributions . . . 167
6.2 Future Directions . . . 170
6.2.1 Efficient Wearable Computing . . . 170
6.2.2 Location-based Activity Recognition . . . 171
6.2.3 Device-free Activity Recognition . . . 171
References . . . 174

LIST OF TABLES

Table — page

2–1 Common feature extraction approaches and example applications in human activity recognition 17

2–2 Common types of activity classes used for human behaviour analysis. . 34

2–3 Examples of diversity of machine learning algorithms used by human activity recognition systems. 39

3–1 Activity class distribution. 54

3–2 Accuracy of individual baseline classifiers. The highest accuracy in each section is in bold. 55

3–3 Comparison of recognition accuracy. 62

4–1 Comparison of different strategies to infer proximity 111

5–1 Detection delays of different features, in seconds, when a users switch between *standing* and *walking* (averaged over 10 rounds) 146

5–2 Detection delays of different features, in seconds, when a users *enters* an empty sensing area (averaged over 20 rounds) 149

5–3 Accuracy (%) of different learning techniques for location identification from histogram-based features. 164

LIST OF FIGURES

Figure | page

3–1 Accuracy rate and standard error demonstration for each individual activity . . . 58

3–2 Overall structure of the active algorithm . . . 59

3–3 Influence of (a) θ and (b) δ on the accuracy of the activity recognition task. . . . 62

3–4 Algorithm performance on a segment of data. Top: Certainty measure of the classifier in use. Middle: The corresponding true activity labels at each time frame. Bottom: The algorithm's decision of the classifier to engage . . . 63

4–1 Graphical model of a Dirichlet process mixture model (reproduced from [147]) . . . 73

4–2 Left: Graphical model of a Hierarchical Dirichlet process, Right: example of a HDP structure for $J = 3$ (reproduced from [147]) . . . 75

4–3 Graphical representation of a Chinese restaurant process and Chinese restaurant franchise as presented in [147], p.11. Each rectangle represents a restaurant. . . . 77

4–4 Representation of topic modeling approach . . . 81

4–5 Representation of mobility trace clustering using the proposed HDP-based approach . . . 82

4–6 Different branches in the organization have their own specific space. Each participating employee was assigned a unique ID and base stations with unique IDs were placed at fixed locations for RRSI records [50]. . . . 87

4–7 Example of complex trajectories. The green and red trajectories are the mobility traces of two different employees performing "configuration" and "pricing" tasks, respectively. 89

4–8 Trajectory quantization demonstration. 90

4–9 Average quantization errors for different number of subregions. 91

4–10 Log-likelihood over training data, as the HDP algorithm processes. . . 94

4–11 Cluster visualization from different iterations of HDP algorithm. The probability distribution over each subregion is represented by a filled circle on its centroid, various colors denote different clusters and the size corresponds to the probability. 95

4–12 Schematic view of spatial features extraction from mobility traces, with every user represented by a different color. 98

4–13 Architecture of the proximity prediction algorithm. The left and right side demonstrate user-specific modeling and global modeling, respectively 101

4–14 Performance comparison between HDP and KLD-HDP on synthetic data from plain mixture of Poisson model 104

4–15 Average log-likelihood over training and test data as a function of the number of HDP iterations, in a) the global model and b) the User-specific model. 106

4–16 Visualization of the HDP clustering. Each horizontal line with different color represents the probability distribution of user's *interest* over the locations. The radius of each bubble is proportional to the probability assigned to the particular user for that location. . 107

4–17 Bottom: D_{US} over all possible pairwise user model comparisons. Top: Positive proximity was decided based on setting a threshold on the D_{US} score. The best true positive rate, about 78% is achieved when setting the threshold to 3. 108

4–18 The proposed global model with the automatically inferred number of clusters vs. k-means with different numbers of clusters. Averaging the results over 5 runs, k-means exhibits the best performance using 45 clusters. The proposed approach constantly outperforms this baseline in terms of true positive rate. 110

4–19 illustration of (a) pairwise communication between cell towers, (b) a schematic example of individual trajectory data, and (c) an example of communication sub-graph 112

4–20 Hourly and daily distributions of call records 114

4–21 Log-likelihood over training and testing data, as the algorithm progresses. Learning is quick and converges to good solutions 117

4–22 Visualization of antenna clusters (learned based on call numbers) for nighttime, from 12am to 5am, respectively 118

4–23 Visualization of antenna clusters (learned based on call numbers) for daytime, from 6am to 11am, respectively 119

4–24 Visualization of antenna clusters (learned based on call numbers) for daytime, from 12pm to 5pm, respectively 120

4–25 Visualization of antenna clusters (learned based on call numbers) for daytime, from 6pm to 11pm, respectively 121

5–1 Schematic representation of multipath effect 131

5–2 Graphical model of the LDA (reproduced from [24]) 136

5–3 General overview of the proposed device-free smart indoor system . . 137

5–4 An example segment of CSI magnitude of one subcarrier, before and after applying the noise-removal filter. This segment is collected while one person is *walking* in the sensing area. 142

5–5 An example of CSI magnitudes of all subcarriers. In this example, for each packet the CSI values were measured from 6 data streams and 30 subcarriers per stream. These plots are exhibiting CSI readings of all streams in the same sensing area and the same time-interval. 144

5–6 Histogram of CSI magnitudes of particular subcarriers for two different users performing *standing* and *walking* at the same locations. . . . 146

5–7 CSI magnitudes of all subcarriers of one stream and the extracted features using moving window for *standing* and *walking*. 147

5–8 An example of extracted features for a segment of data including the state of empty spaces and an *entrance* event. 148

5–9 CSI magnitudes of *walking* on the same path performed by 3 different users. 151

5–10 CSI magnitudes of *walking* of a user on 2 different locations within the same sensing area. 153

5–11 Histogram of CSI magnitudes of *walking* of a user on 2 different locations within the same sensing area. 155

5–12 Location layouts and device setup: a) apartment (about $650ft^2$), b) office 1(about $200ft^2$), c) lab (about $600ft^2$), d) TV room (about $250ft^2$) e) conference rooms (about $300ft^2$) 156

5–13 Normalized confusion matrices of entrance detection at; Top left: office 1, Top right: conference room, bottom: TV room. 158

5–14 Cross-validation over conference room data for choosing the best number of neighbors for k-NN. 159

5–15 Performance of the user identification system with different group size of participants and types of activities at; Left: conference room, Middle: office 1, Right: lab. In all locations, the classifiers trained on combined activities including *walking*, *standing* and *sitting* outperform classifiers trained only on *walking* data 162

5–16 Clustering accuracy for different vocabulary size. 165

LIST OF ACRONYMS

CFR	Channel Frequency Response
CRF	Conditional Random Field
CRP	Chinese Restaurant Process
CSI	Channel State Information
D4D	Data for Development
DBN	Dynamic Bayesian Networks
DP	Dirichlet Process
DTW	Dynamic Time Warping
DWT	Discrete Wavelet Transform
ERP	Edit Distance with Real Penalty
FMCW	Frequency Modulated Carrier Waves
GPS	Global Positioning Service
HDP	Hierarchical Dirichlet Process
HMM	Hidden Markov Model
ICA	Independent Component Analysis
IoT	Internet of Things
KL	Kullback-Leibler Divergence
LCSS	Longest Common Sub-sequences
LDA	Latent Dirichlet Allocation
MIMO	Multi-input Multi-output
MSP	Mobile Sensing Platform
NPB	Non-Parametric Bayesian Model
OFDM	Orthogonal Frequency Division Modulation
PCA	Principal Component Analysis
POI	Points of Interest
POMDP	Partially Observable Markov Decision Processes
QE	Quantization Error
RSSI	Received Signal Strength indicator
SVM	Support Vector Machine

CHAPTER 1
Introduction

1.1 Motivation

Recent progress in wireless technologies and advanced electronics and sensors has stoked a great interest, both in academic and commercial circles, in ubiquitous information storage and analysis. The growing popularity of smart mobile and wireless devices enables the collection of large volumes of a wide variety of data about human everyday activities, which creates important research opportunities in context and location aware computing. These pervasive devices accurately and continuously sense characteristics of the human environment, movements and behaviours in real-time and have the capability of logging the information and transferring it to data servers. The richness and extent of this information, coupled with state of the art machine learning techniques, can be the basis for context aware activity recognition. Human activity recognition is a complex and challenging problem, whose goal is to discover meaningful patterns in human data. It has a broad variety of applications in areas such as health and fitness monitoring, smart homes and assisted living, social network analysis, surveillance and security, urban planning and mobile advertising. For instance, patients with movement disabilities or elderly people often need to be monitored at home and have a precise exercise routine as part of their daily life. Automatic recognition of their daily physical activities can provide feedbacks on the routines and detect abnormal events [164, 148]. Or, environmental sensors embedded in mobile phones have the potential to collect significant information that can provide traffic suggestions, climate prediction or environmental monitoring.

Over the last decade there has been a significant amount of research attempting to model user behaviour and predict their mobility patterns, often relying on heavy deployment of complex sensing infrastructure (including cameras, accelerometer, proximity sensors, GPS) to continuously collect sensor readings and utilize learning algorithms to identify activities, movements or gestures. Most of the classical studies in activity recognition can be broadly divided into two categories based on their sensing technologies. The first category is vision based systems (e.g. webcam-style cameras, security cameras, depth sensing cameras, infrared cameras, tomographic-based systems), which use computer vision algorithms to infer human activities from recorded image sequences. An interesting example is Microsoft Kinect [22] that allows the user to interact with console games by means of gestures. Although vision based approaches have been suitable for some public place monitoring, surveillance or gesture recognition applications, they raise serious privacy concerns when it comes to constantly monitoring people's personal and professional lives. In addition to being intrusive, video streams and images are very high dimensional signals and their long-term processing and analysis techniques are relatively infeasible, complex and computationally expensive. Another issue of camera based technologies is their sensitivity to illumination variations, occlusion and background changes that make them impractical in certain applications. The second category is sensor-based methods that take advantage of compact sensors, which collect significant information about physical activities and movement patterns of the user [19, 90, 94]. Sensor-based activity recognition itself can be generally classified into two categories in terms of the type of sensor used for data gathering. The first is called as wearable sensor-based computing, where information can either be collected by mobile smart phones that come equipped with a wide range of embedded sensors (e.g. accelerometer, gyroscope, GPS, wifi) or by using other smart wearable devices specifically designed for the collection and storage of human activity monitoring. The global growth of the wearable technology market (e.g. smart bracelets and smart watches) in recent years, illustrates the significance of this trend and the user preference for these devices due to their compact size, low cost, non-invasiveness and power efficiency. The second category of sensor-based activity recognition relies on environmental variable computing that unobtrusively infers human physical status from changes of environmental variables (e.g. proximity, barometric pressure, temperature, humidity, RFID, wifi signals) [99, 75, 28]. Ambient assisted living and remote care applications in the context of smart home systems are remarkable examples of such intelligent technologies where normal and abnormal physical activities and environmental variables are automatically obtained for evaluation of performance and safety [90, 148].

Despite advances in analyzing human behaviour from sensory data, there are still technical challenges that motivate the development of new techniques to improve performance in more realistic scenarios.

1.2 Contribution

The goal of this thesis is to propose and evaluate algorithms for learning human behaviour patterns from data obtained from mobile and wireless sensory systems. In this spirit, different topics in human activity recognition are described and studied, and new, practical machine learning solutions are proposed. In the following, we highlight the novel contributions described in this thesis.

1.2.1 Sensor Selection for Efficient Activity Recognition

Activity recognition using real-time information obtained from embedded modalities in smart devices is especially valuable, owing to the fact that these devices have become ubiquitous, and they are capable of recording a large amount of data. Mobile phone sensing and wearable devices are preferable in a wide range of applications such as health-care, fitness and safety, due to their low cost, compact size, non invasiveness and low power consumption. However, wearable devices impose some restrictions in terms of computational and energy resources, which need to be taken into account by a machine learning algorithm. As opposed to classical approaches that include all available sensory information on the device to obtain high accuracy in activity recognition, we propose to use a real-time learning method, which interactively determines the most effective set of modalities (or sensors) given the task at hand. The proposed approach actively selects a smaller subset of sensors that are the most informative yet cost-effective for each time frame. We evaluated the performance of algorithm on real data collected using the Intel Mobile Sensing Platform (MSP) [42], which contains a number of sensor modalities and 6 different activities. The empirical results show that the proposed online classification method provides good power efficiency without significant loss in prediction accuracy.

1.2.2 Location-based Activity Recognition

The growing popularity of location-sensing mobile devices and the recent progress in wireless technologies and satellite-based navigation systems enables the collection of a wide range of human trajectories and location data. These data have the potential to reveal information about user context and mobility patterns, as well as to provide insight into social behaviour. Different approaches have been put forward to explore spatial characteristics of user behaviour [18], learn from individual location history [76] and infer similarity or diversity among users [51, 53]. Application areas in which this type of information could be very useful, include advertising (e.g. sending ads to a user that lingers in front of a particular store), surveillance and security, health monitoring (e.g. recognizing if an elderly person living alone is having abnormal mobility pattern and needs help), urban planning, and social network analysis.

Most previous works on location-based activity analysis were focused on low-level tasks (such as next location prediction, model-based trajectory analysis and distance-based trajectory similarity detection) where the main goal is to define parametric spatio-temporal models that synthesize or summarize movement patterns. However, these approaches are limited due to the geometrical complexity of human mobility traces (which vary in shape and size) and due to stochasticity in movement patterns. The problem becomes even more challenging when users go to new, unseen places or when people with different mobility patterns share the same location information. Therefore, the goal of this study is to introduce methodology that offers a higher-level understanding of human location data based on modeling "places of interest" of individuals and constructing behavioural signatures for different groups of users. We work under the assumption that the most frequently visited places by each individual are indication of interests or intentions and that people can be grouped based on these interests. More precisely, we develop a novel unsupervised learning framework for clustering and labeling mobile data trajectories based on Hierarchical Dirichlet Processes (HDP) [147]. HDPs are a powerful probabilistic, mixed-member model for the analysis of grouped data, which adapt the number of clusters identified in a dataset to the complexity of the data. The proposed framework was evaluated on three real datasets from different application scenarios. The results support the idea that the proposed model is capable of learning the underlying structure of human mobility behaviour, even in the presence of noisy and complex human trajectories.

1.2.3 Wifi-based Activity Recognition

In recent years, a growing interest for activity identification through device-free approaches has emerged since this does not require people to carry around the sensing infrastructure. The alternative to using mobile sensor-based information for activity recognition is to use radio frequency sensing interfaces, where the key idea is to monitor the influence of human body movements and gestures on the strength and pattern of wireless communications between a transmitter and receiver [153, 130, 105]. There are various technologies for obtaining such data (including wifi, RFID, Zigbee, etc.) with different characteristics and processing steps. For example, authors in [7, 5, 130] have proposed monitoring the Doppler shifts and multipath distortions of wifi signals originated by human physical activities or capturing radio reflections bounced off human body in order to detect and classify different movements and gestures in the environment. However, most of the existing approaches need specific custom hardware, such as transmission radar in order to employ their solutions.

A very recent research area focuses on activity recognition by employing off-the-shelf wifi devices, e.g., access points, laptops, smart TVs. This is mainly motivated by wireless technology improvements and the fact that wifi signals are pervasive in daily life at home, work and even public places. Studies suggest that information gleaned from the physical layer in wireless infrastructures (e.g. wifi signals), such as channel state information (CSI) and received signal strength indicator (RSSI) values have the potential to characterize the environment they pass through, which includes both ambient objects and human movements and gestures. However, the current design and implementation of this novel technology exhibits some limitations due to the complexity of the wireless signal propagation in indoor environments and due to the challenging nature of human behaviour itself.

In this thesis we leverage this new technology to introduce a new activity recognition systems within an intelligent indoor environment, created by wifi network coverage of off-the-shelf wireless devices. In this study, several analytic and modeling procedures are considered withing the concept of device-free smart homes, security surveillance applications including localization, entrance detection and user identification. The proposed design and implementation have been evaluated on real data gathered from diverse living environments under multiple realistic scenarios.

1.3 Overview

The thesis is organized as follows:

Chapter 2 provides a general overview of existing machine learning techniques for the specific application of context aware activity recognition and human data analysis. Also, we briefly introduce different types of human data and present existing techniques for analyzing each data type.

Chapter 3 studies challenges and restrictions of wearable sensor-based activity recognition and then presents a novel sensor selection strategy for efficient activity recognition.

Chapter 4 contains the presentation of location-aware activity recognition problem. First, we provide an overview of existing location-aware approaches, and present the necessary background material for our solution. Then, we describe our proposed algorithm for high-level mobility data analysis.

Chapter 5 includes a review of radio-frequency based activity recognition and describes our approach to creating an intelligent ambient that learns human activities from wireless signal transmissions and distortions.

Chapter 6 summarizes the main findings and results, provides conclusions and discusses some future suggestions for extensions of this work.

CHAPTER 2
Background

Monitoring human daily activities and movements plays an important role in context aware and ubiquitous computing in both academia and industry, where it combines disciplines such as computer science, statistics, sociology and human-computer interaction. The goal of activity recognition is to interpret activities and gestures of users from data collected through a sensing infrastructure in order to assist the users with their tasks. User activity recognition is a well studied and challenging research area because of the diversity and complexity of human behaviour [29]. Researchers have investigated different data analysis approaches depending on the underlying sensing technology that is employed for collecting the data and the environment in which the activity is performed.

The first works on activity recognition took place within the computer vision community and leveraged video cameras as sensing devices [35]. In fact, there have been extensive studies of vision based activity recognition systems which exploit the rich information contained in video. However, in some specific applications such as surveillance and security vision based activity recognition suffers from privacy issues and its computational cost can also be prohibitive.

Therefore, in this thesis we focus our attention on challenges and research problems related to sensor-based activity recognition systems. In this chapter, we discuss related work and the necessary background on sensor-based activity recognition.

Section 2.1 introduces the specific characteristics and challenges of human activity recognition. Section 2.2 provides background on machine learning algorithms used in this area. Section 2.3 presents related work on human activity learning systems. Further background on specific methods is provided in each subsequent chapters as needed.

2.1 Human Data Analysis

The early works on sensor-based activity recognition date back to the 1990s [59, 135, 13] when small body-worn sensors were used for detection of basic physical activities and body postures under controlled conditions. Further advances in miniaturization and computational power motivated steps towards more challenging problems and realistic application scenarios where the human data was collected under naturalistic circumstances. Particularly, the automatic recognition of human activities has become interesting and beneficial to several real-world domains such as smart homes [144] and assisted living [158, 174], medical diagnosis, rehabilitation and physical therapy [86, 32, 85], the entertainment and sport sectors [138, 17, 55] and security [164]. In recent years, activity recognition has become a key component in many human-centric industrial applications. For instance, most of the leading smart phone manufacturing companies like ©Samsung and ©Apple release their devices preloaded with integrated health platforms that collect and analyze sensory data to help users track their own activities and hence fitness and wellness. Another example is sport products such as the ©Nike+ running shoes and SportWatch, the ©Speedo wristwatch, *E-textile* (garments made from smart textiles that monitor heart rate), ©ImpactSport's ePulse armband (heart rate monitor) or ©BodyMedia's GoWear unit (equipped with accelerometer, heat flux, galvanic skin response and skin temperature), which provide users with feedback on their performance using compact integrated motion and body vital sensors.

Despite considerable advances in collection and analysis of human data, designing and developing activity recognition systems for real life applications remains a challenging task, due to some unique requirements such as data collection protocol, sensor selection and placement strategies and ground truth annotation. We know outline some of these issues.

2.1.1 Common Issues in Human Data Analysis

2.1.1.1 Data Collection Protocol

Any comprehensive human-centric study should contain a large number of participants with diverse characteristics to ensure flexibility and robustness of the system. However, in many behavioural and activity monitoring studies, the techniques are evaluated on data collected in controlled experimental settings (e.g. laboratories) where participants are asked to perform predefined and staged actions. This setup usually allows dense sensor installation which provides useful data but can not be replicated in real practical scenarios. As a result, the performance of the algorithms significantly decrease when used under naturalistic circumstances, because the controlled environment may artificially restrict, simplify or influence the learning process [19]. A relatively small number of studies have collected data under longer term and in real-life naturalistic (or semi-naturalistic) conditions where the participants perform their normal everyday activities. These out-of-lab natural settings usually result in lower recognition accuracy.

Another aspect of human data collection is the extent to which the user is actively engaged in the sensing process [94], which allows the development of two different types of experimental design: participatory sensing or opportunistic sensing. In the former, the users actively perform the activities and may also provide true labels to facilitate the learning process. In the latter case, the participants are not involved in the data collection activity, which is particularly useful for community sensing where individual sensing could be hard and time-consuming.

2.1.1.2 Sensors Selection and Placement

One of the preliminary steps in activity recognition is to determine the most informative and cost-effective set of sensors that allows optimum separation of activity patterns in the feature space. There are many factors that influence the sensor selection process for activity recognition infrastructures such as price, size, energy consumption, obtrusiveness, ease of installation and integration with existing platforms and type of data it generates. In the context of wearable (on-body) sensors, the challenge of sensor selection is coupled with the number and placements of sensors (especially for accelerometers), since the accuracy of activity recognition is affected by position setup depending on the application. For example, different studies suggest that for body posture and activity analysis from accelerometer data, the best position to attach the sensors are on thigh and hip to help distinguishing walking, sitting, standing and cycling, and on chest and wrist to help distinguishing typing, eating and lying [66, 19, 55]. On the other hand, there are scenarios and applications in which it is required, convenient or beneficial to carry-on activity recognition from context environmental information such as proximity, pressure, and radio-based signals(e.g. wifi and RFID). Although these platforms are more comfortable and unobtrusive, there are many concerns and challenges that need to be considered while dealing with non-wearable sensors. Depending on the application, a broad variety of parameters influence the performance of sensing infrastructure such as ambient noise, inference impact, energy efficiency and range, occlusion and orientation sensitivity. For instance, authors in [28] have used RFID tags to explore the interaction of users with everyday objects and they have performed a wide range of experiments to discover the optimum setup of RFID tags (in terms of spatial density, orientation, proximity to users, etc.) for activity monitoring in a realistic home environment.

2.1.1.3 Preprocessing

Human-related sensor modalities are acquired from multiple sources of sensing infrastructures with different characteristics such as sampling rate, power consumption and required operating system. For example, the sampling rate for accelerometer depending on the application lies between 20 Hz to 100 Hz [95], where as GPS is typically sampled at a relatively slower rate about(5 Hz) [29]. Moreover, raw sensor data may be disturbed by artifacts due to sensor displacement or malfunction and electronic noise. Thus, after collecting raw data from different sensors, the next step is pre-processing that begins with synchronization and noise reduction. This stage employs signal processing techniques such as filtering and smoothing to remove artifacts and transforms the nonsynchronous sequences of raw samples into a set of synchronized time-series.

At the same time, human activities are often performed in time units of seconds or minutes, which is much longer compared to the sensors' sampling rates; thus, activity recognition needs to be carried out during a time window (rather than by samples). Furthermore, humans perform their activities consecutively and even concurrently,and the transitions between actions are smooth rather than being clearly separated by pauses. The segmentation step identifies the boundaries of activities and separates the segments of the data stream that contain information about activities. Segmentation techniques can be divided into two categories: overlapping and non-overlapping approaches, where the size of the segments can be fixed or optimally adapted to the application characteristics. In the literature, the most commonly used methodology for stream segmentation is the time-based sliding window, where a window of fixed size moves over sensor events in order to take into account temporal variations and time dependency for activity learning [19, 81, 90, 92]. Another group of studies, have used contextual sources of information (e.g., additional sensors) to identify segments of different activities from the recorded data. For example, mobile phone usage context (such as text messages and call logs) or location information obtained from GPS traces [14] may be used as an external source of information about the start and duration of activities.

2.1.1.4 Feature Computation

The raw sensory data gathered from sensing devices are usually an ordered sequence of observation values at consecutive time steps. After preprocessing, a feature extraction module transforms the raw observation samples into features (or attributes) that help discrimination between different category of activities. In activity recognition, the feature selection criteria are usually problem-dependent. In general, two class of approaches have been proposed to extract features for activity recognition:

- **Statistical Features:** The classical methods of feature extraction are signal-based statistical analyses due to their simplicity and high performance across different applications. In these approaches, a variety of quantitative characteristics, including wavelet, time- and frequency-domain information, are generated from raw signals that best describe human activities.
- **Knowledge-driven Features:** These approaches intend to exploit semantic relationships and prior knowledge in the domain of interest in order to capture deeper interpretation of humans' behaviour and activities in real life scenarios and applications [39].

Table 2–1 represents the most commonly used features and examples of their applications.

The feature extraction step is often followed by a feature selection process or a dimensionality reduction that prepares the data for the machine learning stage. The goal of the feature selection or dimensionality reduction step is to increase accuracy and reduce computational cost. The more features are involved in the learning process, the more computational effort, data and memory are required. Moreover, not all of the extracted features are equally informative and discriminating between activities. Therefore, a variety of feature ranking and feature selection approaches, such as Independent Component Analysis(ICA), Local Discriminant Analysis (LDA)

Table 2–1: Common feature extraction approaches and example applications in human activity recognition

Type	Features/Examples	References
Statistical Features	Time-domain (Mean, standard deviation, variance, entropy, kurtosis)	[141, 19, 125]
	Frequency-domain (Fourier transform, discrete Cosine transform)	[141, 19, 125, 89]
	Time-Frequency domain (Wavelet transforms)	[110, 118]
	Others (Principal Component Analysis, Linear Discriminant Analysis, HAAR filters)	[73, 11, 40]
Knowledge-driven Features	Event-based (neighboring event, situation similarity, activity duration, step detection)	[112, 20, 21]
	Body model(sensor position, cross acceleration)	[15, 21]
	Environmental variables (audio, light, location, date)	[18, 60, 125]

and Principal Component Analysis (PCA), have been explored in the activity recognition literature [16, 12]. For example, PCA is a well known and widely used statistical method that discovers the optimal linear combination of the features and maps data points from a high dimensional space to a lower dimensional space while keeping all the relevant linear structure intact [11, 163]. Some other studies considered using sequential forward/backward feature selection (SFFS/SBFS) algorithms that adds/removes extracted features one at a time such that the performance is maximized [11, 129]. As an alternative, some machine learning techniques such as Lasso or AdaBoost include built-in feature selection mechanisms that automatically select a relevant subset of feature.

2.1.1.5 Ground Truth Annotation

Accurate annotation or labeling is an important step towards activity recognition, since it has a big impact on supervised learning models that are widely employed. Unfortunately, directly observing users can be expensive and inaccurate and scales poorly for the monitoring of large populations. In many cases researches use data annotated with subject self-report labels. In this method, the participants are asked to manually note the activities that they perform, which can potentially increase the amount of collected data and help to reduce the effects of individual variation. However, the reliability of self-report annotation has been the subject of considerable debate, due to its error-prone nature. A least interfering and erroneous alternative would be using cameras or microphones while the activities are carried out. The downside of this method is that it often requires human observers to review the data in order to annotate it with corresponding activity labels. In recent years, researchers have focused on investigation of learning techniques (i.e. unsupervised and semi-supervised algorithms) that reduce the need for labeled data and at the time, introduce methodologies that automatically detects the activity segments from the raw data with minimum user involvement. For example, authors in [150] have used speech recognition techniques to detect the start and end point of activities from predefined set of commands.

2.1.2 Sensing Technologies

Learning human behaviour and activities can take advantages of both low-level sensor data and high-level context information. Advances in pervasive computing along with machine learning algorithms engendered the paradigm of inexpensive and unobtrusive smart environment for gathering functional status of users. Recently, many projects have been developed that resulted in smart infrastructures for monitoring and learning from human behaviour at both individual and organizational scales. Mobile Sensing Platform (MSP) an embedded activity recognition system [42], MavHome and CASAS smart home projects [45, 47] and PlaceLab a radio beacon-based localization system [93] are examples of collaborative efforts between academia and industry for real-life deployment of human activity recognition.

One important aspect of designing a sensing platform for activity recognition is the level of user engagement in the process of human-related data collection, which categorize the existing systems into two types: *participatory sensing* and *opportunistic sensing* [94]. In the former, the participants are voluntary incorporated into the sensing system and consciously interact with the sensing devices. In the latter case, sensing system automatically collects data without the participants' involvement or intervention. Depending on the characteristics of the problem at hand (tolerance of users to endure interruption, sensitivity of the information and other practical consideration), a system decides to what extent participants should be involved in the data collection stage.

Researches in sensing technology has resulted in a wide range of prototypes including, but not limited to: on-body sensing, environmental variables, location, physiological data and Context sensing.

2.1.2.1 On-body Sensing

For many years, wearable sensors, small size mobile devices equipped with miniature internal sensors, have been the most important and broadly used source of information for body posture and human activity analysis. These internal sensors include motion and orientation sensors (i.e., 3D accelerometer, gyroscope, compass and magnetometer), proximity sensors and Bluetooth, and have been widely used as sensing technologies for activity recognition [42, 94, 29, 99, 90]. Classical studies in wearable sensors-based activity recognition used to place a single type of sensor, typically accelerometer, in multiple positions on body to infer the physical movement of users. For example, several studies have reported high recognition accuracy rates (up to %98 [87]) for ambulation activities (e.g., walking, lying, bicycling, climbing stairs, shaking hands) only from acceleration data, under different application scenarios and evaluation methodologies [59, 13, 19, 163, 15].

Over past decade, with advances in microelectronics several research groups developed small wearable sensing platforms that packaged multimodal sensors into a compact device, providing the capability to store raw sensor data, communicate and even locally process the data. For instance, *Xsens* [134], MSP and *SenseWear* [132] provide versatility by detecting different modalities for deployment of many applications of monitoring human body movement and posture for home automation or healthcare. Recently, we are on the new era of context-aware computing where smart mobile devices are a ubiquitous part of modern society, and they are capable of recording a wide range of sensing modalities. These mobiles incorporate many diverse and powerful sensors and due to their convenient size people carry them close to their body (e.g. in their pockets or) almost everywhere. Hence, there is an increased interest in using off-the-shelf mobile phones for collecting human-related data and studying human movement and behaviour from these realistic source of information [92].

2.1.2.2 Environmental Variables

In many instances intelligent environment can assist monitoring the state of users. A large group of studies have explored the influence of user activities on ambient variables (such as barometric pressure, temperature, audio signals, speech, humidity, ambient light), and then aimed to statistically verify the changes in these measurements with respect to each activity [107, 42, 99, 125, 84]. Particularly, passive environmental variables are proven to reflect the characteristics of user's physical motions (e.g., walking, running, resting) and their associated locations.

2.1.2.3 Location Data

Knowledge of user's location and mobility pattern provide fundamental information for estimating human activity. Even though human mobility and movement models have a high degree of variation, they also exhibit structural patterns due to some restrictions such as temporal, geographic and social constraints. Different approaches have been put forward to derive user location [91, 6, 84, 31], explore spatial characteristics of human behaviour [18, 75, 171], and learn from individual and social location histories [51, 53, 41, 14]. At the application level, location-based activity recognition can potentially benefit many research areas, such as advertisement, health monitoring, social, urban planning and transportation.

The framework of system locations can be briefly divided into tree classes: mobile phone-based location systems (GSM), radio-based location systems (wifi beacons, RADAR, Zigbee, RFID tags), satellite-based location systems (Global Positioning Service (GPS), GLONASS satellite navigation systems). Another aspect of location sensing infrastructure, which plays an important role in the accuracy and performance of a location-based activity recognition system, is the range of coverage through different spaces. For example, GPS receivers are ineffective and unreliable indoors, while infrared proximity badges perform poorly in the presence of direct sunlight and cannot be used outdoors [31]. As an alternative, Intels Universal Location Framework (ULF) [67] fuses readings from an outdoor GPS receiver with indoor WiFi signal-strength triangulation when user moves between indoor and outdoor environments.

2.1.2.4 Physiological Data

Physiological signals were expected to play a larger role in revealing information about human behaviour. A few work considered combining vital signs data (such as heart rate, electrocardiogram ECG, skin temperature, eye-blinking, respiration rate and blood oxygen saturation) with other sensors for monitoring physical activities of users [143, 125, 96, 30]. However, these signals often response to activity changes with a delay which makes them more correlated with the intensity level of the activity instead of the type or duration of the activity [125]. Moreover, biosignals are very sensitive to sensor placement, subject-dependent and noisy measurements, which makes the data analysis more challenging. In spite of that, the physiological signals are proven to accurately reflect the emotional and biological states of users and have been vastly used in affective recognition and health monitoring studies.

2.1.2.5 Radio-based Sensing

There is a recent research interest towards device-free sensing technologies that relax deployment requirement and active user involvement, and reduce the energy consumption for data collection. These infrastructures (such as wifi, Zigbee badges, RFID tags, FM broadcast signals, microwaves) consider the fact that any moving object in a radio field attenuates the radio strength and influence the characteristics of communication channel between transmitter and receiver [153]. Most of the related works in activity recognition from radio-based sensing devices focus on body motion and gesture detection in home automation and assisted living applications (e.g., [5, 105, 7, 130]), where the wireless communications coverage are expected to be more stable and have higher quality.

2.1.2.6 Context sensing

Modern studies of activity recognition have adopted a new range of context information to exploit human behaviour and actions. Depending on the application scenario, researches have used various sources of information from phone context (e.g., call, text, email, web surf, application usage, calendar entries) to social network (e.g., ©Facebook and ©Twitter profiles) and social media communications (e.g., shares, likes, follows, messages) in order to facilitate understanding and monitoring human behaviour [51, 83, 114, 52, 107]. This contextual information along with other sensory data lead to creating user-specified models that accurately discover and identify structure in routine of user behaviours and learn about their interests, attentions and actions.

2.2 Taxonomy of Machine Learning

Similar to any pattern recognition algorithm, a typical activity recognition pipeline comprises a front-end, whereby relevant features are extracted from the data stream, and a back-end, where the task of learning takes place. So far, we had focused on th challenges and properties of the front-end. In this section we will shift the focus more towards the learning mechanism. Machine learning is a large field in artificial intelligent that aims to evolves knowledge and patterns in empirical data to generalize from limited amount of instances. The general role of machine learning in activity recognition is to build a comprehensive model that accurately outputs a label of *activity* given a set of input *observations*. Machine learning and probabilistic modeling techniques that have been used for activity recognition varies akin to the variety of the sensing technology and activity itself. Based on the type of obtainable information, learning techniques can be broadly divided into four main classes: *supervised*, *semi-supervised*, *unsupervised* and *reinforcement* learning. The supervised learning approaches learn from given instances, which means each point in training data is associated with an output class label. Unsupervised learning approaches are considered when labelled data is not available and we aim to discover structure or similarities among observed instances. In the semi-supervised learning approaches the assumption is that we have a small amount of labelled data and a large amount of unlabelled data, where the goal is to improve accuracy of supervised learning tasks using the readily available unlabelled data. In reinforcement learning, the goal of the learner is to learn which action to take based on a rewards/punishment system. Unlike the previous types, reinforcement learning is not commonly used in activity recognition literature.

Learning approaches can also be categorized based on the role of learner. Traditional learning techniques, which take a given batch of data and produce a hypothesis or model are called *passive learning* algorithms. Alternatively, *active learners* are the algorithms that have access to a small amount of labeled data, but can interactively query the source of information to achieve optimal outputs. There are several strategies for determining which data points are most informative and should be labeled, including [137]:

- **Uncertainty sampling**, that selects the data points the model is least confident about.
- **Query by committee**, that involves a committee of models trained on the current labelled data, and selects the data points which they disagree the most.
- **Expected model change**, that selects the data points which would convey the greatest change to the current model.
- **Expected error reduction**, that selects the data point which would most reduce the model's generalization error.

Here, we provide an overview of existing learning approaches for treating human-data in activity recognition context.

2.2.1 Supervised Learning

As mention earlier, when we are fortunate enough to obtain observations with activity labels (training set), we consider supervised learning algorithms to produce a classifier that can assign label to unseen data (test set). The majority of classical works in activity recognition have used supervised learning approaches, including both *discriminative* and *generative* models. Discriminative models (such as Support Vector Machine(SVM) and decision trees) simply provide explicit boundaries between classes, whereas generative models (e.g. Naive Bayes, Hidden Markov Model(HMM)) learns the distribution of each individual class and how the data is actually generated.

Suppose we have a set of N training samples of the form $\{(x_1, y_1), \cdots, (x_N, y_N)\}$, where each x_i denotes the feature vector of the ith sample and y_i denotes its associated output value or label. There are tree type of activity recognition problems that can be addressed using supervised learning framework depending on the type of output domain:

- **Classification:** that is the general case in activity recognition where the training data is used to assign and predict an activity class label, $\boldsymbol{Y}$, from a finite set of nominal variables or categories. In this case, if there are only two choices of activities(e.g. talking vs. not talking) [114], we have binary classification and when there are more categories (e.g., *walking*, *running*, *jogging*), the problem is multi-class classification [131, 92, 141, 171].
- **Regression:** the problem where the learning goal is to predict a real-valued output, $\boldsymbol{Y} = \mathbb{R}$, such as heart rate and body fat percentage [97], instead of class label.

- **Anomaly detection:** where the goal is to identify an *unusual* event, such as fall or intrusion detection [164, 71], given an imbalanced dataset including many samples of normal events and a small minority of abnormal events.

More formally, given a training set $D \subseteq \boldsymbol{X} \times \boldsymbol{Y}$ consists of N samples, a supervised learning algorithm seeks a function $h : \boldsymbol{X} \rightarrow \boldsymbol{Y}$, usually called a *hypothesis*, such that $h(\boldsymbol{X})$ is a good estimation of $\boldsymbol{Y}$ values with small *generalization error*. In supervised machine learning, the generalization error is a measure to evaluate the accuracy of an algorithm in predicting labels for previously unseen data. The hypothesis space $\boldsymbol{H} \subseteq \boldsymbol{Y}^{\boldsymbol{X}}$, is a subest of possible functions out of which the learner selects its hypothesis depending on the properties of the feature space and prior knowledge on the learning task. In order to choose the hypothesis that best fits the training data, a loss function $L(y_i, \hat{y})$ is needed to be defined to measure how much the predicted estimations, $\hat{y}$, differ from the true y values. The risk of function h is then defined as the expected loss of h, and can be estimated from the training data as

$$R(h) = \frac{1}{N} \sum_{i} L(y_i, h(x_i)). \tag{2.1}$$

In supervised learning framework, the goal is to choose the hypothesis that minimize the risk function. There are two basic approaches to choosing h: *empirical risk minimization* and *structural risk minimization*, while both approaches assume that the training set are N drawn i.i.d. pairs according to the distribution D. In empirical risk minimization, the learning algorithm aims to search for the h that minimizes the $R(h)$, so an optimization algorithm can be applied to find the h.

In probabilistic models where h takes the form of a conditional probability model $h(x) = P(y|x)$, the loss function is the negative log-likelihood $L(y, \hat{y}) = -\log P(y|x)$, and the empirical risk minimization is equivalent to maximum likelihood estimation.

When the training set does not include sufficient samples, empirical risk minimization may lead to high variance and poor generalization. Choosing a more complex hypothesis may yield to finding a fit that perfectly predicts the training samples, but it does not generalize well to new unseen data. Structural risk minimization seeks to prevent the learner from memorizing the training samples without generalization, i.e. overfitting, by incorporating penalty into the optimization. The main idea is to change the error function that we intend to minimize into

$$J(h) = R(h) + \lambda C(h) \tag{2.2}$$

where λ is regularization coefficient and controls the bias-variance trade-off, and can be chosen empirically via cross validation. $C(h)$ is the regularization penalty that is employed to penalize complexity of the hypothesis.

2.2.2 Unsupervised Learning

Unsupervised learning is considered where the training set contains only unlabelled data points. Consider a dataset of N input samples of the form $\{x_1, \cdots, x_N\}$, where each x_i denotes the feature vector of the ith sample and $\boldsymbol{X}$ denotes the input space. In this case, there is no corresponding desired output $\boldsymbol{Y}$ provided in the data, and the goal of the machine is to develop a formal framework to estimate a model that represents the probability distribution for new input given previous inputs. Indeed, discovering structures and routines in human behaviour from completely unlabelled data is a complicated problem. Therefore, the goal of unsupervised algorithms is to find underlying patterns or similarities in the data. Different approaches of unsupervised learning based on their application include:

- **Clustering:** Cluster analysis is the general task of grouping a set of observations according to their *similarity*. Many activity recognition application have used different clustering methodology for finding activities such as typical ambulatory activities and context identification from the unlabelled observations [111, 89, 114].
- **Dimensionality reduction:** This is the process of replacing a high-dimensional feature space by its projection onto a smaller space to avoid the effects of curse of dimensionality and improve the performance of the learning model. This process summarizes all data points in the same way regardless of their class. Many activity recognition systems have used dimensionality reduction techniques as a pre-processing step to the main classifiers [12, 11, 163].
- **Quantization:** This has often been used as a pre-process step for discretization of continuous variable spaces.

2.2.3 Semi-supervised Learning

Semi-supervised learning consider the application scenarios where obtaining fully labelled data is hard or expensive, and therefore, the goal is to learn from partially labeled instances. In this case, the small amount of labelled instances are used to train a classifier and create decision boundary between the classes, and later, the decision boundaries are enhanced based on the distribution of the larger amount of unlabelled instances. In fact, semi-supervised learning is halfway between supervised and unsupervised learning [37]. In many human-centric data analysis applications learning with semi-supervised algorithms are preferable since they can take advantages of large quantities of observations without considering the annotation costs. For example, many recent activity recognition studies have benefited from different strategies of semi-supervised learning in their works to avoid the complications of getting fully labelled data [119, 139].

As in the supervised learning framework, we are given a set of N training samples with the form of $\{(x_1, y_1), \cdots, (x_M, y_M)\}$, where each x_i denotes the feature vector of the ith sample and y_i denotes its associated output value. Additionally, we are given M samples $\{x_{N+1}, \cdots, x_{N+M}\}$ without corresponding labels or outputs $\boldsymbol{Y}$, with typically $N \ll M$. Semi-supervised learning algorithm attempts to leverage combined information to improve the performance of individual classification (on only labelled part of the data) or clustering (on all of the data while discarding the labels) models.

In general, most of the semi-supervised algorithms used in activity recognition field can be organized into two classes:

- **Generative models:** These commonly used models usually assume some additional information is available on the probability distribution of the observations and try to estimate the parameters of this probability density function. These approaches seek to estimate $p(x|y)$ the distribution of data points belonging to each class, and assume that the distribution takes a particular form of $p(x|y, \theta)$ parameterized by the vector θ. The parameterized distribution can be written as $p(x, y|\theta) = p(y|\theta)p(x|y, \theta)$, and each parameter vector θ corresponds to a predictor function $f_\theta(x) = \underset{y}{\operatorname{argmax}}\ p(y|x, \theta)$. To identify the f_θ that both fits the labelled data and unlabelled data, the log likelihood of θ is maximized, with λ as a balancing weight,

$$\underset{\Theta}{\operatorname{argmax}} \left(\log p(\{x_i, y_i\}_{i=1}^{l}|\theta) + \lambda \log p(\{x_i\}_{i=l+1}^{l+u}|\theta)\right) \tag{2.3}$$

Examples of this models applied for activity recognition problem include Gaussian mixture models for activity recognition in healthcare monitoring systems [124] and Hidden Markov model for daily routine activity recognition [88, 150].

- **Graph-Based models:** In this class of algorithms the data is represented by the nodes of a graph, where the edges are labelled with pairwise distance of the incident nodes and then the distance of two points is computed by minimizing the aggregate path distance over all connecting paths. The intuition is to construct a graph using domain knowledge or similarity of instances. One common method is to connect each data sample to its k nearest neighbors or to samples within some distance ϵ, where the weight $\boldsymbol{W}ij$ of an edge between each two samples x_i and x_j is set to $e^{\frac{-||x_i - x_j||^2}{\epsilon}}$. For instance, in [140] authors have proposed and explored a graph-based semi-supervised technique for decreasing the level of experience sampling interruptions in a scalable activity recognition system.

2.3 Learning Human Activities

Human related data acquisition and inferring semantic activity (including gesture, action, behaviour and routine) labels have created specific research challenges.

In this section, we aim to briefly study the common issues and properties pertaining to designing and developing an activity recognition system. The related works in human activity recognition can be categorized from many point of view. Here we describe the most important characteristics of design and implementation of activity recognition systems.

- **Goal of the system:** Activity recognition systems can be categorized into three types in terms of their intended tasks; *predictive* systems that learn from the past events and need to predict the future events such as next location or user action, *decision making* systems that learn from the current state and situation of the user and are required to derive a real-time decision, e.g., location-based mobile advertising services, and *analytic* systems that usually learn from a large-scale or long-time observations of user activities and infer higher-level information about their behaviuoral routines and interactions.
- **Response time:** According to their response time activity recognition systems can work *offline* or *online*. The former record the data from sensing devices and afterwards, process the data and provide feedback. The latter systems acquire data from sensing devices and process it in real time, and hence, can provide immediate feedback on the performed activity.
- **Subject-dependency:** An activity recognition system can be personalized for each individual or be a general flexible model that works for all individuals. User specific models imply that the system should be re-trained for each new user and tailored to the user specific characteristics. Furthermore, some systems take advantages of combining the personal and general models to form

a hybrid model that maximize the benefit from labelled data. Moreover, the activity recognition system can only be focused on individual activities and actions or study the organizational behaviour of a group of users. The latter case targets a group of human subjects and aim to learn patterns in their social interactions, interests, habits and participations.

- **System model:** When there is enough labelled data available for studying an application, in general a statistical machine learning approach is utilized to train a model based on observations, called *data-driven* approach. In contrast, *knowledge-driven* approaches (e.g., rule-based models) depend on rich domain knowledge and heuristics to capture and encode the characteristics of the problem using a set of logical rules found by greedy searches. There are also hybrid models which bridge the gap between theoretical domain knowledge and a set of classified examples of a problem.

2.3.1 Activity Recognition Challenges

While activity recognition application faces the general research challenges of all pattern recognition systems, they also need to cope with a number of specific domain-specific issues.

Definition of activities. The first specific challenge of the activity recognition problem is the complexity and diversity of human activities that can be performed in many different ways. Although state-of-the-art researchers have reported good performances on many types of activities, providing a clear taxonomy of human activities is still an on-going investigation. Therefore, the activity recognition field is far from reaching a good understanding and definition even for simple activities.

Table 2–2: Common types of activity classes used for human behaviour analysis.

Activity class	Examples	References
Ambulation	walking, running, sitting, standing, cycling, climbing stairs, driving a car, lying, treadmill, rowing, jumping, cross trainer, riding elevator	[131, 92, 87, 143]
Everyday routines	eating, drinking, brushing teeth, watching TV, vacuuming, hygiene	[28, 90, 131]
Location-based	proximity, localization, mobility tracking, mobility pattern, next-location prediction	[84, 31, 6, 67]
Healthcare	breathing, eye movement analysis, body position, abnormal events (falling, high blood pressure)	[7, 30, 158, 71]
Human-computer interaction (HCI)	typing keyboard, making call, texting, moving a computer mouse	[51, 53]

Table 2–2 briefly summarizes important examples of activity classes introduced in the related works.

Complex and overlapping activities. Different kinds of activities have their own specific characteristics that introduce completely different problems to the field. The challenge explained in previous paragraph became even more complicated when complex routines or behaviours are to be recognized. Complex activities are composed of several consecutive simple sub-activities that might be performed with certain logical sequence, speed and duration. For example, *preparing food* is composed of several instances of washing objects, chopping, shredding, cooking, among others, that might be carried out in an interleaved manner.

Moreover, studies in the field generally assume that a user only perform one activity at a time which can be true for some certain activities such as *cycling* and *sitting*. However, in real-life scenarios boundaries between activities are fuzzy since they can take place concurrently, alternating or overlapping. For example, an individual can be *watching TV* while *cooking* or may *fall* when trying to *sit*. Therefore, capturing activity boundaries and transition points brings additional uncertainty to the problem.

The work in [80] is an interesting study, which first uses supervised learning to assign low-level simple activity labels to the sensor data, and then employs an unsupervised probabilistic modeling, i.e. Latent Dirichlet Allocation (LDA), to discover structures in activity patterns. In this work, a fairly large set of high-level complex events, such as *discussing at whiteboard*, *picking up cafeteria food* and *preparing food* are discovered using data from wearable sensors. LDA is an example of a *topic modeling*, which is a type of probabilistic modeling for discovering abstract *topics* that occur in a collection of text data. This family of unsupervised learning are excellently suited where the tasks is to discover and model structure of multiple events in different levels. In Chapter 4 and 5 of this thesis, we have adopted this framework to extract high-level structure in users activities from low-level data.

Long-term monitoring. There are several real-life situations wherein human activity recognition or monitoring needs to provide long term supervision or assistance to the users. A few examples of such scenarios include assisted living and medical applications for people with chronic problems, group behaviour and routine analysis or sport application for recording daily performances. Long-term continuous measurement of activities dramatically increase the cost and complexity of the recognition system and impose additional constraints to confidentiality and intrusion concerns. In recent years, many studies suggest utilizing inertial sensing infrastructures, where taking measurements do not interfere too much with user's normal life, and ultimately, shifting the research direction towards device-free solutions. For example, the study in [7] suggests using a radio-based radar technique, called FMCW (Frequency Modulated Carrier Waves), to separate the reflections arriving from human body to continuously, yet non-intrusively, monitor chest movements in order to extract breathing pattern and heart rates. Another example, is the work in [65], where acceleration data from a wristwatch is incorporated with prior knowledge regarding the duration of the activities, by coding them as constraints and sequence patterns, in order to monitor daily routines (e.g. *showering*, *dinning* and *brushing teeth*) of users.

Predictive learning. Another important aspect of human activity recognition is how well the system is capable of predicting user's future move/activity/location. Several statistical models (mostly based on Markov models, Bayesian networks and Neural networks) have been proposed in the literature to learn the temporal/spatial patterns in sequences of events and and try to predict the upcoming activities/locations based on the previous observations. For instance, the work in [127] investigates the feasibility of indoor next location prediction using sequences of previously visited locations and employing a recurrent neural network, i.e., Elman networks, in the context of smart office building. This ability to predict one or few next activities or locations could be extremely helpful in applications such as advertisement, transportation and risk management. For example, a location-based recommendation model is proposed based on collaborative filtering and genetic algorithms in [48], which leverages user's preferences and context information to predict the potential items from available products or services.

Activity recognition datasets. If we intend to develop a comprehensive and clear understanding of human activities and behaviour, a large amount of high quality and diverse data should be available. In contrast with many application fields (such as speech recognition and computer vision) in which various standard datasets are available to evaluate the effectiveness and accuracy of proposed algorithms, there are limited number of publicly available datasets that contains information about people's activities. Many research groups in this field try to design their customized setting and collect their own problem-specific data under laboratory settings with limitations such as few number of human subjects and staged activities. However, for reproduction and quantitative comparison it is crucial to utilize standard datasets as benchmarks to evaluate new approaches. Nevertheless, a few research institutions and open data challenges have started joint efforts to prepare standard data and invite other researchers to participate in the data analysis process, e.g,:

- **Nokia Mobile Data Challenge:** MDC dataset is a large dataset generated from embedded sensors in mobiles of nearly 200 participants and it was designed to reveal information about behaviour of individuals and social networks [98].
- **Reality Mining - MIT:** Several mobile datasets colleted by MIT Human Dynamics Lab which contain the dynamics of several communities in order to contribute to human behaviour and interactions studies [3].
- **Activity Recognition Challenge:** Include a publicly available benchmark database of daily activities recorded in a sensor rich environment [1].
- **Data 4 Development (D4D) Challenge:** An open *Big Data* challenge including anonymous data extracted from the mobile networks of African countries (Senegal,2014 and Ivory Coast,2013) to gain new insights on social and economic solutions for these underdeveloped regions [2, 25].

2.3.2 Machine Learning for Human Data Analysis

In general, there are several problem-dependent properties that should be taken into consideration when determining the learning strategy including feature space dimension, computational budget, time constraints and number of classes. Researchers in activity recognition field have assessed a wide range of machine learning techniques to reveal the most efficient methods for inferring human actions, behaviours and movements from sensory data. Table 2–3 exhibit several examples of different supervised and unsupervised methodologies used by state-of-the-art activity recognition systems.

Table 2–3: Examples of diversity of machine learning algorithms used by human activity recognition systems.

Learning Method	**Algorithm**	**Act. num & type**	**Acc.%**
Neural Networks	Multilayer Perceptron [92]	6 ambulation	91
	Artificial Neural Networks [12, 87]	19 ambulation, 15 ambulation-static	96, 98
Temporal probabilistic	HMM [44, 173]	6 daily routine-ambulation, 12 ambulation	93, 90
Bayesian	Naive Bayes [46, 82]	11 daily routine-ambulation,6 ambulation	92,80
	Bayesian Networks [122]	7 office daily routine	97
Support Vector Machine	non-linear SVM [12, 79, 108]	19 ambulation, 16 daily routine-ambulation-high level(e.g., preparing for work) , 6 fall detection-ambulation	98, 79, 96
Ensemble	Random forest [33, 108]	5 ambulation, 6 fall detection-ambulation	94, 94
	Adaboost [99]	8 ambulation-daily routine	90
Instance Based	k-nearest neighbor(kNN) [12, 106]	19 ambulation, 7 daily routine-HCI	99, 98
Unsupervised	Latent Dirichlet Allocation [80]	10 ambulation-daily routine	72

However, the nature of human activities poses specific challenges and limitations to the learning procedure. In particular, we describe the most common challenges of machine learning with human activity data as follows.

Activity detection. Many applications of activity recognition assume that the stream data has been segmented ahead, at the start and end points of an activity, and the system only requires to predict the activity occurred in the segments. Alternatively, in some scenarios the system is expected to automatically spot the occurrence of an activity in the streaming data. In the latter case, a hierarchical learning processes is usually employed, where an extra layer of *activity detection* is embedded in the system. The activity detection layer highlights the segments where the activity performance have happened and then at next layer the classification unit build a model for activity prediction. The detected even can be either an abnormal situation like *falling* or a normal daily activity like *walking*. For example, a wifi-based device-free fall detection system is recommended in [71], which monitors wireless signals within a wifi-covered area to detected anomalous data points by measuring a local outlier factor, the local density of a given data point with respect to its k-nearest neighbours.

Time series modeling. Most of the human related data are naturally indexed over time, implying that the input of activity recognition systems are consecutive sequence of observations, i.e., time series. Time series data introduces new requirements to all of the processing steps including preprocessing, feature computation and learning algorithm. The recognition system used for analyzing time-series data should be able to capture temporal variations in measurements originated from human movement and behaviour, while being robust to external variations such as sensor displacement. Additionally, trajectory data, which is the path of a moving objects through space as function of time and can be captured as a time-stamped series of location-points, has become an increasingly important research theme in human activity recognition. Therefore, various types of probabilistic models have been proposed on spatio-temporal analysis of human data to capture underlying structure of mobility behaviour of the users. For example, in [126] a hierarchical inference model for location-based activity recognition and significant place discovery is proposed that classifies GPS trajectories of users into a sequence of activities such as *walking*, *driving*,*sleeping*, and then, identifies a user's significant places (e.g., work, home and bus stops) from the pattern of these sequences. In Chapter 4, we will elaborate on different time series data analyses methodologies and related work on trajectory-based activity recognition systems. Also, we introduce a new framework for long-term analysis of trajectory data within a few different real-world application scenarios.

Another considerable aspect of human data time series is the temporal pattern of activity's occurrence, which broadly creates three groups of activities; periodic, static and sporadic activities [29]. Periodic activities, e.g. *walking*, *cycling*, *swimming*, and *rowing*, contain regularly repeating sequences of actions. For recognition of these activities usually time and frequency domain features are extracted and sliding window segmentation is used for classification. The work in [63] introduces geometric template matching, which is an efficient algorithm based on a combination of feature extraction using time-delay embedding and supervised learning. The proposed feature extraction techniques can model and reconstruct the essential states and dynamics of an underlying dynamical system from a short sequence of measurements (evenly spaced in time) of the system, while different periodic activities such as walking and running are performed.

Static activities, e.g. *sitting* and *standing*, are isolated simple actions that do not exhibit periodicity or temporal pattern. This class of activities usually initiate with an activity detection step and then the type of activity is learned. For instance, in [155] off-the-shelf wifi devices are used as sensing infrastructure in a wifi-based activity recognition system to capture wireless signals and identify 7 different static in-place activities, such as *washing dishes* and *sleeping*, from the data. Sporadic activities happen at irregular intervals in time and interspersed with other activities, e.g. human gestures. Similar to anomaly detection, the recognition of this group of activities includes an accurate segmentation step in order to isolate the intended action, followed by a classification step to identify the activity.

Cost-sensitive learning. There are many applications in activity recognition (e.g. healthcare and medical monitoring), where the system needs to be sensitive toward misclassification of some specific activities. For example, for systems monitoring elderly people at home, confusion between *falling* with *sitting* is not tolerable. These learning models incur different penalties by incorporating a cost matrix C_{ij}, where the values in this matrix represent the cost of predicting activity i given the actual activity j [54]. In this regard, the learning algorithm is manipulated to a learner that outputs the activity class with minimum misclassification cost. For example, the abnormality detection system in [78] uses RFID-based sensor networks to observe daily activities of elderly people and employs a hybrid SVM/HMM approach to identify abnormal events.

Unbalanced data problems. In human activity learning, like many other learning applications, class imbalance is a frequently occurred and considerable problem. In activity recognition this is more challenging for long-term behavioural monitoring where only a few number of abnormal events might happen compared with usual activities and events. In general, class imbalance can be addressed by many strategies such as oversampling the smaller class or undersampling the bigger class to achieve equal class distributions in the training data, or alternatively, generating synthetic data to expand the smaller classes.

Interclass variability and similarity. This challenge arise from the fact that the same activity, such as walking, may be perform differently by different individuals and it also may occur when the same individual perform the same task differently [29]. Many environmental and generic factors can contribute to this effect in the performance of activities and the learning system should be prepared to cope with intraclass variability. To deal with this problem, one solution can be to used larger amount of training data from and also include data from different users in the training phase. Another alternative solution would be to include more specific features (e.g., full-body models) that could potentially discriminate not only among activities but also among users.

On the other hand, the inverse problem occurs when different activities with very similar characteristics (e.g., climbing stairs vs. descending stairs) have to be distinguished. This type of problem can usually be resolved by incorporating more sensing modalities (including modalities such as barometric pressure or gyroscope in the stairs example).

2.3.3 Hierarchical Activity Recognition

The primary focus of many activity recognition systems have been on accurate detection of simple, *low-level* and often predefined activities through low-level sensory observations. However, in recent studies, the concept of human activity recognition has evolved from a simple event or action modeling process to a higher level conceptual understanding of human behaviour. For example a *High-level* activity may refer to a complex activity (e.g. *cleaning the house*) that consist of several simple activities (e.g. *walking, vacuuming, standing*) or to a long-term functional status (e.g. *daily routines, social habits*) inferred from extended monitoring of short-term events (e.g. *eating, launch time, visited locations*) of a user. Therefore, the hierarchical structure of activity learning algorithm allows layered inference of high-level activities based on combination of low-level subcomponents by adding an additional layer of semantic representations. A wide range of research topics can benefit from the fusion of different levels of human activities. For instance, a notable number of studies have focused on discovering high-level mobility patterns, through long-term monitoring of the users' daily location traces and behavioural routines [103, 161].

Related works in this area have proposed different learning approaches to infer high-level and/or long-term activities of human daily life. One natural solution is to use HMM-based models that have been successfully applied in sequential pattern recognition problems, where the sequential pattern can be decomposed into piecewise stationary segments and an underlying stochastic process (i.e., the sequence of states) is assumed that is not observable, but affects the observed sequence of events. In many studies, layered HMM representations have been employed to efficiently model different level of activities and at the same time take into account temporal dependency of activities within events at different levels [122, 43, 100] However, the HMM based models are particularly useful when certain (temporal) dependencies between variables can be assumed, and moreover, the accuracy of these algorithms are seems to be very sensitive to unobserved changes.

Alternative approaches including Dynamic Bayesian Networks (DBN) model, Conditional Random Field (CRF) and Latent Dirichlet Allocation (LDA) model (see Section 2.3.1, Complex and overlapping activities), have been investigated for higher-level activity recognition, depending on specific characteristics of the input features and ultimate intention of the study [122, 150, 80]. For example, DBN models have been shown to offer a structural framework that combines prior knowledge and observable data, provides efficient reasoning under uncertainty and allows handling of incomplete data [122].

Nevertheless, there are several challenges that prevent many researchers to extend the study and analysis of human behaviour patterns beyond low-level, short-term activity recognition. In addition to general difficulties of human activity recognition (mentioned in Section 2.3.1), a realistic study on high-level semantic recognition includes a tedious and time-consuming data collection step. Moreover, analysis of human behaviour in this scope significantly increase the amount of annotation effort. Finally, making these recognition strategies scalable is a nontrivial task since many parameters, including individual and cultural variations, may extremely influence the semantic definitions and thus, the performance of the systems.

The lack of comprehensive study in high-level activity recognition and limitations of existing techniques motivate us to consider the development of new realistic and scalable solutions. In Chapter 4 of this thesis, we extend the study of high-level human behaviour analysis in the context of location-aware computing and present a novel practical step toward recognition of semantics in human mobility data. Also, in Chapter 5, in addition to low-level detailed activity recognition using wifi signals, a high-level analysis of these measurements in the context of smart environments is presented.

CHAPTER 3
Sensor Selection for Efficient Activity Recognition

Recognizing everyday activities is an active area of research in machine learning, human-computer interaction and context-aware computing. Information obtained from embedded modalities in smart wearable devices is especially valuable, due to the fact that these devices have become ubiquitous and come equipped with many modalities such as GPS, accelerometer, digital compass, gyroscope, barometer, WiFi and infrared sensors, which can query the local environment and yield information about the user's context and activities. Furthermore, these devices are popular for human data collection applications due to their compact size, low cost, non-invasiveness and low power consumption. Therefore, human activity recognition from data gathered by such devices has become a great interest among researchers in the field. However, activity recognition using wearable devices impose restrictions in terms of computational and energy resources, which need to be taken into account by the learning strategy.

On the other hand, most of existing studies in user activity recognition tend to include as many sensing modalities as possible in order to maximize the context information provided to the learning module. Subsequently, the feature extraction unit produces a large and complex feature vector including both low-level features (e.g., time and frequency domain content) and higher-level measures (e.g., number of objects detected by the proximity sensor) [42, 19, 109]. At the end, a feature selection phase is needed to determine which of these features are most useful, relevant and informative for classification. However, in realistic scenarios where the activity recognition application is expected to function along with many other applications on a mobile device with limited power, it is crucial to avoid unnecessary or unproductive sensing effort. Therefore, the problem of efficiently selecting the sensing modalities (or prioritizing information) is an essential issue for activity recognition using wearable devices.

In this chapter, we propose an efficient design of sensor selection that takes the variation of activities into consideration and maintains a desired level of performance while avoiding excessive computation burden to the system. We propose a real-time learning strategy, which interactively determines the most effective set of modalities (or sensors) considering the activity in action. Inspired by the *active learning* framework, the activity classification task starts with an initial, small subset of sensors. Then, at each timestep, a probabilistic *certainty measure* (or *confidence score*) is computed, which determines the certainty of the classifier on the predicted activity label. If the classifier is not confident enough about the labeling decision, the system queries the smart device to incorporate more sensors, until the most informative yet cost-effective subset of sensors is involved. Our approach is computationally simple and allows accurate classification of activities while minimizing energy consumption for small portable devices.

3.1 Background

In this section, we present the necessary background material for this chapter. We begin with an brief overview of existing approaches for sensor selection, followed by a review of the Random Forest algorithm that is used in our approach as the classification technique.

3.1.1 Related work

There are many application fields, including sensor networks [168] and computer vision [145], where senor selection is an important decision that must be made based on information utility and cost. The techniques of sensor selection can be broadly classified into two main categories [167]. First, there are greedy approaches that consider sensor selection as a heuristic search problem and aim to find the best solution among all possible combinations [69]. These class of approaches usually begin with a heuristic initial combination of sensors and then iteratively improve the solution, for example through random perturbations [8] or adaptive entropy-based aggregation [56], until it converges. Second, there are decision-theoretic techniques, which treat the sensor selection process as a decision making problem with multiple solutions and limited resources. There are models such as partially observable Markov decision processes (POMDPs), which can be solved using stochastic dynamic programming [34].

However, both types of methods are computationally expensive due to combinatorial explosion, and are not practical in a real-time implementation except for small problems. We are interested in the sensor selection problem in the context of real-time activity recognition for smart portable devices, where beside the sensor selection unit, a lot of computing power is required by other different parts of the activity detection application, including heavy-duty feature extraction/selection and classification tools. Moreover, regardless of the sensor selection technique category, most of the existing algorithms usually select a final optimum subset of sensors for the classification of all classes. This often requires all the selected sensors to be engaged all the time and ignores the fact that the impact of each sensor varies depending on the class of activities.

3.1.2 Random Forest: A classifier with built-in certainty score

Our approach will leverage Random Forests [26], an effective ensemble learning method that applies bootstrap aggregating, or bagging technique to decision tree learners. The main idea of ensemble learning for classification is to merge a group of regular learning models and combine their decisions in some manner in order to efficiently improve performance of the learners. These ensemble methods are widely employed in activity recognition applications due to their high overall accuracy and their ability to handle diverse features and noise. Following provides examples of the most popular fusion methods:

- **Boosting**, is a special case of model averaging method which first creates a 'weak' classifier, and then iteratively rebuild a stronger model from a dataset, in which misclassified points by the previous model are given more weight. Finally, all of the successive models are combined by some weighted majority voting schemes. Many algorithms, such as AdaBoost, fit into this framework where a weak model (i.e., slightly better than random guessing) is repeatedly forced to concentrate on the modified version of the data.

- **Bagging or bootstrap aggregation**, is a another model averaging strategy where multiple version of training set (produced by sampling with replacement) are used to train a set of classifiers and the, the output of these models are combined by voting to provide a single decision. As an example, Random forests algorithm combines random decision trees with bagging to reduce variance and stability in the learning process and construct a significantly improved estimates.

Random Forest classifiers are used in many applications because beside classification, they also provide information on attribute importance and confidence level in the classification. Intuitively, a Random Forest builds many classification trees, where each tree votes for the class label. The forest chooses the classification with the most votes over all the trees. Let's assume we have N instances in the training set and there are M features for each instance. In order to grow each tree of the forest, the algorithm proceeds as follows:

1. Sample N instances at random with replacement to create a subset of the original data for training the tree.
2. At each node, a given number of $m \ll M$ features are randomly chosen out of M and the best split is determined only according to these m features. The value of m is constant during the construction of the forest.
3. Each tree grows until all leaves are pure, i.e. no pruning is performed. In practical implementations, usually the depth of the trees is a fixed number and can be tuned using cross-validation.

When the training set for each tree is drawn by sampling, normally about one-third of the N instances are left out of the bootstrap (bag) to be used as a validation set, in order to get a running estimate of the classification error as trees are added to the forest. This out-of-bag (OOB) error is also used to get estimates of variable importance by monitoring the number of correctly classified OOB samples, while the m attributes are randomly shuffled for each single tree.

In practice Random Forests have been used extensively due to their ease of training and very good generalization performance as well as due to the fact that they provide insight into the importance of different features and in the uncertainty of labels. In our application of Random Forest, in particular, we are also interested in the confidence level of the classifier's prediction for each instance, which can be assessed by the proportion of the votes given by all trees for the correct class. We will use this ratio as an estimate of the confidence level, in order to score the different sensors under consideration in the sensor selection procedure.

3.2 Active Learning for Sensor Selection

In this section we propose a real-time activity recognition algorithm which actively selects a smaller subset of sensors that are the most informative, yet cost-effective, for each time frame. First, we use a greedy process to discover sets of sensor modalities that most influence each specific activity. These subsets of sensors are then used to build different models for the activities. We use the baseline models to develop an algorithm that decides on-line which model is suitable for recognizing the activity in each given time frame. Our algorithm has the flavor of active learning [137], but instead of asking for labels on new data points, we start the recognition task with a small set of sensors and then interactively send queries for more features, as needed. In this way, we can afford to run the activity recognition engine on a low-powered device without sacrificing the accuracy. We present empirical results on real data, which illustrate the utility of this approach.

3.2.1 Dataset

The dataset used for activity recognition was collected by Dieter Fox [142], using the Intel Mobile Sensing Platform (MSP) [42], which contains several sensors, including 3-axis accelerometer, 3-axis gyroscope, visible light photo transistor, barometer, and humidity and ambient light sensors. Six participants wore the MSP units on a belt at the side of their hip and were asked to perform six different activities (*walking*, *lingering*, *running*, *climbing upstairs*, *walking downstairs* and *riding a vehicle*) over a period of three weeks. Annotation was acquired through observers who marked the start and end points of the activities. Table 3–1 demonstrates the non-uniform percent distribution of activity class labels. The working dataset is 50 hours of labelled data (excluding the beginning of each recording which was labelled as *unannotated*) and also some long sequences (over 1 minute) labelled as *unknown*. There were also some short unlabelled segments, which we smoothed out using a moving average filter. We computed the magnitude of the acceleration $\sqrt{x^2 + y^2 + z^2}$ based on components sampled at 512 Hz. We also used the gyroscope (sampled at 64 Hz), barometric pressure (sampled at 7.1 Hz) and visible light (sampled at 3 Hz). These four measures were all up-sampled to 512 Hz in order to obtain synchronized time series with equal length. To prevent overfitting to characteristics of the locations, we did not include the humidity and temperature sensors, as they could potentially

Table 3–1: Activity class distribution.

Downstairs	Drive	Linger	Running	Upstairs	Walking
1%	33%	42%	1%	1%	22%

mislead the classifier to report a false correlation between location and activities. For example, if a lot of *walking* data were collected under hot sun in a warm day, the classifier would see temperature as a relevant feature to the action of *walking*.

3.2.2 Baseline Classifiers

First, we wanted to investigate the effect of different subsets of sensors on the accuracy of recognizing the six different activities. Through this experiment we seek to answer the following questions:

- While all of the embedded sensors on a smart device are potentially interesting and can provide some useful information about the user's context, is there any modality that does not significantly contribute information about a particular activity (or a set of activities)?
- Can we decrease the power consumption by discovering a subset of sensors that can adequately provide effective information about a particular activity (or a set of activities)?
- In this case, how much the recognition of each particular activity is influenced by the omission of each sensor?

We began by examining all possible combinations of sensors on the entire dataset. We treated each time sample as an instance and used raw sensor data as features for building baseline classifiers. Random Forest was employed as the classification technique, due to the reasons stated in Section 3.1.2. We performed cross-validation

Table 3–2: Accuracy of individual baseline classifiers. The highest accuracy in each section is in bold.

No.	Feature Set	Accuracy
1	$\{Acc, Bar, Gyro, VisLight\}$	**86.16**
2	$\{Acc, Bar, Gyro\}$	75.16
3	$\{Acc, Bar, VisLight\}$	**86.50**
4	$\{Acc, Gyro, VisLight\}$	84.33
5	$\{Bar, Gyro, VisLight\}$	78.33
6	$\{Bar, Gyro\}$	54.00
7	$\{Acc, Gyro\}$	69.50
8	$\{Acc, Bar\}$	74.83
9	$\{Acc, VisLight\}$	**77.66**
10	$\{Bar, VisLight\}$	74.00
11	$\{Gyro, VisLight\}$	74.00
12	$\{Acc\}$	**48.16**

over users (leaving in turn each user's dataset aside as the test set and combining and randomizing all other datasets to use as training set). The accuracy of the classifiers for all 12 possible combination sets of four sensors [1] is given in Table 3–2. From now on, instead of full sensor names, we use the abbreviations Acc, Bar, $Gyro$ and $VisLight$ for accelerometer, barometric pressure, gyroscope and ambient visible light, respectively.

Preliminary Evaluation. For classification, we have used the Random Forests implementation by Leo Breiman and Adele Cutler in [27]. In both training and testing we used 100 trees in the forest and the depth of the trees was set to 2. It is recommended to prune (limit the depth of) the trees when dealing with noisy data.

[1] Single features except the accelerometer are excluded from the results due to poor performance.

The overall results are competitive with prior activity recognition results that used complex feature sets, even though we used the raw sensory values [42, 19]. The recognition accuracy is the number of correct predictions over total number of data points, and is defined as $\frac{\sum_i A_{ii}}{\sum_i \sum_j A_{ij}}$, where $\boldsymbol{A}$ is a contingency matrix whose element $A_{i,j}$ is the number of times that a data point from true label i was classified as label j. The classification results are show in Figure 3–1. Each plot is for one specific activity and shows the accuracy and standard error of all baseline classifiers. It is clear that not all sensors are contributing equally to the performance. For example, comparing results from classifiers No.1 and No.3 in Figure 3–1 shows that data from the gyroscope did not provide useful information about this set of activities. Moreover, this sensor seems to lead to similar or, in some cases, even smaller improvement in recognition accuracy compared to the barometer sensor. So we decided to prune the gyroscope.

The contribution of each sensor varies among different activities. For example, accelerometer data is key in discriminating physical activities such as *running* and *walking*. However, the classifier using only accelerometer data (No.12) performs poorly while recognizing some activities like *riding a vehicle* or while distinguishing activities with similar dynamics (e.g. *upstairs* vs. *downstairs*). Nevertheless, this classifier is the cheapest one (in terms of energy consumption) and it is reliable enough to be used as a default classifier for our active learner.

Table 3–2 shows accuracy results for all the test sets, using subsets of sensors. These results demonstrates that a smaller subset of sensors exist that can learn the same set of activities reasonably well, with less information. Classifiers No.3 and

No.9 achieved 86.5% and 77.66% accuracy rate, respectively, whereas classifier No.1 (using all sensors) only obtained 86.16%. Hence, using subsets of sensors is feasible and desirable.

3.2.3 Smart Classifier Selection

In contrast to supervised learning, which requires a large number of labelled samples to build an accurate classifier, active learning attempts to obtain better performance with less labelled data, and only queries the source of information for more labels upon request. There are many problems in which obtaining labelled instances is difficult, time-consuming, or expensive. In such scenarios, active learning starts with a small number of labelled instances, and then interactively queries for more labelled data. These algorithms are highly motivated in many modern machine learning problems, since the learner needs fewer examples to learn a concept, therefore minimizing the cost. We would like to adopt this idea in order to reduce the overall power consumption of activity recognition for low-power devices. However, instead of more labelled samples we want to query more sensory information under uncertain situations.

Here, we introduce a real-time algorithm that optimally selects the best classifier for each time frame, using as guidance the results that we presented in the previous section. The main idea is to start the activity recognition task by acquiring data just from the single most informative sensor and building a cheap classifier. The certainty measure provided by this classifier is then used to identify points in time when uncertainty is high, so using more sensors could be beneficial. Other classifiers can then be invoked. Figure 3–2 presents an overview of the information flow.

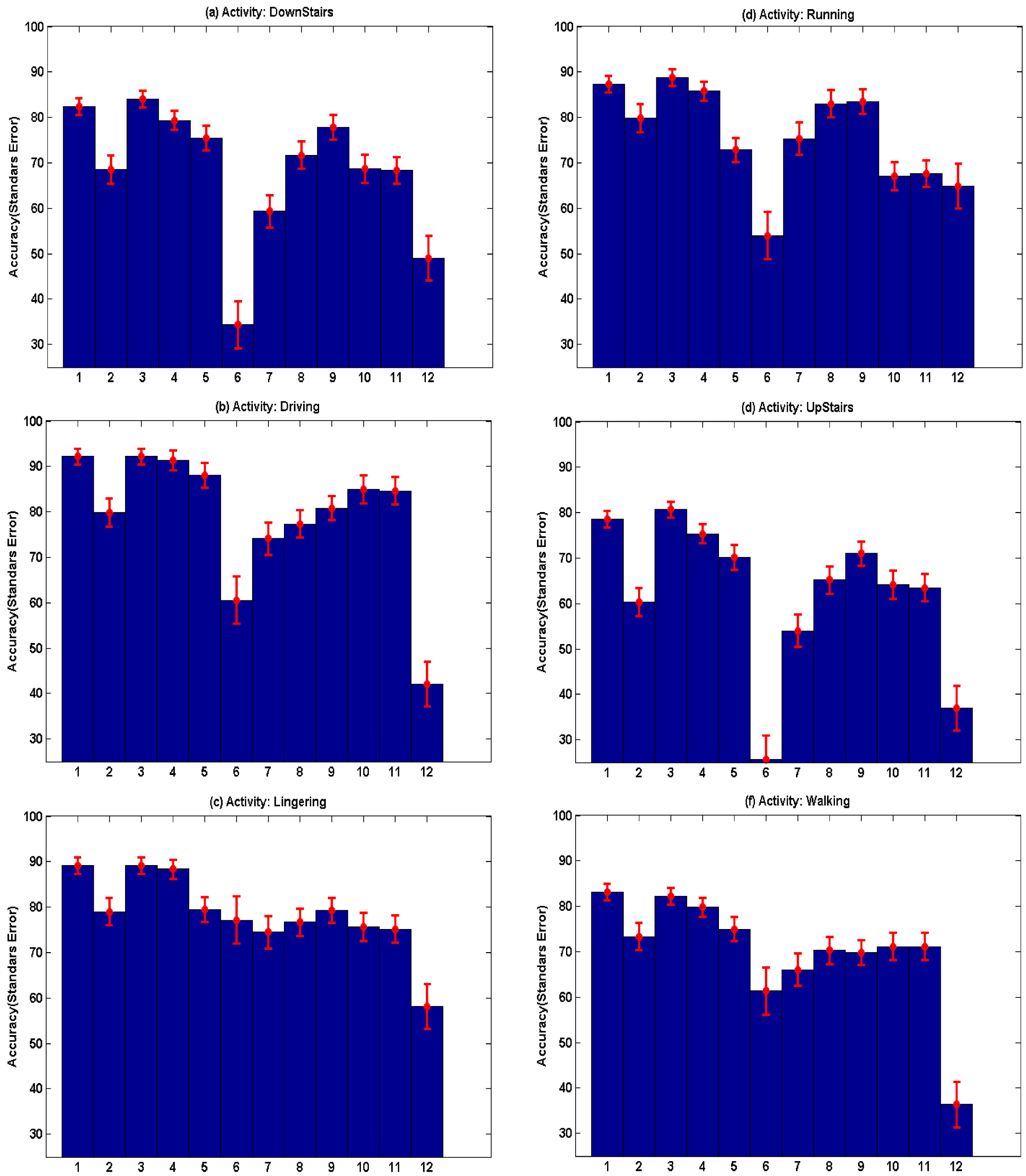


Figure 3–1: Accuracy rate and standard error demonstration for each individual activity

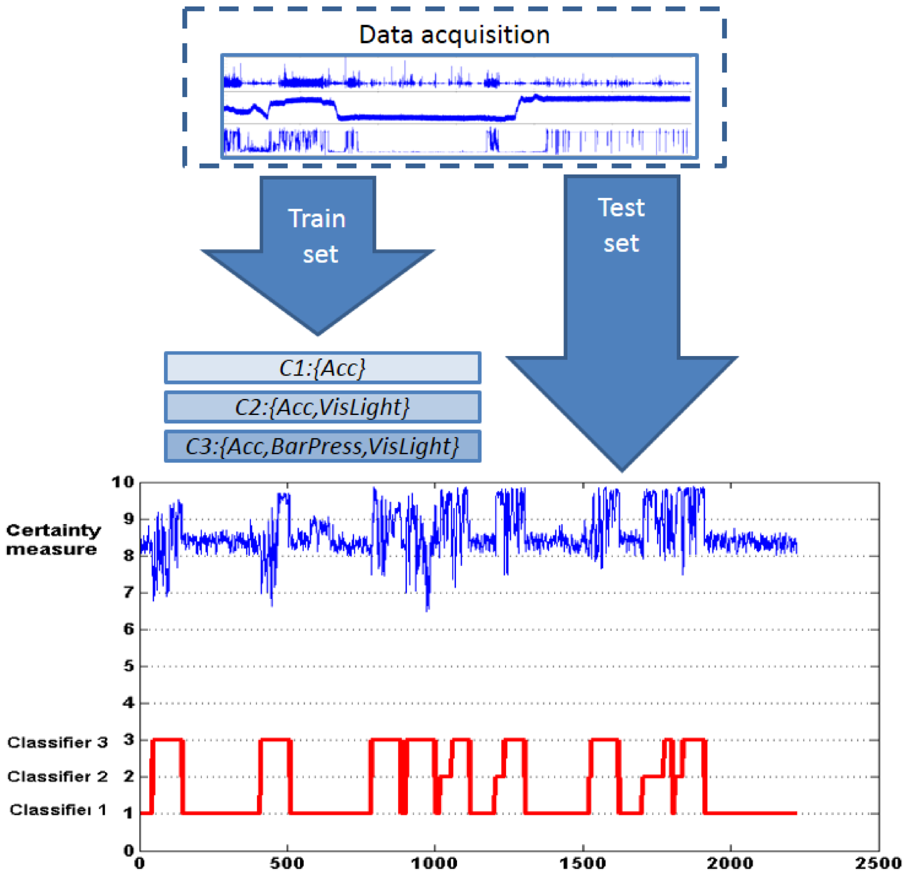


Figure 3–2: Overall structure of the active algorithm

The algorithm begins by training a set of classifiers, each using a feature set selected in advance, based on application-specific criteria. The experiment in the previous section is an example of how the best feature sets can be chosen, though one would not need to be so exhaustive. In general, the set of classifiers should contain at least one cheap classifier that can run all the time, and an expensive classifier with very good accuracy. Also, if there is a large number of sensor modalities, it is useful to have some classifiers that use different types of resources, not only for energy consumption, but also to ensure that the application is robust with respect to sensor failure, or unusually noisy readings.

When the training phase is over, the algorithm will have to process new time series. It starts by sliding a fixed-width window (of length $w = 200$), with 10% overlap, over the data, in order to obtain data intervals. We would like to keep the length of these intervals as small as possible, in order to avoid mixtures of activities, but large enough to capture the essence of the activity. Each interval is initially labelled by the cheapest classifier. We compute the running average of the certainty measure over each frame, to indicate if the classifier is confident enough about the labelling decision or not. If the measure drops below a given threshold, the algorithm will query other sensors, and upgrade the classifier to a more complex one, which works with the new information. The procedure continues until the most informative subset of sensors is chosen and the best classifier provides the most accurate prediction.

The algorithm will switch back to a cheaper classifier as soon as its certainty measure rises above the threshold. To do this, the algorithm simultaneously computes and compares the confidence level of both classifiers at each time frame, and switches back when the threshold is exceeded again. Ideally, we want the algorithm to have smooth transitions between classifiers, so we also use a control parameter, which allows the algorithm to switch from one model to another only after δ frames.

3.3 Evaluation

We evaluated the proposed algorithm on the dataset and selected subsets from the experiment in Section 3.2.2. The number of classifiers used is $N = 3$, where the classifiers are $C_1 = \{Acc\}$, $C_2 = \{Acc, VisLight\}$ and $C_3 = \{Acc, Bar, VisLight\}$.

Hence, the algorithm will first use only the accelerometer data for classification, then incorporate visible light, and in the worst case, barometric pressure as well. As we explained in Section 3.1.2 the certainty measure is the ratio between the votes that match the majority label and the overall number of the trees in the forest. In both training and testing we used 10 trees in the forest and the depth of the trees was set to 2.

There are two parameters that were chosen empirically, and which influence the accuracy rate:

- θ, the threshold for the certainty measure, which may depend on the overall accuracy rate
- δ, the number of frames before switching to another classifier is allowed. Tuning the parameter δ is important for controlling the balance between the speed of switching between the classifiers and the delay before choosing the efficient classifier.

Figure 3–3 shows how θ and δ affect the overall accuracy of the system. One can see that performance is stable for a fairly large range of these parameters. In this experiment we set $\delta = 90$ and $\theta = 8.2$ to efficiently maintain the high accuracy.

In practice, we found that switching between two classifiers, instead of three, yields better accuracy and smoother transitions. This happens because the algorithm does not stay with C_2 for long and tends to switch between C_1 and C_3. Figure 3–4 shows a run of the algorithm on a segment of data from one specific user, using classifiers trained on the other users' data, with $\delta = 90$ and $\theta = 8.2$. The delay that happens in detection of the correct classifier is caused by the parameter δ that

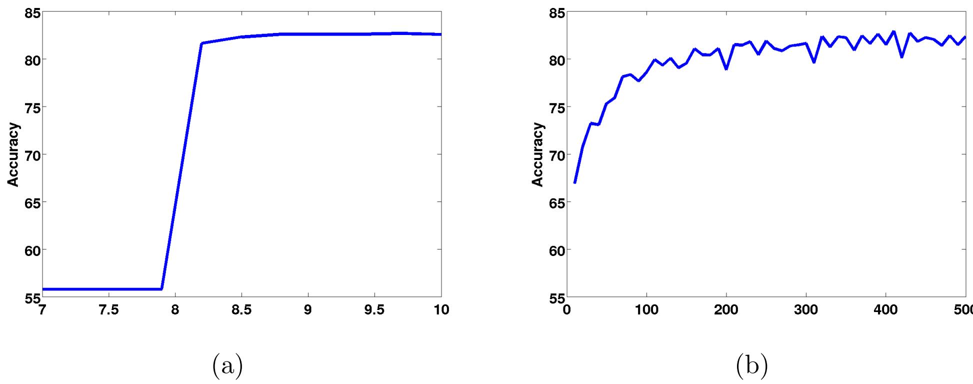


Figure 3–3: Influence of (a) θ and (b) δ on the accuracy of the activity recognition task.

Table 3–3: Comparison of recognition accuracy.

Algorithm	Accuracy	Proportion of time
Classifier C_1	48.16	100%
Classifier C_2	77.66	100%
Classifier C_3	86.50	100%
Active alg.(C_1, C_2, C_3)	71.78	9%,32%,59%
Active alg.(C_1, C_3)	80.14	35%,65%

forces stability to the system and does not allow quick switching between classifiers. This delay can be critical in time-sensitive systems, such as medical monitoring application, and is required to be measured and taken into consideration while tuning the parameters. Our application is fairly flexible to such small delays, which are approximately in the order of 2-10 milliseconds.

Table 3–3 shows the classification results of the proposed algorithm and the baselines from the first experiment, as well as the proportion of the time the algorithm used each classifier. The overall accuracy of the active algorithm (combination of 2 classifiers) is just 6% lower than the best baseline (C_3). While this is a given dataset

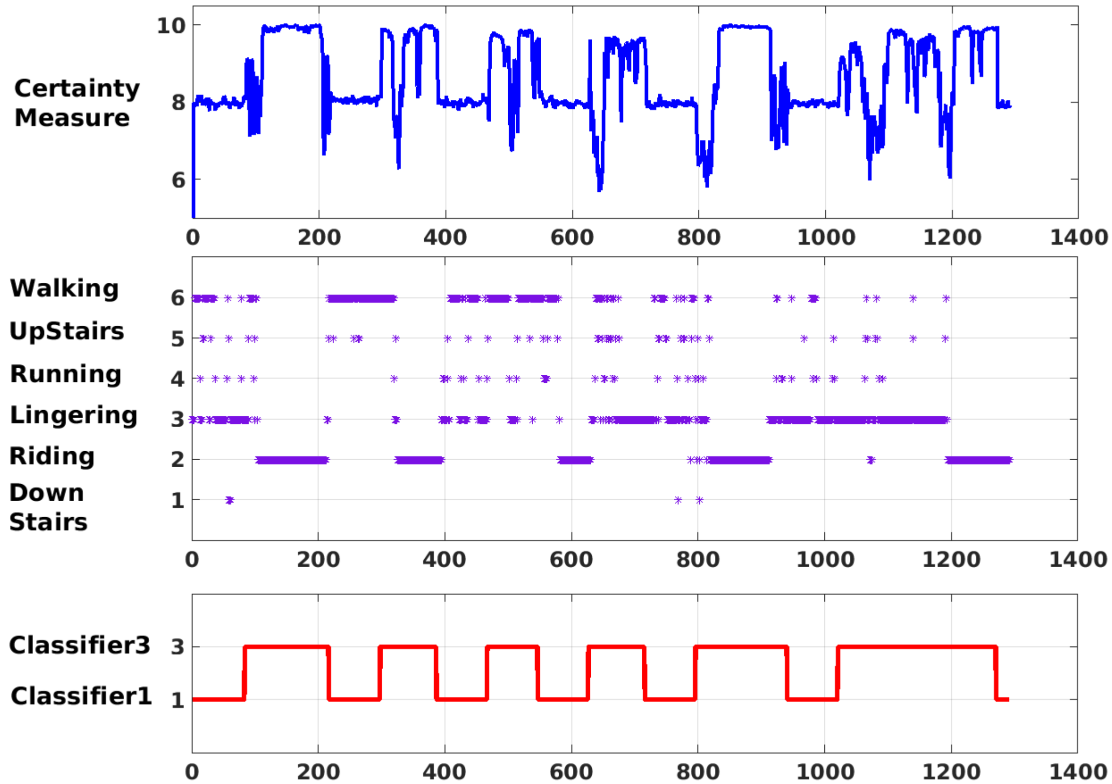


Figure 3–4: Algorithm performance on a segment of data. Top: Certainty measure of the classifier in use. Middle: The corresponding true activity labels at each time frame. Bottom: The algorithm's decision of the classifier to engage

and we can not measure energy consumption directly, using the cheap classifiers 35% of the time suggests that large energy savings may be possible.

3.4 Conclusion

We presented an approach that can be used to select among classifiers with different features (and power consumption) in activity recognition tasks. The active-learning-style idea is to use a certainty measure in the result of the classification to decide if a more "expert" classifier should provide labels. However, no input from a user is required, as the algorithm is fully automatic. We evaluated the performance of our algorithm on a large, noisy dataset including four different modalities, gathered while subjects were performing different activities under naturalistic circumstances. The empirical results show that our approach can successfully switch between complex and simple classifiers, on-line and in real-time, yielding power savings without significant loss in accuracy. In chapter 6 of this thesis, we will provide some directions for future possibilities of this work.

CHAPTER 4
Location-based Activity Recognition

The growing popularity of mobile computing devices with integrated location sensing technologies has promoted a huge interest in analyzing data collected from such devices. These modern intelligent devices have the ability to accurately track moving objects, store mobility data and process or transfer this information. Besides, many people voluntarily carry their location-aware smart devices (e.g., smartphones and tablets) everywhere they go, which allows long-term analysis and discovery of mobility patterns. We consider the problem of analyzing people's mobility and movement patterns from their location history gathered by portable mobile devices. Human mobility traces or trajectories can be extremely complex and unpredictable, by nature, which makes it hard to construct accurate models of mobility behaviour. In this chapter, we study location-based human activity recognition and present a novel approach for user mobility analysis using an unsupervised learning framework. We evaluate our algorithm on 3 different real-world datasets that have been collected from people over the course of 1 to 9 months. This chapter is organized as follow. First, in Section 4.1, we provide an overview of related work for location trajectory analysis, and a brief introduction of non-parametric Bayesian models. Section 4.2 presents the proposed methodology, a hierarchical clustering approach based on hierarchical Dirichlet processes (HDPs) for location data analysis. Sections 4.3, 4.4 and 4.5 introduce three different application scenarios, where the HDP-based approach is applied on real datasets and evaluated with different criteria. In each application, a real-world dataset is presented, the specific problem is described, and then, the empirical results are discussed. Finally, Section 4.6 concludes the evaluation and results of the proposed approach, with a discussion on potential future directions of this work as well.

4.1 Background

In this section we briefly introduce the problem of trajectory analysis in the context of location-based human behaviour recognition. We review a number of existing approaches for learning from trajectory data. This is followed by a brief review of non-parametric Bayesian models, which form the basis of our proposed approach for trajectory mining.

4.1.1 Location Data Mining

Recent advances in wireless technologies and satellite-based navigation systems allows a large amount of geo-spatial trajectory data to be generated. Location history data enables discovering valuable knowledge about human preferences and behaviour, and consequently, applications that utilize such information can offer customizable services according to the dynamics of their users' surroundings. Mobile advertisement, surveillance and security, health monitoring, urban planning and social network analysis are a few examples of application fields wherein location-aware computing has significantly improved performance.

Given this massive volume of information, in academic communities, wide research efforts have been put towards different directions in spatial data analysis.

One interesting direction is localization (and positioning), which focuses on the problem of determining the location of a moving object using observable fine-grained or coarse-grained data, such as GPS readings and wifi signals [6, 31]. Another direction includes developing algorithms for analysis and mining moving object traces and location trajectories [170, 18]. We are interested in the second category of studies, where the goal is to infer semantic patterns in users' long-term mobility traces.

Many existing studies in location trajectory analysis fall into distance/similarity-based approaches, which aim to summarize the shape of the whole trajectory into a representative and then, define a distance metric between such representatives and actual data [151, 115]. Several distance functions have been proposed for different applications including Dynamic Time Warping (DTW) [36], Edit Distance with Real Penalty (ERP) [38] and Longest Common Sub-sequences (LCSS) [151]. For instance, DTW was proposed to allow alignment between trajectories that are not of the same length. In practice, however, there are complex and long trajectories that convey important and interesting knowledge on their partial segments instead of the whole route. Therefore, these solutions are difficult to use for the long-term analysis of human location history, where one needs to use very long traces of data.

Another considerable amount of work has focused on defining a spatial-temporal models that synthesize individual or group movement patterns into statistical representations, which are also expected to predict future moves. These models are usually used as a basis for creating a similarity measure between different moving objects based on the sequence of visited locations. Since trajectories can be interpreted as sequences of location points, these approaches leverage existing sequence inference models. For example, [127, 64, 126] have used Markov models, including Markov chains and hidden Markov models (HMMs), to represent the mobility behaviour of an individual and predict their next location based on the previously visited places and the probability distribution of the transition between places. However, these approaches are limited while facing geometrical complexity of long trajectories, such as long-term human moving history, (which vary in shape and size) and stochasticity in movement patterns.

The problem becomes even more challenging when moving subjects go to new, previously unseen location points, or when trajectories belonging to different mobility patterns share a large set of common location points. To address this limitations, an increasing number of studies began to consider employing semantic factors, such as personal interests and preferences, for the analysis of long and complex human trajectories. These approaches suggest a hierarchical representation of human mobility behaviour where beside low-level detailed location data, high-level semantic information is leveraged to discover meaningful properties about users. Some interesting examples include the studies presented in [101, 162, 53, 169] where semantic information such as user similarities, social ties and friendship networks have been inferred from user location trajectories. This line of studies usually transforms the trajectories into other data structures such as graphs, matrices and tensors, which allow borrowing from a broader range of approaches, such as graph mining, matrix factorization and collaborative filtering, for trajectory analysis [104, 170, 169]. While these solutions are promising for some specific applications, many open challenges remain unsolved in high-level human location analysis. For instance, there is a great need for scalable approaches that minimize the amount of user involvement or annotation, or generalized systems that automatically distinguish the preferences and interests of large group of users based on their interactions.

The focus of our work is to infer interpersonal interactions of a group of people by analyzing their mobility traces and location histories. We seek to mine location histories in order to discover and characterize the points of interest (POIs) frequently visited by users, and then estimate a similarity measure between users based on the proximity of their POIs. To this end, we use unsupervised learning methodology, which applies hierarchical clustering for grouping people and their mobility behaviour based on their coarse-grained location trajectories.

4.1.2 Non-Parametric Bayesian Modeling

In this section we briefly describe Hierarchical Dirichlet Processes (HDPs) [147] which are the basis of our proposed approach. The HDP is a non-parametric Bayesian (NPB) mixture model typically used for clustering large collections of grouped data, e.g., topic modeling from text documents. The non-parametric term means that the number of clusters is open-ended [123]. The NPB mixture model offers flexible model selection, which adapts the number of clusters identified in a dataset to the complexity of the data. The HDP is a model of this type, which also allows sharing of components between clusters, as well as sharing of the clusters. The following presentation is based on [58, 116, 147] and [61, Chapter 2], where these topics are expounded in more detail.

4.1.2.1 Dirichlet Distribution

Let $p(\boldsymbol{y}|\theta)$ denote a probability model for a set of observed data $\boldsymbol{y}$ given the model parameters θ. The Bayesian framework place a prior distribution on the latent parameter θ, in order to predict the future observations based on the already observed data. Given a set of N i.i.d. observations, the *predictive likelihood* of future observations is:

$$p(\boldsymbol{y}|\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N, \lambda) = \int_{\Theta} p(\boldsymbol{y}|\nu)p(\nu|\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N, \lambda)d\nu, \tag{4.1}$$

where Θ is the space of parameters and λ is the parameter of the prior, called *hyper-parameter*, and can simply be seen as tuning parameters. Given a set of observations and the properly tuned hyper-parameters, the distribution over parameter θ or the *posterior density* on θ is:

$$p(\theta|\boldsymbol{y}, \lambda) = \frac{p(\boldsymbol{y}|\theta)p(\theta|\lambda)}{\int_{\Theta} p(\boldsymbol{y}|\nu)p(\nu|\lambda)d\nu}. \tag{4.2}$$

Consider a random variable $\boldsymbol{y}$ that can take non-negative elements from $\{1, \cdots, K\}$. In addition, suppose that $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_K)$ is an associated probability distribution with π_i representing the probability of observing the value i, and $\pi_i \geq 0, \sum_{i=1}^{K} \pi_i = 1$. According to the *multinomial* distribution the probability of a sequence of observations over N trials, $y_1, \cdots, y_N$, is:

$$p(y_1, \cdots, y_N|\boldsymbol{\pi}) = \frac{N!}{\prod_k N_k!} \prod_k \pi_k^{N_k} \tag{4.3}$$

where N_k denotes the number of times k happens in the sequence. Intuitively, the multinomial distribution models the distribution of the histogram vector over k possible outcomes. The *Binomial* distribution is a special case when $K = 2$.

In Bayesian probability theory, when the posterior distribution remains in the same family of distributions as the prior probability distribution for all possible observations and likelihoods, the prior is referred to as *conjugate prior*. Given this definition, the Dirichlet distribution is a conjugate prior for the multinomial distribution, where the samples from a Dirichlet distribution can be treated as parameters for a multinomial distribution and the posterior remains a Dirichlet. This means that if a data point has a multinomial distribution, and the prior distribution of the distribution's parameter is distributed as a Dirichlet, then the posterior distribution of the parameter is also a Dirichlet.

A K-dimensional Dirichlet distribution with parameter vector $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_k)$, which we denote by $Dir(\boldsymbol{\alpha})$, has the form of:

$$p(\boldsymbol{\pi}|\boldsymbol{\alpha}) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k \pi_k^{\alpha_k - 1}, \tag{4.4}$$

where $\alpha_k > 0$ and $\Gamma(.)$ denotes the Gamma function. We have a symmetric Dirichlet when the parameters $\alpha_1, \cdots, \alpha_k$ are all equal. The *Beta* distribution is the special case of this distribution when $K = 2$, which we denote by $Beta(\alpha_1, \alpha_2)$.

The conjugacy property of the Dirichlet distribution implies that, given N multinomial distribution $y_1, \cdots, y_N$, the posterior distribution of $\boldsymbol{\pi}$ is also Dirichlet:

$$p(\boldsymbol{\pi}|y_1,\cdots,y_N,\boldsymbol{\alpha}) \propto p(\boldsymbol{\pi}|\boldsymbol{\alpha})p(y_1,\cdots,y_N|\boldsymbol{\pi}) \tag{4.5}$$

$$\propto \prod_{k+1}^{K} \pi_k^{\alpha_k+N_k-1} \propto Dir(\alpha_1+N_1,\cdots,\alpha_K+N_K). \tag{4.6}$$

Therefore, the predictive likelihood can be derived as

$$p(y=k|y_1,\cdots,y_N,\boldsymbol{\alpha}) = \frac{N_k+\alpha_k}{N+\alpha_0}, \alpha_0 \triangleq \sum_{k=1}^{K}\alpha_k \tag{4.7}$$

where N_k represents the actual number of times k occurs in the observations and hyperparameters $\boldsymbol{\alpha}$ can be considered as pseudocounts, i.e. representing the number of counts added to the observations in each category that we have already seen.

4.1.2.2 Dirichlet Process

The Dirichlet process (DP) [58] is a stochastic process whose domain is the space of probability measures. Suppose we have a measurable space (Θ, β). A DP is defined as a distribution of probability measures G over (Θ, β), with the property that for any finite measurable partition $(A_1,\cdots,A_K)$ of Θ, the random vector $(G(A_1,\cdots,G(A_K))$ is a finite-dimensional Dirichlet distribution:

$$(G(A_1,\cdots,G(A_K)) \sim Dir(\gamma H(A_1),\cdots,\gamma H(A_K)) \tag{4.8}$$

where H is a *base probability measure* on Θ and $\gamma \in \mathbb{R}^+$ is a *concentration parameter*. We write $G \sim DP(\gamma, H)$ if G is a random probability measure drawn from a Dirichlet process. Figure 4–1 shows the graphical representation of a Dirichlet process.

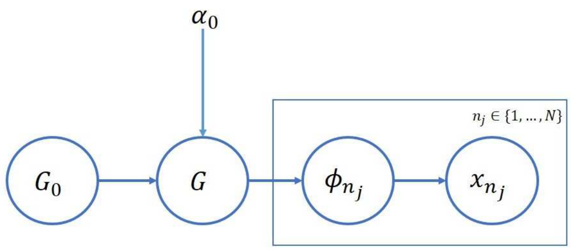


Figure 4–1: Graphical model of a Dirichlet process mixture model (reproduced from [147])

The Dirichlet process prior exhibits an important clustering property when used in the *mixture model* setting. There are several representations of the Dirichlet process that help understand its clustering property, such as Chinese Restaurant process (CRP) [9], stick-breaking process [136] and Polya Urn [23]. Here, we briefly review the stick-breaking process and the CRP.

The base distributions $G \sim DP(\gamma, H)$ can be written as:

$$G = \sum_{k=1}^{\infty} \beta_k \delta_{\phi_k} \tag{4.9}$$

where $\phi_k \overset{iid}{\sim} H$ and $\beta = (\beta_k)_{k=1}^{\infty}$ is a vector of weights obtained by a stick-breaking process:

$$\beta_k = \upsilon_k \prod_{l=1}^{k-1} (1 - \upsilon_l) \tag{4.10}$$

where $\upsilon_l \overset{iid}{\sim} Beta(1, \gamma)$.

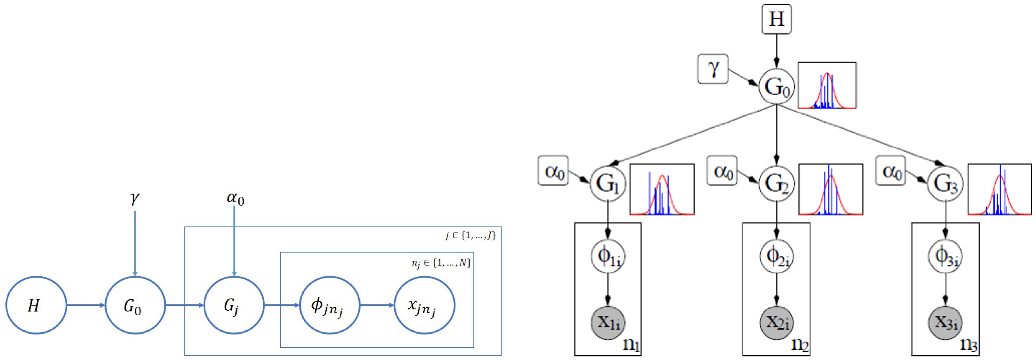
The construction for β_k can be seen as starting with a unit-length stick and repeatedly breaking off a portion of the remaining stick according to υ_k. This construction also gives insight into how the concentration parameter γ controls the relative proportion of the weights β_k. For small values of γ, we expect to see the majority of the mass in the first few weights β_k, so we should observe the same values frequently while drawing samples from G. For large values of γ, we would expect the mass to be more evenly distributed among the weights β_k.

Hence, DP mixture models can be viewed as having a countably infinite number of mixing components, which potentially allows an infinite number of clusters. In other words, consider each ϕ_k as the parameter of the k^{th} mixture component, with mixing proportion given by β_k.

The DPs are used across a wide variety of applications of Bayesian analysis including Bayesian model validation, density estimations and clustering via mixture models and other hierarchical Bayesian models [147]. The most common application of the DP is in clustering data using mixture models, where the non-parametric nature of the Dirichlet process translates to mixture models with a countably infinite number of components. For example, [113] used DP mixture models to cluster genes with similar expression patterns in the analysis of DNA microarray data.

4.1.2.3 Hierarchical Dirichlet Process

In many clustering problems, it is useful to model groups of exchangeable data points jointly, allowing them to share their generative clusters in order to remain linked. An exchangeable sequence of random variables is a sequence of random variables such that for any finite permutation of the indices, the joint probability distribution of the permuted sequence is the same as the original sequence, meaning that any order is equally likely. A inspiring example is the *topic modeling* problem and bag-of-words assumption, where the order of words is ignored and the main goal is to discover topics that are distributions over words, while documents are viewed as



(a) (b)

Figure 4–2: Left: Graphical model of a Hierarchical Dirichlet process, Right: example of a HDP structure for $J = 3$ (reproduced from [147])

distributions across topics. The Hierarchical Dirichlet process, a mixture of DPs in a hierarchically coupled architecture, is a model for groups of data that are assumed to be generated by linked but distinct generative processes.

Suppose there are J groups of observations, each consisting of n_j exchangeable data points $(x_{j1}, \cdots, x_{jn_j})$. We want to model the data points in each observation group with a DP mixture model. While different groups share the same set of mixture components, each mixture has a mixing proportion specific to the group.

In an HDP, the base distribution G for a set of Dirichlet processes is itself drawn from a Dirichlet process. Specifically, the HDP is a distribution over a set of random probability measures, including a global probability measure G_0 and a probability measure G_j for each group j, over the space (Θ, β). The global measure G_0 is distributed as $DP(\gamma, H)$ and each G_j is conditionally independent given G_0, with distribution $G_j \sim DP(\alpha_0, G_0)$:

$$G_0|\gamma, H \sim DP(\gamma, H) \tag{4.11}$$

$$G_j|\alpha, G_0 \sim DP(\alpha_0, G_0) \tag{4.12}$$

Figure 4–2 depicts the graphical representation of the HDP model. According to the stick-break construction, G_0 can be expressed as weighted sum of point masses, which reveals its discrete nature. Since all individual G_j are draws from base distribution G_0, the atoms of each G_j are samples from G_0:

$$G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\phi_k} \tag{4.13}$$

$$G_j = \sum_{k=1}^{\infty} \pi_{jk} \delta_{\phi_k} \tag{4.14}$$

where the posterior ϕ_k is the parameter of the k^{th} mixture component and $\pi_j = (\pi_{jk})_{k=1}^{\infty}$ is the mixing proportion. In next section we discuss how to obtain the posterior π_k.

4.1.2.4 Chinese Restaurant Process

In [147], a Chinese Restaurant Process (CRP) is used to develop inference algorithms for the HDP mixture model based on Gibbs sampling. Intuitively, a CRP is a discrete stochastic process corresponding to seating n_j customers at tables in J Chinese restaurant franchise with unbounded number of shared table/dishes where first customer sits at the first table and the next customer chooses uniformly at random to sit at an occupied table, with probability proportional to the number of customers already there, or at the next unoccupied table. The probability distribution of this random partition leads to a clustering of the values π_k: the dishes

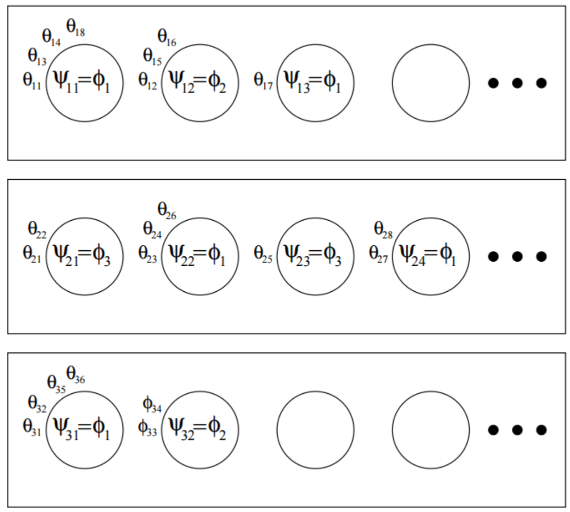


Figure 4–3: Graphical representation of a Chinese restaurant process and Chinese restaurant franchise as presented in [147], p.11. Each rectangle represents a restaurant.

are "common" to all restaurants but the customer seatings are restaurant specific. In the Chinese restaurant franchise, the metaphor of the CRP is extended to allow multiple restaurants which share a set of dishes.

More formally, a Chinese restaurant franchise is defined by the variables $\boldsymbol{t} = (t_{ji})$ (table taken by customer i in restaurant j), $\boldsymbol{k} = (k_{jt})$ (dish ordered by table t in restaurant j) and $\boldsymbol{\phi} = (\phi_k)$ (dish k shared among franchise), given data $\boldsymbol{x}$. Imagine we have J restaurants, each with n_j costumers who sit at tables $\boldsymbol{t}$. Each table is served a dish ϕ_k from a menu common to all restaurants. Figure 4–3 exhibits a representation of the Chinese restaurant process and the clustering property of the DP.

In restaurant level, customer i in restaurant j sat at table t_{ji} where n_{jt} is the number of customers currently at table t. A subsequent customer sits at an occupied table with probability proportional to the number of customers already there, or at the next unoccupied table with probability proportional to α_0 ($G_j \sim DP(\alpha_0, G_0)$). The conditional distributions are:

$$t_{ji}|t_{j1}, \cdots, t_{ji-1}, \alpha_0 \sim \sum_t \frac{n_{jt}}{\sum_{t'} n_{jt'} + \alpha_0}\delta_t + \frac{\alpha_0}{\sum_{t'} n_{jt'} + \alpha_0}\delta_{t^{new}}, \tag{4.15}$$

Performing this process independently for each group (restaurant) j, integrates out all the G_js and provides explicit assignments of costumers to local table partitioning at the restaurant. Each local partitioning contains a parameter copied from some global partitioning, which is indicated by variable ϕ_k. Considering that all local assignments are simply i.i.d. draws from G_0, which is again distributed according to $DP(\gamma, H)$, now the same CRP partitioning process may be applied but in the global level. Similarly, the conditional distributions are:

$$k_{jt}|k_{11}, \cdots, k_{1n_1}, k_{21}, \cdots, k_{jt-1}, \gamma \sim \sum_k \frac{m_k}{\sum_{k'} m_{k'} + \gamma}\delta_k + \frac{\gamma}{\sum_{k'} m_{k'} + \gamma}\delta_{k^{new}}. \tag{4.16}$$

4.1.2.5 Gibbs Sampling in Chinese Restaurant Processes

Here we describe the inference procedure for the HDP mixture model based on Gibbs sampling t, k and ϕ given observation data x. Let $f(.|\phi)$ and h be the density functions for $F(\phi)$ and H respectively, n_{jt}^{-i} be the number of $t_{ji'}$'s equal to t except t_{ji}, and m_k^{-jt} be the number of $k_{j't'}$'s equal to k except k_{jt}. The conditional probability for t_{ji} given the other variables is proportional to the product of a prior and likelihood term. The prior term is given by 4.15 and since this is an exchange process, t_{ji} can be the last one assigned. The likelihood is given by $f(x_{ji}|\phi_{k_{jt}})$, where for $t = t^{new}$, $k_{jt^{new}}$ can be sampled using 4.16, and $\phi_{k^{new}\sim H}$. The distribution then has the form of:

$$p(t_{ji} = t|\boldsymbol{t} \setminus t_{ji}, \boldsymbol{k}, \boldsymbol{\phi}, \boldsymbol{x}) \propto \begin{cases} \alpha_0 f(x_{ji}|\phi_{k_{jt}}) & \text{if } t = t^{new} \\ n_{jt}^{-i} f(x_{ji}|\phi_{k_{jt}}) & \text{if } t \text{ currently used.} \end{cases} \tag{4.17}$$

And similarly, the conditional distributions for k_{jt} and ϕ_k are:

$$p(k_{jt} = k|\boldsymbol{t}, \boldsymbol{k} \setminus k_{jt}, \boldsymbol{\phi}, \boldsymbol{x}) \propto \begin{cases} \gamma \prod_{i:t_{ji}=t} f(x_{ji}|\phi_k) & \text{if } k = k^{new} \\ m_k^{-t} \prod_{i:t_{ji}=t} f(x_{ji}|\phi_k) & \text{if } k \text{ currently used.} \end{cases} \tag{4.18}$$

$$p(\phi_k|\boldsymbol{t}, \boldsymbol{k}, \boldsymbol{\phi} \setminus \phi_k, \boldsymbol{x}) \propto h(\phi_k) \prod_{ji:k_{jt_{ji}}=k} f(x_{ji}|\phi_k) \tag{4.19}$$

where $\phi_{k^{new}} \sim H$. While H is conjugate to $F(.)$ we can integrate out ϕ. This completes the generative process, where G_0 and G_j's are marginalized out.

The HDPs have been widely used in clustering applications, where the number of latent variable in not known or unbounded. A classic example of the HDP application is in topic modeling, where the goal is to project documents into a topic space that facilitates effective document clustering [146]. HDPs have also found uses in the applications beyond clustering. For example, in the problem of speaker diarization, which involves segmenting an audio recording into time intervals associated with individual speakers, the number of the number of true speakers is typically unknown, and may grow as more data is observed. This application seems like a natural fit for the HDP, where the goal is to infer the number of speakers as well as the transitions among speakers. In [62], a hybrid model of HDP-HMM is proposed for speaker diarization, where the HDP is used as a prior distribution on transition matrices of the hidden Markov models.

4.2 Clustering Location Traces with HDP

In this section, we explain how the HDP framework is employed for clustering location history and mobile data. As briefly mentioned earlier, we are interested in high-level analysis of human mobility behaviour. We aim to take advantage of an unsupervised learning algorithm to automatically infer semantic social ties such as similar interests and physical interactions, from their location histories. The key strategy is to discover and characterize the places or points of interest (POIs) frequently visited by each individual, and subsequently build a similarity measure between individuals based on the physical proximity of their POIs. The working hypothesis is that physical proximity has an essential effect on social ties; the probability of social interaction quickly rises with decreasing spatial distance between people. We assume that different social groups will exhibit distinct profiles in terms of the places where they hang out, so individual users can be clustered into behavioural profiles based on their distribution of POIs. In a realistic setup, there are a number of practical challenges that motivate the usage of non-parametric Bayesian clustering framework for clustering location history and mobile data. First, individual users belonging to different behavioural profiles may share a large set of common location points, such as popular places in as city or dining area of a workplace. Second, each individual user may not follow the same mobility pattern for a long-time, and our model should be capable of capturing the diverse characteristics of human

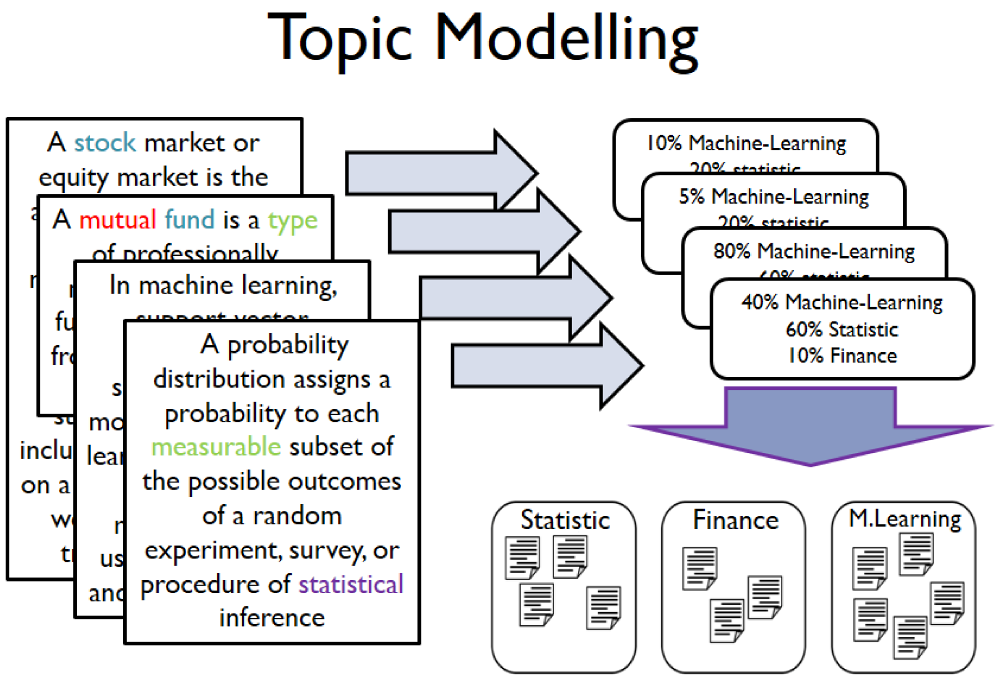


Figure 4–4: Representation of topic modeling approach

movement over a long period of time. Moreover, in many scenarios we may not have *prior* knowledge on the number of behavioural profiles in advance, or new clusters of mobility behaviour may appear over time.

This naturally fits into the hierarchical Dirichlet process setup, which allows sharing of components between clusters (behavior profiles) that are actually probability distributions over all possible location points. While all individuals travel across the same set of locations, each daily trace has its own characteristics due to a specific combination of visit frequencies, and also, the number of identified clusters is automatically adapted to the complexity of the data. Inspired by the topic modeling approach used for text documents, we model the location traces (i.e., "documents")

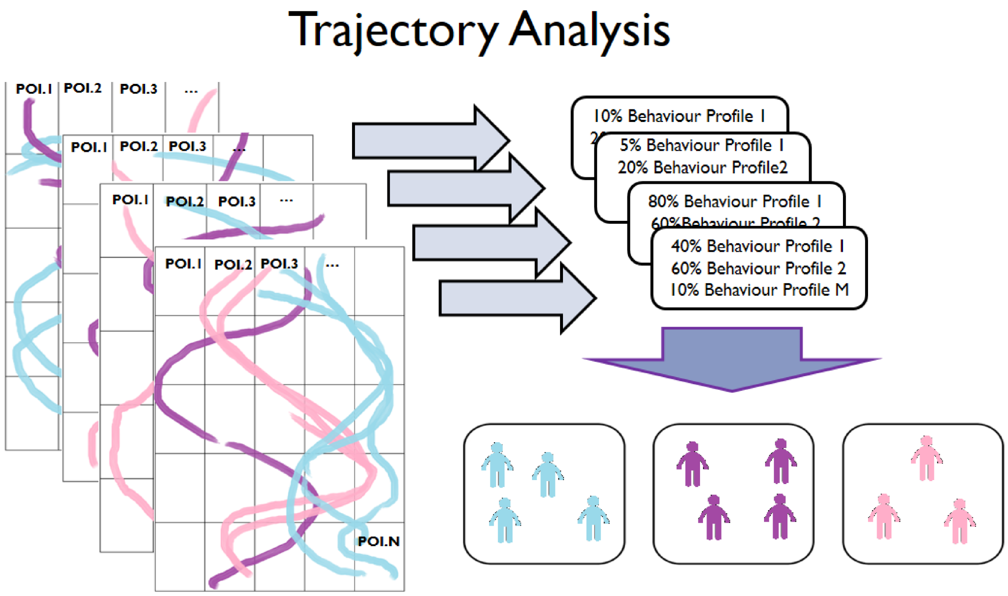


Figure 4–5: Representation of mobility trace clustering using the proposed HDP-based approach

using a Dirichlet process mixture model, in which each mobility behavior profile (i.e., "topic") is a distribution across the user's visited location points (i.e., "words"). Figures 4–4 and 4–5 present intuitive representations of the HDP framework for *topic modeling* as well as the proposed *trajectory analysis* technique.

We are interested in analyzing long and complex human trajectories, in order to discover important locations or places for individuals and groups of users. We believe that for this particular task, the order or temporal sequence of locations is not informative. Hence, we treat each trajectory as an unordered set of locations, rather than a sequence of places, and the "interest" in a particular place is estimated by the number of timestamps associated with the user at a location. This will transform trajectories into a bag-of-word data structure where the *topics* are inferred from word distribution across the document and word order does not play a significant role.

4.2.1 Inferring Places of Interest

Suppose we have J groups of location traces (or observation group), each consisting of n_j exchangeable location points, $L_j = (l_{j1}, l_{j2}, \cdots, l_{jn_j})$, from total N possible locations. These J observation groups represent the mobility records of U individual users during different events, where each user has J_u entries in the data, such that $\sum_{u=1}^{U} J_u = J$.

We propose to model the location points or POIs in each observation group with a Dirichlet process mixture model. Hence, a given location trace can be related to several clusters and is modeled as a sample from a mixture of corresponding clusters. The actual mixing proportions are defined by location counts and the importance of a POI is the total number of timestamps during which the user was at that place.

The HDP defines a conditional distribution over cluster assignments $P(\boldsymbol{c}|\boldsymbol{L})$ where $\boldsymbol{L} = \{L_1, \cdots, L_J\}$ are the location traces and $\boldsymbol{c} = \{c_1, \cdots, c_r\}$ are r assigned clusters. Using Bayes rule,

$$P(\boldsymbol{c}|\boldsymbol{L}) = \frac{P(\boldsymbol{L}|\boldsymbol{c})P(\boldsymbol{c})}{P(\boldsymbol{L})} \tag{4.20}$$

where $P(\boldsymbol{L})$ is computed from the location traces by counting the occurrences of each location and $P(\boldsymbol{c})$ from the clustering of the entire observation groups. The distribution $P(\boldsymbol{L}|\boldsymbol{c})$ is estimated from the visited location frequencies of each trace, as:

$$P(L_j|\boldsymbol{c}) = \sum_{i=1}^{N} P(L_j, l_i|c) = \sum_{i=1}^{N} P(L_j|l_i, c)P(l_i|c), \tag{4.21}$$

where l_i is the ith POI location in L_j. In $P(\boldsymbol{c}|\boldsymbol{L})$, each particular cluster represents a *mobility behaviour profile* among user mobility patterns. The HDP assigns to each mobility trace a distribution over locations, which emphasizes the popular locations for each behaviour profile, and intuitively reveals dominant behaviour profiles for each individual.

4.2.2 Regularization

Due to the nonparametric nature of HDP models, the number of clusters is a random variable whose mean grows at a logarithmic rate with respect to the number of data points. This means at each step of cluster generation, a DP mixture model can either assign a data point to a previously-generated cluster or can start a new cluster. As an advantage of this property, the HDP allows a very fine level of discrimination in mobility patterns, which may lead to the emergence of too many similar clusters that represent the same or slightly different behavior profile. Moreover, the resulting posterior probability distribution from the HDP will rarely provide an explicit correspondence between clusters and location points.

We address these problems by introducing a regularization step on top of the HDP, in order to *a*) measure the distance between derived DP mixture models, in order to prune clusters generated by the HDP that may be too similar; and *b*) compute the similarity between individual users by comparing their posterior distribution of popular location points (POIs).

Depending on the specific data type and the application scenario, many different similarity/divergence measures have been proposed for estimating the statistical distance between two probability distributions. In the following sections we will introduce three different application scenarios, in which the HDP is used to generate behaviour profiles from location history and mobile data. In each working example, we investigate the effectiveness of different metrics for the proposed regularization step, and in each case we choose the distance *score* that best reflects the correlation between clusters and users. In practice, we discovered that distance scores computed based on ℓ_2-norm and KL-Divergence are the most effective approaches for the comparison of DP mixture models.

4.3 Task Inference from Location Data

In this study, we consider the problem of human trajectory analysis, where the mobility traces were collected from employees at a Chicago-area IT facility for one month. Participant employees with different job titles were given related tasks to complete while wearing electronic sensing devices, and were asked to submit a report when they completed their tasks. One important characteristic of this organization is that the tasks are information-sensitive and thus required the employees to discuss their problems with other employees in their group. Hence, fulfilling a task usually requires employees to follow fairly long trajectories. In this naturalistic setting, completing a task was not time-restricted so the length of a trajectory, corresponding to one completed task, would vary from minutes to hours and even days. We aim to perform *task* recognition based on the hierarchical clustering of these complex location trajectories.

Following the strategy described in previous section (see Section 4.2), first, we need to re-form or segment the fine-coarse location trajectories into a sequence of location points and their corresponding counts. Then, the transformed trajectories will be modeled using the HDP mixture model, followed by a regularization step that infers the “task” (e.g. topics) from trajectory clusters.

4.3.1 Reality Commons - Badge Dataset

This dataset was collected using Sociometric Badges (Zigbee) [121] recording the behaviour and interactions of employees at Chicago-area server configuration firm over one month [50]. It includes multiple real and synthetic measures that reflect the performance and dynamics of the organization with different temporal resolutions. The wearable computing platform produces several types of information e.g., received signal strength indicator (RSSI), speech features and 3-axis accelerometer data. In this work, we are mainly interested in time-stamped RSSI readings to anchor nodes with fixed positions, which convey instantaneous locations of employees in the workspace.

More precisely, the dataset documents the work of 23 participating employees in an IT facility that were given computer system tasks on a first-come, first-serve basis. Each employee was asked to take a client's IT configuration requirements and produce IT products according to these specifications, while wearing a badge. In total, 1,900 hours of data were collected with a median of 80 hours per employee. We utilized three main data types.

- The *workspace layout* is depicted in Figure 4–6. Participating employees with different departmental roles are indicated at their booths with different colors.

Figure 4–6: Different branches in the organization have their own specific space. Each participating employee was assigned a unique ID and base stations with unique IDs were placed at fixed locations for RRSI records [50].

The base stations (anchor node), on yellow squares, were placed at fixed positions in order to locate badges and timestamps in their area. The booths indicated with letter "N" belong to employees who did not participate in the study. This floormap and the role indications on it will later be used for the verification of the clustering results.

- The *behaviour data* includes the locations of employees estimated from Zigbee RSSI readings of the badges worn by each employee, representing the fixed location (anchor node) to which they went. The synthetic coordinates for each employee were extracted from raw RSSI readings under two constraints. First, each badge at coordinate (x, y) should see RSSI records from at least three base stations for a particular instance in time. Second, the times of these readings should be at most 1 second apart. In [50] the authors claim that solving the optimization problem of finding the best station and corresponding distance under these two constraints led to the coordinate estimation with standard deviation on the order of the radius of one booth. The coordinate trajectories that we used for our evaluation were computed per employee ID per minute.

- The *performance data* includes the assigning time, closing time, difficulty level (basic, complex, or advanced), assigned-to, closed-by, number of follow-ups and role of the employee (pricing or configuration) of each completed task. In total there were 455 task reports in this dataset. We aim to cluster the location trajectories based on the tasks at each specific time step. Therefore, we focused on the following features: assigning and closing time, closed-by employee ID and the role of employee. Also, for each observation we acquire the corresponding coordinate trajectory from the Instantaneous locations dataset.

Figure 4–7 shows two typical examples of our observations. Although the employees from different jobs spend most of their time visiting their own group's specified region, they regularly pass through other regions and common areas as well (see Figure 4–6 for details of workplace layout). These shared areas make it hard to learn the pattern of the "configuration" and "pricing" tasks. On the other hand, we need to extract location-based features instead of temporal or shape-based features in order to discover the correlation between tasks and visited places. Our framework aims

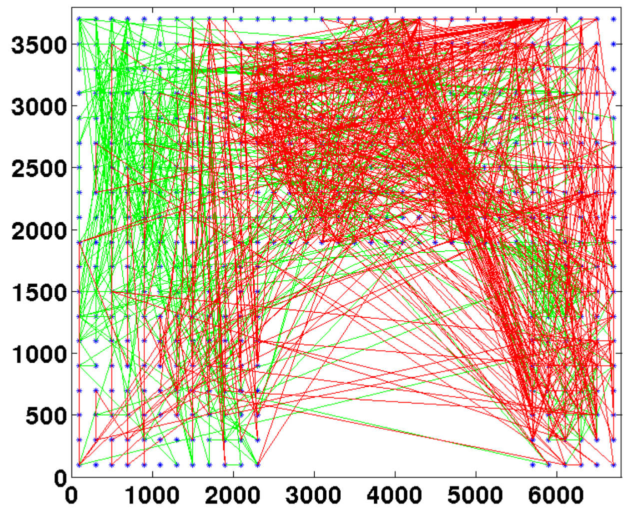


Figure 4–7: Example of complex trajectories. The green and red trajectories are the mobility traces of two different employees performing "configuration" and "pricing" tasks, respectively.

to generate discriminative regional-based features and then carry out hierarchical non-parametric Bayesian clustering to distinguish groups of trajectories.

4.3.2 Trajectory Segmentation

In order to generate segmented trajectories and transform the long, complex traces into sequences of location points, we perform a spatial segmentation on the original 2-dimensional space of the mobility trajectories, which partitions this space into basic regions. We recursively split the original region into *four* identical regions and obtain a grid structure. Then, each trajectory is quantized to a finite number of subregions. For ease of computation, each rectangular small subregion is represented by its centroid point. Ideally, we would want to have homogeneous subregions in the sense that they would mostly contain trajectory parts from the same cluster distribution. However, a good quantization should also concise, which means that

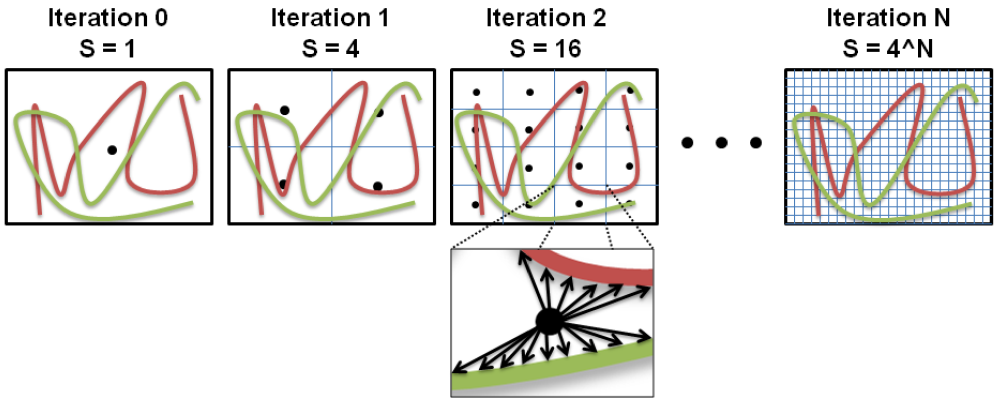


Figure 4–8: Trajectory quantization demonstration.

the number of subregions should stay as small as possible to avoid overfitting and feature space complexity. Thus, it is essential to find a measure that offers a desirable compromise between these two properties. We decided to monitor the "quantization error" (QE) of subregions after each iteration of quad splitting to empirically find the best number of splits. As shown in Figure 4–8, at each iteration of splitting, the average QE of each subregion is calculated by summation of the Euclidean distances of raw trajectory values enclosed in that subregion from the centroid of the region.

More formally, assume we have a set of trajectories $\mathcal{T} = \{T_1, \cdots, T_i\}\,(1 \leq i \leq num_{tra})$, with each trajectory denoted as $T_i = \{(x_1, y_1), \cdots, (x_\ell, y_\ell)\}\,(1 \leq \ell \leq length_i)$. Each quad splitting results in $S(iter) = 4^{iter}\,(itrer \geq 0)$ subregions $(r_1, \cdots, r_S)$ and yields to a quantization of the trajectories into S centroids. The QE of subregion r at each splitting iteration is defined as:

$$QE^2(r) = \sum_{\hat{T}_j \in r} (\hat{T}_j(x, y) - C_r(x, y))^2, \tag{4.22}$$

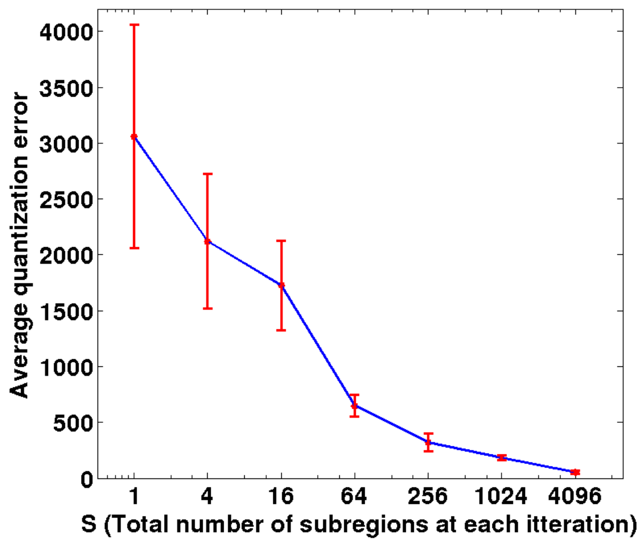


Figure 4–9: Average quantization errors for different number of subregions.

where $\hat{T}_j$ represents all trajectory partitions enclosed in region r that are mapped to the centroid of r (C_r). The average QE of each iteration is then calculated as follows,

$$QE(iter) = \frac{\sum_{r=1}^{S(iter)} QE(r)}{S(iter)}. \tag{4.23}$$

4.3.3 Evaluation

Figure 4–9 demonstrates the QE for six iterations of recursive quad splitting of our working example workplace map. From this figure we can observe that the average QE significantly drops after the 3rd iteration and stays unchanged from there on. Based on this information, we split the workspace to 64 identical regions and therefore, each trajectory is composed of 64 distinct location points (POIs) and the corresponding number of times that trajectory passes from each POI. Note that in this framework, the temporal aspect of the trajectories is not a concern since we only care about the most visited location points by each employee and not the sequence of their motion.

4.3.4 Hierarchical Location Clustering

Prior to the clustering, we generated groups of data from raw data (mobility traces of employees completing their tasks), i.e. each group is a mixture of components (location point) with different mixing proportions (number of passes from POIs) specific to the group. Although in our working example, the true labels of each employee's task and also the total number of task types are known in advance, we want to solve the general in which these parameters are unknown. That is, our goal is to be able to infer that the employees performed task only by observing their mobility pattern. We only use the available annotations from *performance data* to validate our clustering results.

The HDP assigns to each trajectory a distribution over locations which emphasizes the most popular POIs traversed by employees of each branch. As we mentioned earlier, the HDP model does not provide a unique cluster for each mobility trajectory; instead, it computes probability distributions across shared locations. In the language of Bayesian modeling,

$$P(Task|Traj) = \frac{P(Traj|Task)P(Task)}{P(Traj)}, \tag{4.24}$$

where $P(Traj)$ is assumed to be a constant and

$$P(Traj|Task) = \sum_{POI} P(Traj, POI|Task) = \tag{4.25}$$
$$\sum_{POI} P(Traj|POI, Task)P(POI|Task).$$

As expected from the HDP model, the number of clusters grows with the number of mobility traces. Therefore, the popular POIs that appear more often in the data tend to be represented by more clusters. This yields a better level of discrimination in POI distributions. However, the drawback is that we would need a decision boundary across POIs in order to assign a one-to-one correspondence between tasks and trajectories. In practice, to overcome the this problem, a greedy pruning phase is critical in order to achieve meaningful results.

Given the distribution $P(POI|Task)$, the cluster probability given a particular set of location points, we computed pairwise ℓ_2-norm distance to learn similarity and correlation between clusters. The distance score between each pair of distributions in $P(POI|Task)$, is estimated as follows,

$$L2(i, j) = \sum_{m=1}^{M} \|P(POI_m|Task_i) - P(POI_m|Task_j)\|_2^2 \tag{4.26}$$

where $P(POI|Task_i), P(POI|Task_j) \in P(POI|Task)$ and $P_i = \{p(i, 1), \cdots, p_{(i,M)}\}$ represents *task* i with corresponding probability distributions over M POIs. Ideally, we would like to keep clusters with highest $L2$ distances to others, which reveal the underlying structure of *interesting* location points for each branch. This will help to keep the best clusters while pruning spurious ones. These clusters detected by

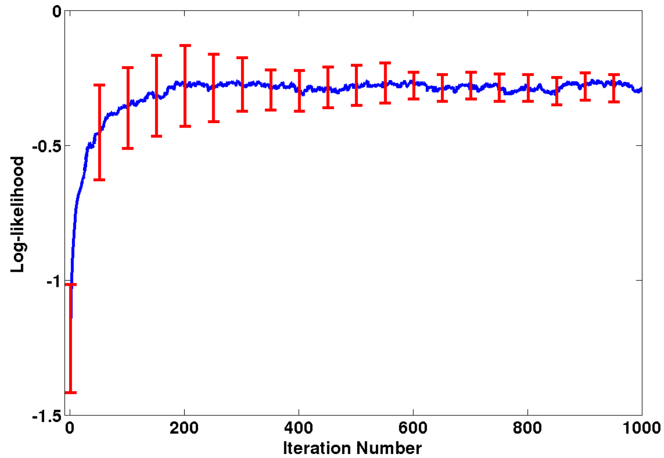


Figure 4–10: Log-likelihood over training data, as the HDP algorithm processes.

the HDP are the *task* labels that we were interested to assign to each trajectory observation.

4.3.5 Evaluation

The original raw trajectory were RSSI recordings mapped into a network of 502 grid points evenly distributed throughout the workspace using the algorithm described in [50]. The trajectories were quantized to 64 subregions based on the experimental results we achieved in Section 4.3.2 and these location points and their corresponding counts were fed into HDP algorithm to produce *task* clusters. We expected the HDP models to learn the mobility patterns that are intuitively correlated to the employees assignments and be able to predict correct labels based on spatial characteristics.

For the experiments, we used the HDP implementation provided by [146]. The HDP requires setting concentration parameters that govern the a priori number of

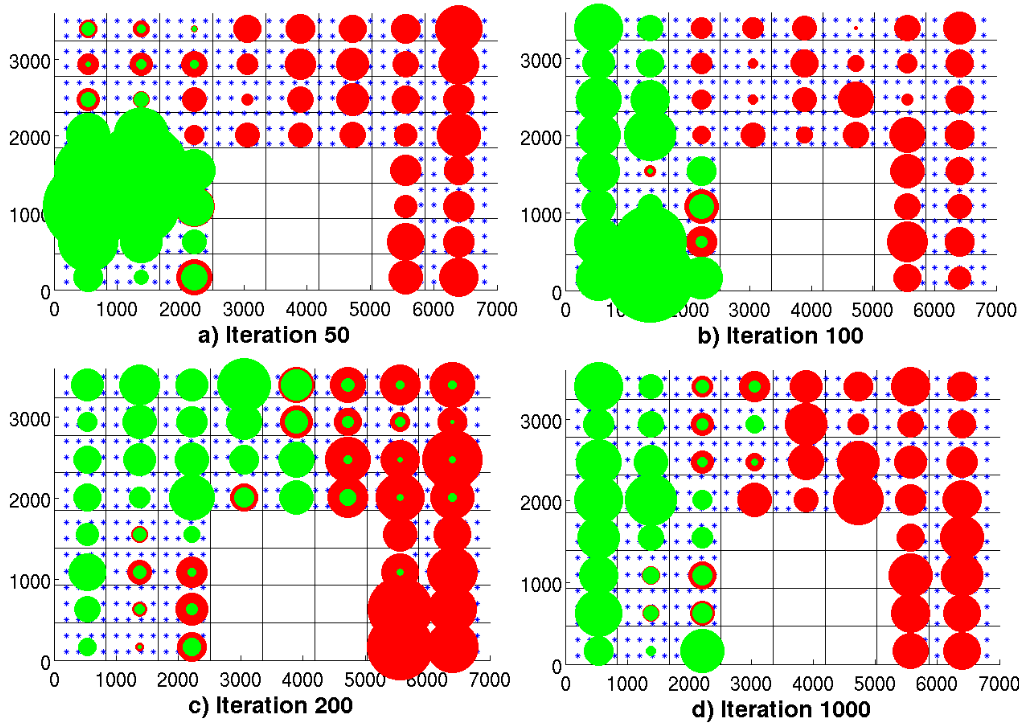


Figure 4–11: Cluster visualization from different iterations of HDP algorithm. The probability distribution over each subregion is represented by a filled circle on its centroid, various colors denote different clusters and the size corresponds to the probability.

clusters, namely, α_0 and α, which we picked by initial exploratory experiments using a small subset of the data. Figure 4–10 presents the average log-likelihood over the training set for $\alpha_0 = 0.2$ and $\alpha = 0.2$ (determined by a line search procedure), as a function of the number of iterations of the HDP algorithm. The algorithm converges quickly and successfully to a solution with good log-likelihood.

As discussed in Section 4.3.4, the HDP does not provide a one-to-one map between trajectories and clusters, hence we used $L2$ score to quantify the correlation between clusters and prune redundant ones. In our experiments, two main independent clusters remain after the pruning step, which had the most distinct probability distribution over the subregion space. Figure 4–11 presents a visualization of the *task* clusters, which reveals a semantic division of trajectories (into “configuration” and “pricing” branches) in comparison with the real layout of the organization. As expected, the HDP allowed sharing location points between clusters. This explains why we have POIs that are assigned to multiple clusters, with different colors. Eventually, we always can pick the dominant cluster (bigger probability) for each subregion.

In addition to visual verification, in order to assess the clustering performance, we used the independent trajectory labels provided in the *Performance data*. Also, we calculated the maximum likelihood of POI assignment by adding up the maximum probability distributions of subregions $P(Trajectory|POI)$ over the winning cluster. The HDP obtained a 72% accuracy and the maximum log-likelihood of subregion assignments was -0.15. The results show that the algorithm mostly fails when the observation sample is extremely long. There were some tasks assignments where more than one working day was needed to complete the task. In this case, the employee would leave and then return to the workplace multiple times during a single task, so many irrelevant places would be repeatedly recorded in the mobility path. The results confirm that the non-parametric hierarchical framework has the ability of offering a high-level viewpoint for learning underlying structure of events, even in the presence of noisy and complex trajectories.

4.4 Learning Social Interaction from Location Data

In the previous section we considered the application of regularized-HDP for fine-grained indoor location trajectory analysis. In this section we use coarse-grained outdoors mobility traces to develop and evaluate our proposed approach. This location history data is borrowed from the MIT Reality Mining dataset [51].

4.4.1 Reality Commons - Reality Mining Dataset

The reality mining project was conducted over the course of nine months from 2004-2005 at the MIT Media Laboratory. This study followed 106 subjects, using their mobile phones as wearable sensors, which continuously collected location information (from cell tower IDs) and Bluetooth device discovery scans (in proximity of approximately five meters). A wide range of other device activities, such as voice calls, text messages, application usage and phone status was also logged by pre-installed software on the phones. 75% of the participants were students and staff at the MIT Media lab, working in the same building; the remaining 25% of the subjects were at the university's business school, adjacent to the Media Lab. The dataset also provides self-reported relational data for individuals; in the survey, subjects were asked about their physical interaction and closeness with others.

The mobility data did not contain the actual geographic coordinates (longitude and latitude) of either the subjects' locations or the cellular towers. Instead, a coarse-grained estimate of the user's location was provided, which was computed from cell tower IDs present in the vicinity of their cell phones. Using a unique tower ID assignment and respective transition timings (timestamps when the phone transitions between cell towers), phone positions were localized to within 100-200

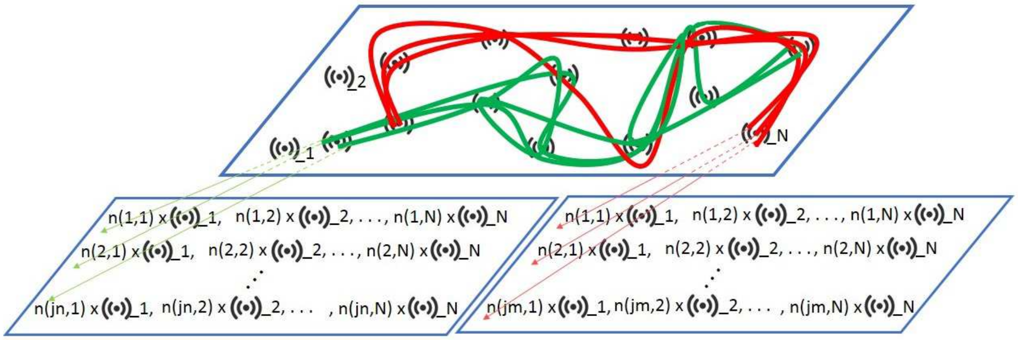


Figure 4–12: Schematic view of spatial features extraction from mobility traces, with every user represented by a different color.

meters. From now on, when we refer to a user's location, we mean the corresponding estimate of the actual location. For each individual, we transform the mobility traces into a set of visited places and the corresponding number of visits (or stays) at these places, over a particular time period (e.g., one day). Figure 4–12 represents the transformation of trajectories into grouped location points or POIs. Similarly to the previous study, we are not interested in the temporal sequence of locations, but in identifying an individual's places of interest and comparing these to those of others.

4.4.2 Dataset Challenges

One of the main challenges of this data is that all of the subjects spend a significant amount of their daytime around a workplace where they share a broad range of cell tower receptions. This could generate erroneous assumptions about their POI similarity and thus, physical interaction level. For example, consider two complete strangers who spend the majority of their time on different floors of a building with similar tower reception distribution. Despite the similarity in their observable locations (visited places), there will be no reported proximity or friendship associated with this pair of users. Therefore, in order to sensibly cluster users into social groups based on their real ties and interests, we need to use a mixture model framework that allows sharing *visited places* across the groups while concentrating on *mixing proportions* of POIs specific to each group.

Another characteristic of the problem is that we intend to discover potential social ties in the data, but there is no *prior* knowledge on the number of social groups. In fact, we want to train a mixture model that allows the model size to grow as new or unseen POIs appear in the mobility traces.

4.4.3 User-specific Modeling and Global Modeling

The procedure of computing the posterior probability distributions from trajectories using the HDP model is similar to what we have presented in Equations 4.20 and 4.21. Following the notation from Section 4.2.1, let $P(\boldsymbol{c}|\boldsymbol{L})$ denotes the conditional distribution over cluster assignments, where $\boldsymbol{L}$ are location traces and $\boldsymbol{c}$ are r assigned clusters. Also, remember that each cluster in $P(\boldsymbol{L}|\boldsymbol{c})$ represents a *behaviour profile* discovered by the HDP mixture model.

In this section, we consider two approaches to building the models. The first idea is to build a global model, which requires that we have access to a global database that contains the entire mobility traces from every user. Equation (4.21) shows the computation of the global $P(\boldsymbol{c}|\boldsymbol{L_j})$ distribution built from all J mobility traces. Afterwards, user-specific weights have to be computed to determine the contribution of each behavior profile in clustering the users into social groups.

However, there are many applications in which the location of the user cannot be shared in one common database, mostly due to privacy preferences. Therefore, in such a situation we would not have the mobility traces of all users and, hence, we would not be able to learn global similarity functions. Rather, we would need to build multiple user-specific models, only from the mobility traces of each individual, and then dispatch the model for comparison with others. In this case, a separate HDP model is computed for each individual, where clusters represent multiple patterns in the individual's behavior profile. The clusters are still shaped over all possible locations and the probability distribution of POIs highlights the importance of different locations for that particular user. Similarly, the individual distribution $P(\boldsymbol{L_u}|\boldsymbol{c_u})$ for user u is estimated from mobility traces, $\boldsymbol{L_u} = \{L_{u1}, \cdots, L_{uJ_u}\}$, as:

$$P(\boldsymbol{L_{uj_u}}|\boldsymbol{c_u}) = \sum_{i=1}^{N} P(L_{uj_u}|l_i, c_u)P(l_i|c_u). \tag{4.27}$$

In this scenario, a distance measure is needed in order to be able to compare the POI distributions between individual user models.

4.4.4 KL-Divergence based Regularization

So far, we have described how a Dirichlet mixture model can be used to model the mobility traces of a network of users and explained two approaches for computing the probability distribution of important places associated with individual or group behavior. Now, we focus on deriving the proximity of users base on their common location interests. In the application presented in the previous section, the $L2$ regularization was introduced for pruning clusters that are too similar and also for estimating a distance score between users base on the distribution of their POIs. In this section, we use this distance measure as a regularizer which controls the excessive growth of HDP clusters.

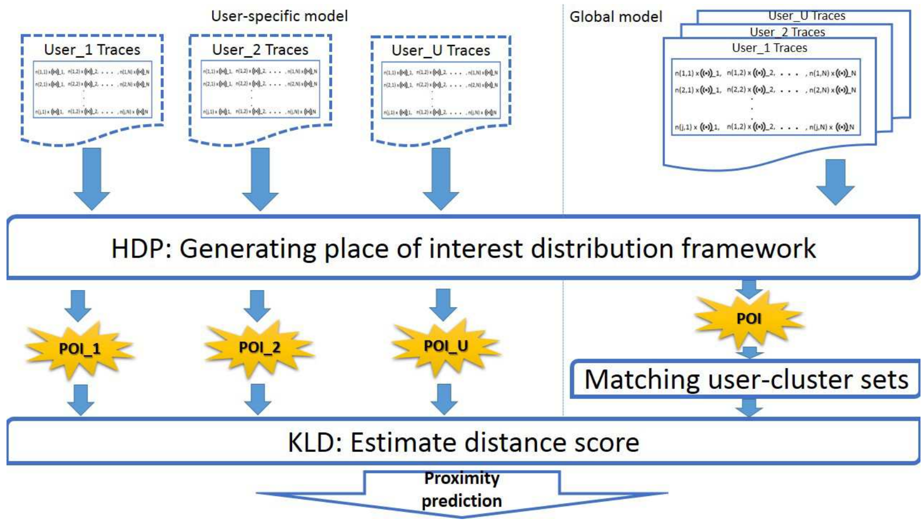


Figure 4–13: Architecture of the proximity prediction algorithm. The left and right side demonstrate user-specific modeling and global modeling, respectively

In the user-specific setting, as a side effect, this will also allows us to compute pairwise distance scores between individual mixture models. In the global setting, we can compute both pairwise distances between users, as well as pairwise distances between POI distributions within the global model, in order to prune excessive clusters. An overview of the proposed system is given in Figure 4–13.

4.4.4.1 Kullback-Leibler Divergence

We will use the KL divergence, also known as the *relative entropy*, as a standard measure for the difference between two probability distributions. The KL divergence for two probability distributions P and Q is defined as:

$$D(P||Q) = \int_{-\infty}^{\infty} P(x) \log \frac{P(x)}{Q(x)} dx \approx \sum_i P_i \log \frac{P_i}{Q_i}. \tag{4.28}$$

The KL divergence between two mixture models is not analytically tractable and Monte Carlo methods are needed to compute it approximately [74]. We recall the approximation presented in [74] for estimating the KL-divergence between two Gaussian mixtures and consider a similar approximation method to compute the distance between two Dirichlet mixtures derived from the HDP. Given n samples of the variable of interest, the approximation to the KL divergence is given as follows. Let X denote a random variable. We draw samples x_i from the distribution $P(X)$. Then, the KL approximation is given by:

$$D_n(P, Q) = \frac{1}{n} \sum_i \log(P(x_i)/Q(x_i)). \tag{4.29}$$

Since, the KL divergence is not a symmetric measure, in order to use it as a distance metric we compute the symmetrized version [128]:

$$SD(P, Q) = \frac{1}{2} \left(D(P||Q) + D(Q||P) \right). \tag{4.30}$$

The idea of using KL divergence for regularization is that we would like to eliminate from the HDP clusters that are too similar to each other. We note that one could simply penalize the log-likelihood of the data by a function based on the estimated KL divergence of the clusters (e.g., the minimum such value over cluster pairs). However, we choose to simply discard clusters, because this keeps the complexity of the HDP and associated inference process much smaller, which is an important practical consideration.

4.4.4.2 Distance-based Proximity Scores

Our application involves computing pairwise distances between Dirichlet mixture models, so we can use the same idea outlined above for this purpose.

In the user-specific setting, given the U individual probability distribution $\boldsymbol{P_u} = [P(l|c_1), P(l|c_2), \cdots, P(l|c_u)]$, where l denotes a possible location for user u, the distance score between each pair of users is:

$$D_{US}(m, n) = SD(P(l|c_n), P(l|c_m)) \tag{4.31}$$

where $n, m \in [1, 2, \cdots, u]$ and c_u are clusters over possible locations. Since the clusters in the user-specific setting reflect the behavior profile of individuals, $D_{US}(m, n)$ reveals the dis-similarity in mobility behavior between users m and n.

In the global setting, given the distribution $P(\boldsymbol{l}|\boldsymbol{c})$ with g clusters across the set of all possible locations $\boldsymbol{l}$, we compute the distance score between users m and n as:

$$D_G(m, n) = SD(P(l|g_n), P(l|g_m)) \tag{4.32}$$

where g_u is a subset of global clusters with at least one non-zero probability distribution over the locations visited by u.

A dissimilarity measure between two location distributions is computed as a symmetrized KL:

$$D(i, j) = \sum_{k=1}^{N} P(l_k|g_i) \log \frac{P(l_k|g_i)}{P(l_k|g_j)} + \sum_{k=1}^{N} P(l_k|g_j) \log \frac{P(l_k|g_j)}{P(l_k|g_i)} \tag{4.33}$$

where $i, j \in [1, 2, \cdots, g]$ and g_i denote global clusters.

4.4.4.3 Illustration on Synthetic Data

In order to illustrate the effect of the KL regularization, we use first synthetic data, for which experimental results are presented in Figure 4–14. We generated a large synthetic dataset in the scenario of grouped data clustering, including 10,000

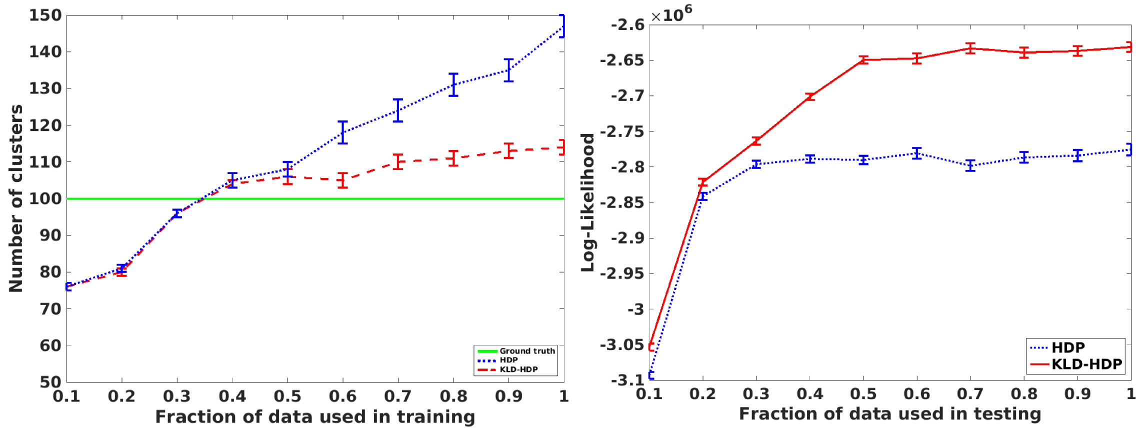


Figure 4–14: Performance comparison between HDP and KLD-HDP on synthetic data from plain mixture of Poisson model

groups of data, each consisting of 500 points, collected from a mixture of Poisson models with a known and fixed number of clusters (100 clusters) and 200 unique exchangeable data points. We used the HDP algorithm to discover the number of clusters using a portion of the data of pre-specified size (between 10% and 100%), running it for 1000 iterations.

Then, we computed D using equation (4.33), and heuristically merged the clusters with pairwise distance $D < 1$. Figure 4–14 demonstrates that the number of clusters discovered by the HDP grows with the size of training data, and keeps growing even when all the real clusters are found. The regularization step merges excessive clusters that helps limiting unnecessary growth in the total number of clusters. The results are compared against the true number of existing clusters in each portion of the data indicated by the green line. The right graph in the Figure 4–14 show the number of clusters of the HDP and regularized HDP, with error bars generated over 10 independent runs. As expected, regularization helps to keep the size of the HDP in check. Without regularization, the number of clusters grows excessively. Moreover, the left graph in the Figure 4–14 illustrates that the log likelihood on text data is improved with regularization, because this helps to avoid overfitting.

4.4.5 Evaluation

In this section, we evaluate the performance of the proposed approaches on the Reality Mining dataset.

4.4.5.1 Reality Mining data

We used a subset of the data collected from 94 subjects that had completed the survey about their physical interaction with others subjects. In total, we had approximately 450,000 hours of information of mobility history, from which about 11,000 days of mobility traces were extracted over all the users. The user location at each timestamp was mapped onto one of the 1028 unique cell tower IDs, based on the signal reception of their mobile phone, and the transition times were provided. This allows computing the amount of time spent, or "interest", for each location per day. The true labels for evaluating the unsupervised model come from a self-reported survey which asked to estimate the physical interaction or proximity (within 3 meters) with other subjects of the study in that work place. The proximity level was chosen from a range between 0-5 where 0 means no proximity, and 1-4 means being close to someone at least 5 min, 10-30 min, 30 min-2 h, 2-4 h and 4 h per day, respectively. In practice, we found that two individuals need to physically spend a significant amount of time (at least 2 hours per day) together, to be able to detect any correlation in their mobility behaviour and thus social interests. Therefore, we transformed the proximity labels into a binary form, where more than 2 hours

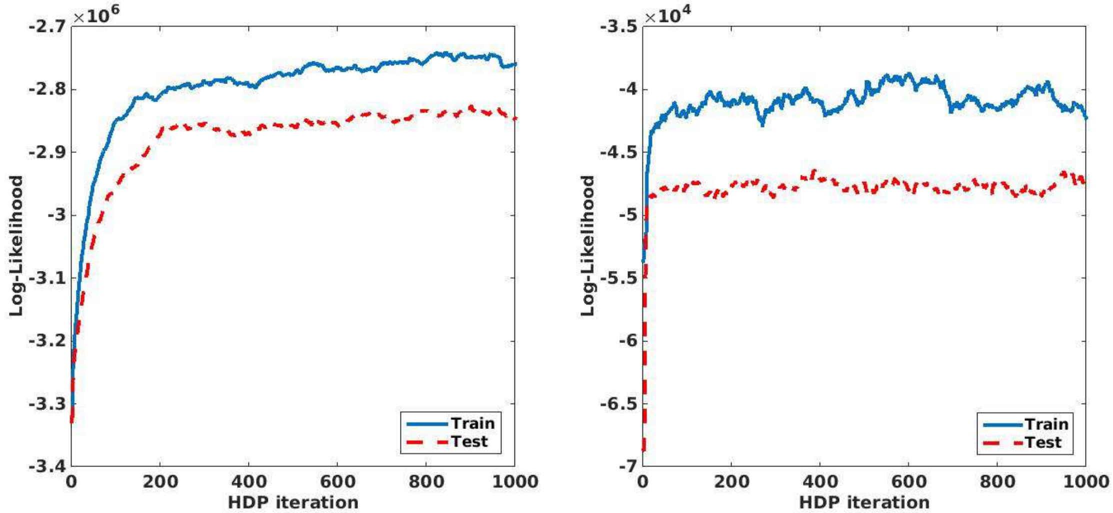


Figure 4–15: Average log-likelihood over training and test data as a function of the number of HDP iterations, in a) the global model and b) the User-specific model.

of proximity per day is considered as 1, and no proximity or less than 2 hours is considered as 0. This also removes some of the subjectivity inherent in the self-reported data. The distribution of the two categories is highly imbalanced, since only 15% of all possible pairwise combination of individuals ended up classified as "close" according to this definition.

4.4.5.2 Proximity prediction from locations

In both the user-specific and the global model setting, we used the HDP implementation provided by [147] with a symmetric Dirichlet distribution with precision parameter of 0.5 for the prior over POI distributions. The posteriors are integrated out using Gibbs sampling for 1000 iterations. In the user-specific setting, one HDP model was built for each individual from their daily traces, which varied from 8-277 days, the mean being 123 days. In this setting, the number of detected clusters for

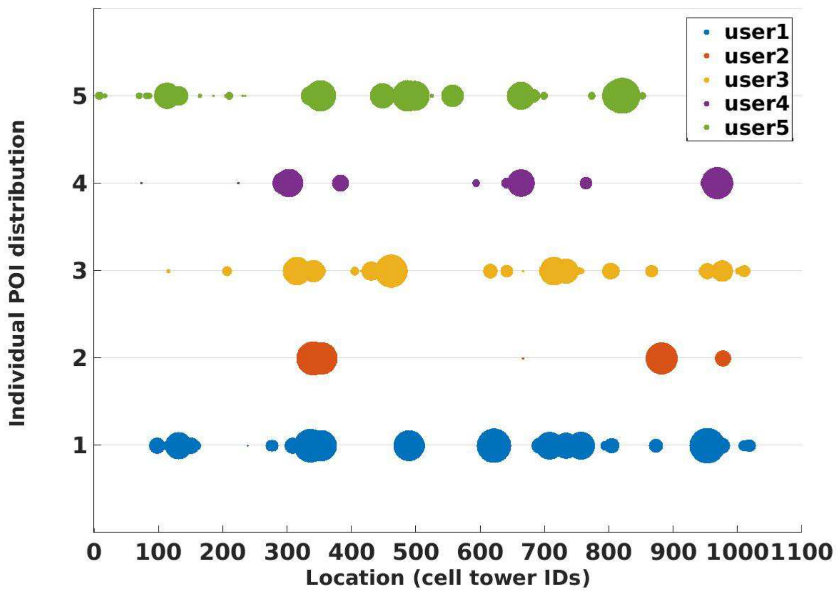


Figure 4–16: Visualization of the HDP clustering. Each horizontal line with different color represents the probability distribution of user's *interest* over the locations. The radius of each bubble is proportional to the probability assigned to the particular user for that location.

each individual varied from 3-17 clusters, with a mean of 6 clusters. In the global setting, one HDP model was built from all 11,000 daily traces, and the model discovered 182 clusters. 10-fold cross validation was used to validate the performance of the HDP model. Figure 4–15 depicts the likelihood over HDP iterations.

Intuitively, clusters in each user-specific model demonstrate the particular behavior of that individual. Figure 4–16 shows an example of the HDP cluster assignments for 5 different individuals. As one can see, the results are quite varied. Now, in order to infer proximity between two individuals, we compared their mixture of POI distributions using the D_{US} distance score described in Equation(4.31). We then

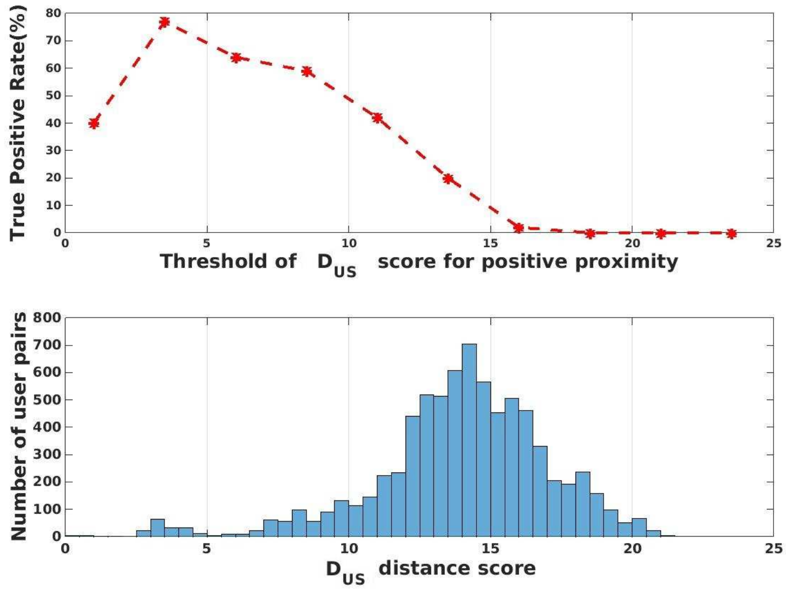


Figure 4–17: Bottom: D_{US} over all possible pairwise user model comparisons. Top: Positive proximity was decided based on setting a threshold on the D_{US} score. The best true positive rate, about 78% is achieved when setting the threshold to 3.

applied a empirical threshold to make a binary decision on the distance between each pair of users. Figure 4–17 gives the histogram of D_{US} distance scores, and shows how the threshold was chosen. We picked a threshold of ϵ as this gave the best true positive rate (TPR) values.

In the global setting, 182 clusters corresponding to general mobility behavior profiles were discovered from all 94 users. A large number of clusters appears in the posterior due to the larger number of data points, and thus, a higher number of redundant clusters occurs in this model. Prior to finding correspondences among individuals, we computed the pairwise KL distance between these clusters in order to

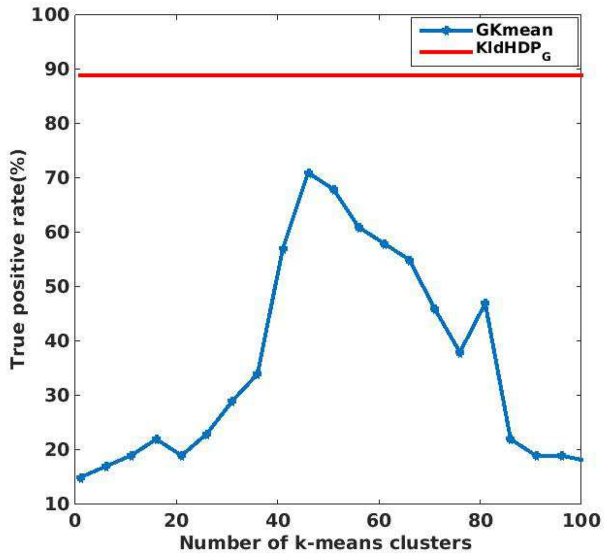


Figure 4–18: The proposed global model with the automatically inferred number of clusters vs. k-means with different numbers of clusters. Averaging the results over 5 runs, k-means exhibits the best performance using 45 clusters. The proposed approach constantly outperforms this baseline in terms of true positive rate.

prune excessive ones. Here, the threshold was heuristically set to merge the clusters with distance < 2, which resulted in 125 clusters.

Afterwards, we computed D_G using Equation (4.32). This score reflects the level of proximity for each pair of users; by applying a threshold, a binary decision of whether two users are "close" or not can be made. In this setting, the best true positive rate of 89% was obtained at a threshold of 4 for D_G.

4.4.5.3 Baseline models

A natural strategy for inferring behavioral profiles from mobility traces is to look at the probability distribution of the most frequently visited places (or POIs). Therefore, we considered extracting a feature set, referred to as *most frequented* *place (MFP)*, by computing the probability distribution of visited locations (i.e., by averaging the amount of time spent at each place) from the traces. Despite its simplicity, this is a very intuitive and strong baseline that has been widely used in the literature to interpret semantic location information from trajectories [172, 41].

The first baseline (denoted USk-means), which is comparable to the user-specific setting, KLD-HDP$_{\mathrm{US}}$, extracts the MFP for each user and employs standard k-means clustering to discover POIs for their traces. Then, the similarity between users was established by counting the number of POIs in common. The second baseline (Gk-means), which is comparable to the global setting KLD-HDP$_{\mathrm{G}}$, extracts the MFP from each trace, then uses the D-score computed from equation (4.33), as a distance function for k-means, in order to cluster the mobility traces into groups. The proximity between two individuals is then decided based on the number of daily traces assigned to clusters in common.

Our algorithm is adaptive, in the sense that the number of clusters is a variable inferred from the mobility patterns of individuals, based on the desired similarity threshold, while in k-means, like in many other clustering algorithms, the number of clusters to be discovered is fixed in advance. We varied the number of clusters in Gk-means from 2-100, while the HDP selected the model size automatically. Figure 4–18 compares the KLD-HDP$_{\mathrm{G}}$ and Gk-means models in terms of correctly predicted positive proximity, averaged over 5 runs.

Finally, in Table 4–1, the TPR, F-score (due to imbalance distribution of clusters) and accuracy (ACC) for all proximity predictors are shown. In general, the

	Global models		User-specific models	
	Gk-means	KLD-HDP$_{\mathrm{G}}$	USk-means	KLD-HDP$_{\mathrm{US}}$
TPR(%)	71	**89**	65	**78**
F-score(%)	77	**90**	74	**83**
ACC(%)	82	**92**	81	**88**

Table 4–1: Comparison of different strategies to infer proximity

difference between the true labels and the predicted ones might come from the differences in the concept of proximity among people. However, the proposed KLD-HDP outperforms the baselines in both settings.

4.5 Mobile Communication Data Analysis

In this section, we present the last application example for the HDP-based approach to analyzing location data. This application scenario has a fairly different anatomy compared with the other applications studied in this chapter, where outdoor and indoor location trajectories were used to analyze the mobility behaviour of a group of users. In contrast, here we introduce a new application of HDP-based clustering, in which mobile communication traffic collected from cell tower antennae is the focus of the analysis. We will use the same approaches in the previous studies for clustering communication traffic for a group of users. In this case, the clusters are distributions of calls across cell tower antennae ("points of importance"), and we aim to find interesting locations corresponding to each cluster.

4.5.1 D4D Challenge Dataset

The *Data for Development* (D4D) dataset [2, 25] was gathered by the Orange Group (a large telecommunications operator) in 2011 to contribute to the socio-economic development and well-being of the Ivory Coast population. This created

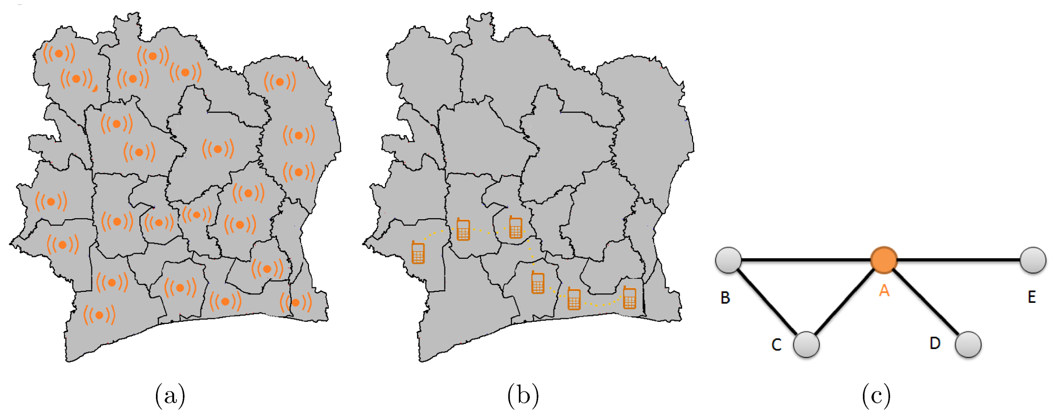


Figure 4–19: illustration of (a) pairwise communication between cell towers, (b) a schematic example of individual trajectory data, and (c) an example of communication sub-graph

a large mobile phone dataset which was made available to scientific researches for a limited time in order to come up, in an open-ended fashion, with ideas to help reduce poverty and to improve living conditions and social development. The D4D data was collected during 5 months from the mobile phone calls of five million customers and resulted in 2.5 billion anonymized calls and text messages records. The released dataset included 3 different types of data and over 80 research teams around the world participated in the challenge by submitting research projects developed using these data. The types of data were as follows:

- **Antenna to Antenna:** The number and duration of calls between each pair of 1196 cell towers for a given hour
- **Individual Trajectory:** The location trajectory of 50,000 randomly selected customers at two different resolutions;
    1. Short term & high resolution: mobility trajectories were shortened to two weeks and locations were identified by 1196 cell tower positions

2. Long term & low resolution: mobility trajectories were collected for five months and locations were identified through 255 administrative regions

- **Communication Sub-graphs:** Communication graphs of 5,000 randomly sampled customers, with first and second order neighborhood for period of two weeks. The first order neighborhood includes all the people each user had communication with, and the second oder neighborhood includes those people's first order neighborhood.

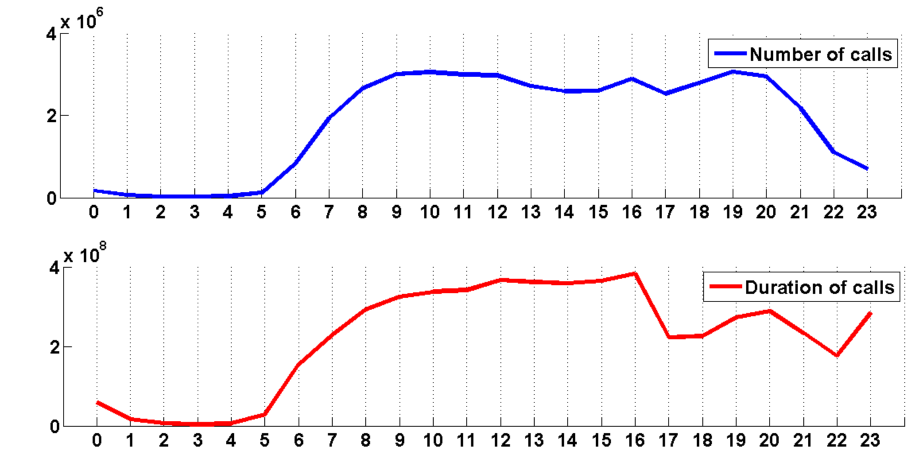
These data types are schematically depicted in Figure 4–19.

4.5.2 Mining Call Patterns

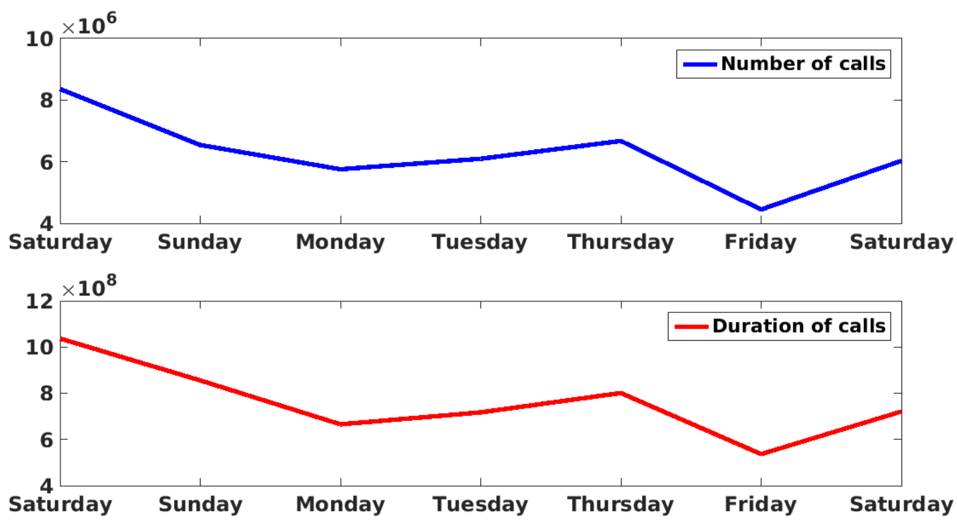
The focus of our work is on the first dataset, antenna to antenna records, where the cell towers were uniquely identified by an antenna ID and their geographic coordinates were provided by longitude and latitude. As described in the D4D data report [25], for technical reasons, some antenna identifiers are not always available. The corresponding communications, which were assigned to the code -1, were removed from our computations due to their potentially misleading effects.

The preliminary observations show that almost every day the number of calls decreases dramatically from midnight until 5:00 am, so from now on we call this duration *nighttime*, and the rest of the day, i.e. from 6:00 to 23:00, *daytime*. Figure 4–20 depicts the distribution of calls during different hours of the day and different days of the week.

The goal of our study is to discover correlations in the call traffic based on both the geographic location of the cell towers and the time of the calls. We employed the HDP model to compute the probability distribution of call traffic between towers,



(a)



(b)

Figure 4–20: Hourly and daily distributions of call records

depending on the time, and to identify "important" locations using the entropy of the distribution. Intuitively, the locations of interest should depend greatly on time of day. For instance, users are more likely to be at home during nighttime and at work during daytime; therefore, the corresponding detected locations are probably residential areas (suburbs) and workplaces, respectively.

Each observation in the data contains information about the originating and terminating antenna ID, as well as the total number of calls between these towers. We aim to compute the probability distribution of calls over each antenna to be able to highlight important clusters or locations for each time step. However the number of clusters are not known a priori and tends to grow as more data points are seen. Further, each antenna is paired with various other antennae at each time step, which means that the clusters we compute have to share components.

This problem formulation fits naturally with the HDP and topic modeling framework. We consider each observation from a pair of antennae to correspond to a *document*, and the *words* are the originating and terminating antennae IDs, while the number of calls corresponds to the word counts.

The HDP assigns to each observation a distribution over antenna locations which emphasizes the most popular locations of the call receivers. However, for each hour of a day, the HDP computes clusters consisting of several call patterns spread over locations. To address this problem, we empirically set a cut-off threshhold to truncate very low probability occurrences for calls.

4.5.3 Evaluation

One of the challenging issues facing big data analysis is how to efficiently subsample the existing data. Using the entire dataset yields significant computational costs, and sub-sampling can lead to imperfect solutions, by ignoring some of the data.

Therefore, we suggested two different strategies for sampling this dataset, which lead to different effects on clustering results.

The entire dataset includes call traffic of 1196 cell towers that could vary within a wide range, from 0 to 200 received calls per minute. In the first attempt, we uniformly sampled observations at each time step, which resulted in a fair distribution of observations with both high and low number of received calls. However, our goal was to look for places of interest that essentially should be identified by observations including high volume of communication traffic. Therefore, the uniform sampling method would reduce the contrast between important places (cell towers) and the rest of the location points. Thus, in the second experiment, we collected all the antennae with high volume of calls (above a certain threshold proportional to the total number of calls recorded in the data for each time step) and then drew uniformly at random from the rest of the observations. This approach worked significantly better, so we present results based on the latter experiment.

Figure 4–21 presents the average log-likelihood over the training set and over an independent test set (from a different week), as a function of the number of iterations of the algorithm. As seen, the algorithm converges quickly and successfully to solutions which have good log-likelihood. We noted that the range of location probabilities in all clusters falls into two categories. First the "important" clusters with probabilities > 0.7, and second the "trivial" clusters with probabilities < 0.2. Figures 4–22 to 4–25 present visualization of the "important" clusters obtained for different hours of a day. In each figure, different colors of the bubbles indicate

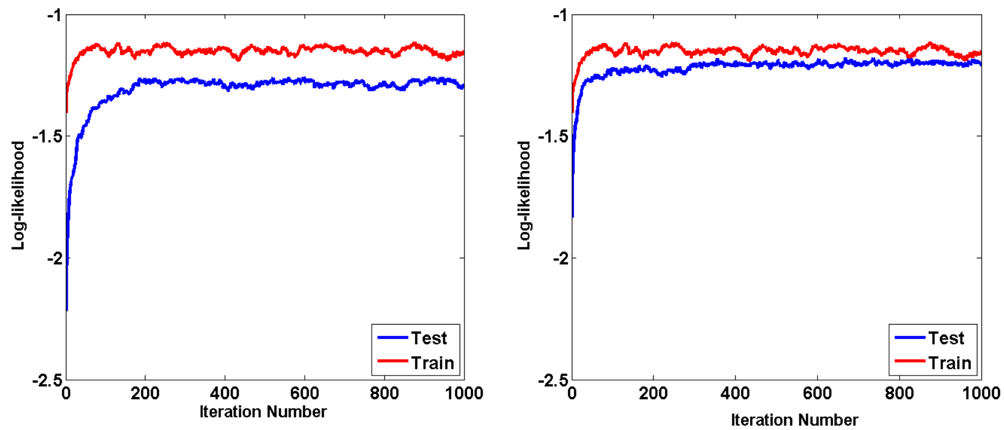


Figure 4–21: Log-likelihood over training and testing data, as the algorithm progresses. Learning is quick and converges to good solutions

different clusters appeared at that hour, and the radius of each bubble is proportional to the probability assigned to the particular cluster for that location.

As expected, the call patterns and geographic distribution of important cell towers dramatically vary during 24 hours. Daytimes show significantly more calls, with more clusters and some very focused clusters. At night, there are significantly fewer clusters, and they are more spread out. It is worth noting that some areas are only visited at certain times during the daytime that can indicate the industrial areas. Unfortunately, this dataset did not include any annotation about the users, the call traffic or the location distributions, therefore, we were not able to take advantage of ground truth labels to validate the inferred interpretations.

This experiment was conducted as a pilot study of the proposed clustering method, during the limited time of D4D challenge, and the preliminary results have

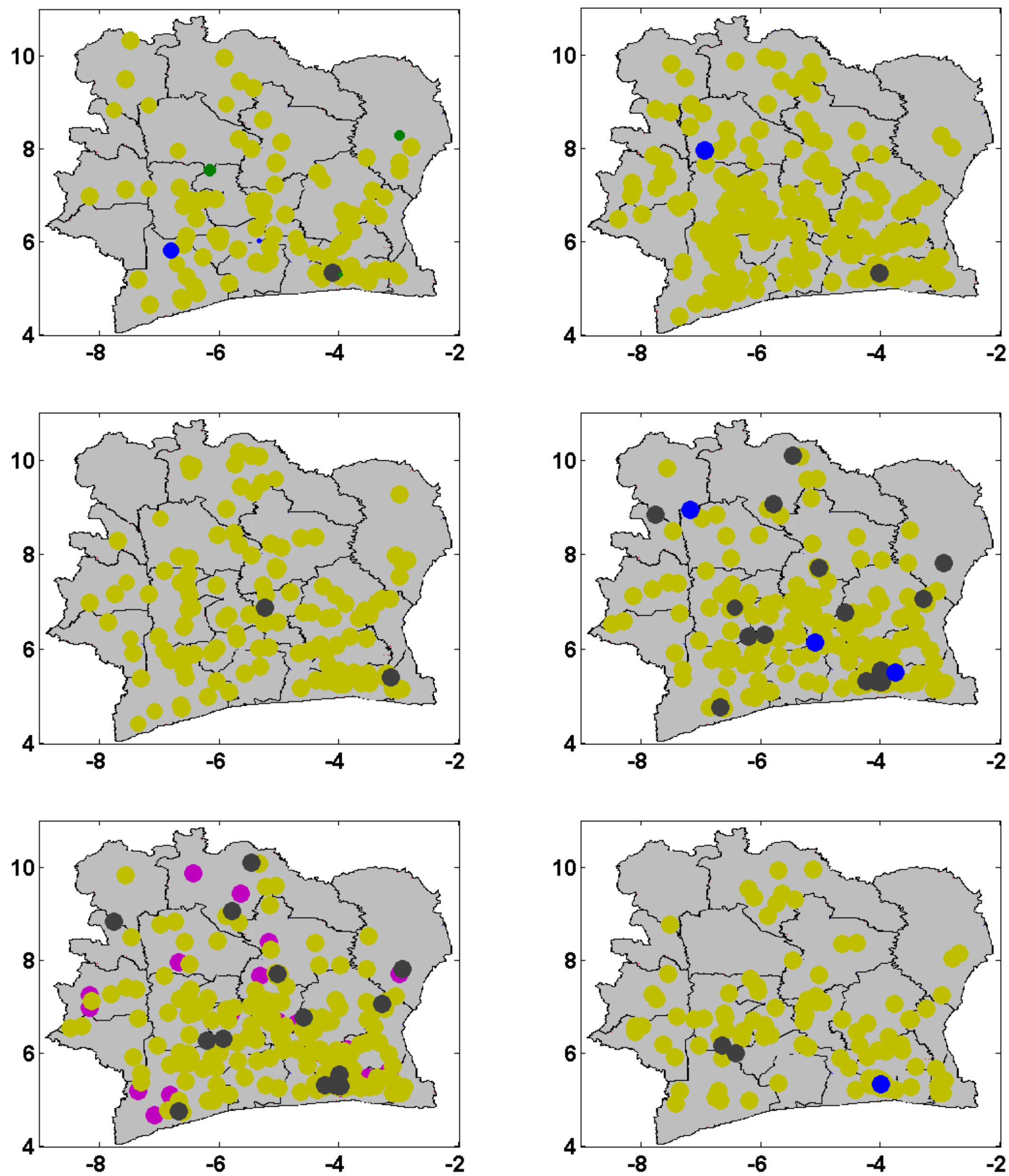


Figure 4–22: Visualization of antenna clusters (learned based on call numbers) for nighttime, from 12am to 5am, respectively

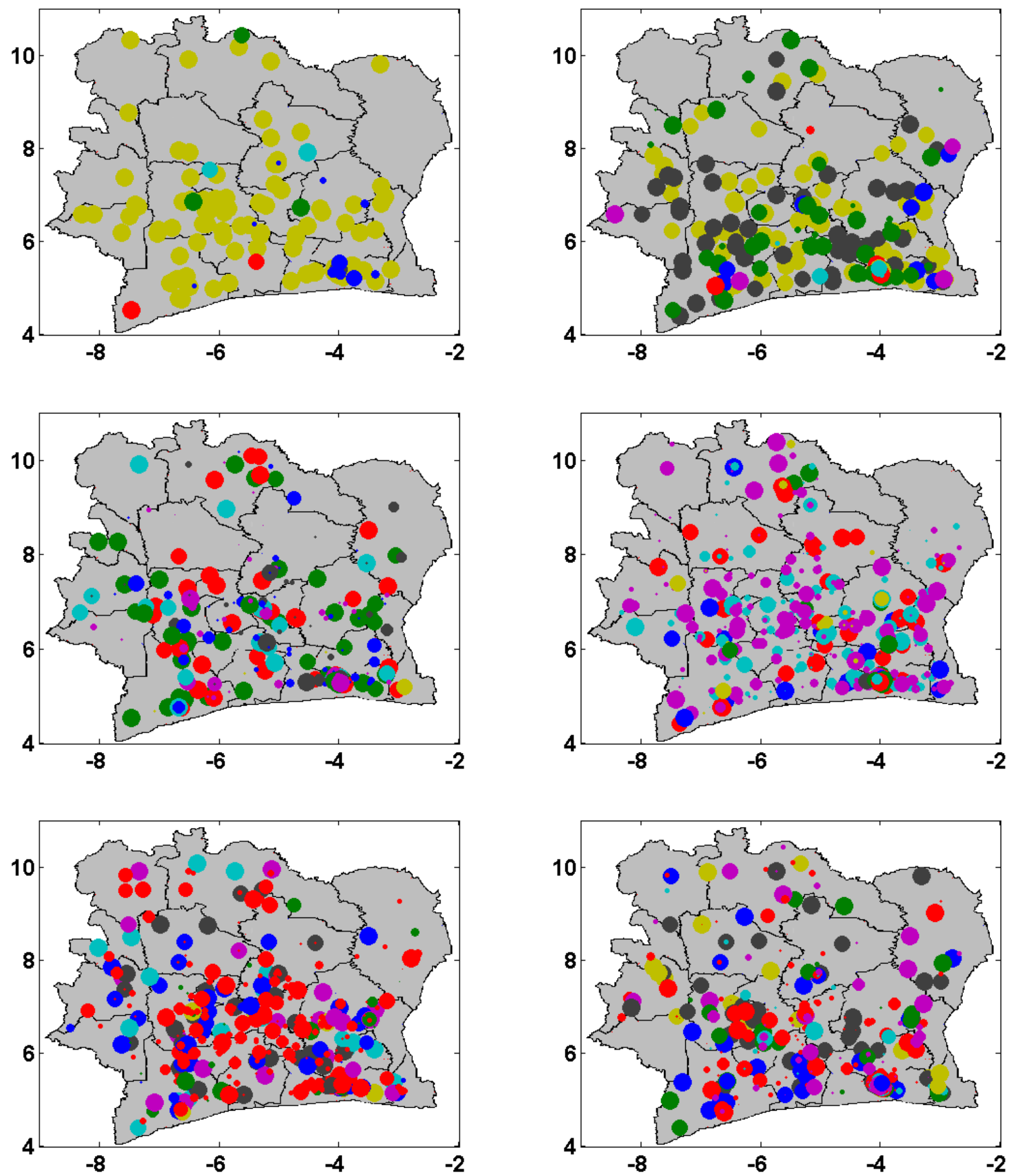


Figure 4–23: Visualization of antenna clusters (learned based on call numbers) for daytime, from 6am to 11am, respectively

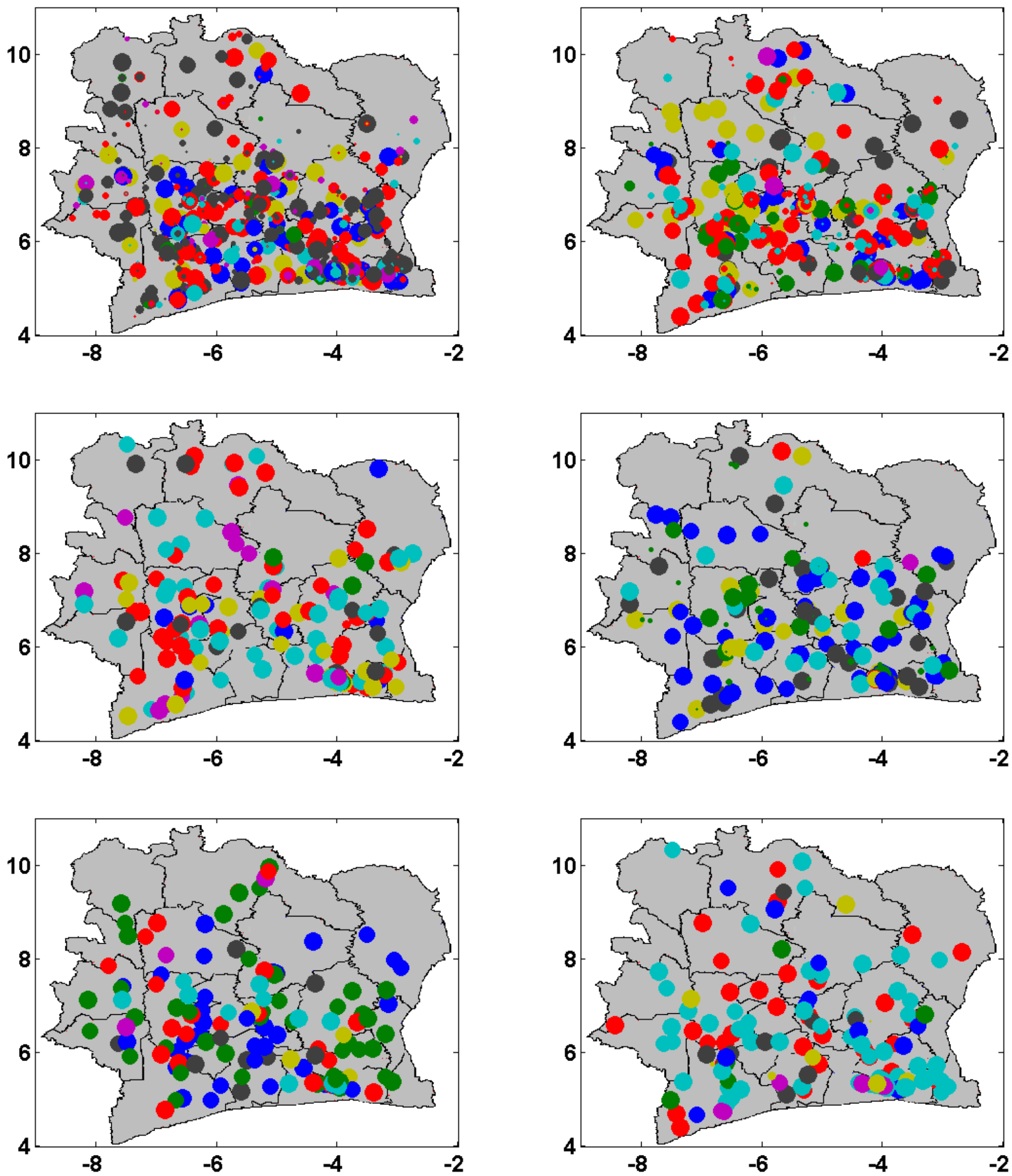


Figure 4–24: Visualization of antenna clusters (learned based on call numbers) for daytime, from 12pm to 5pm, respectively

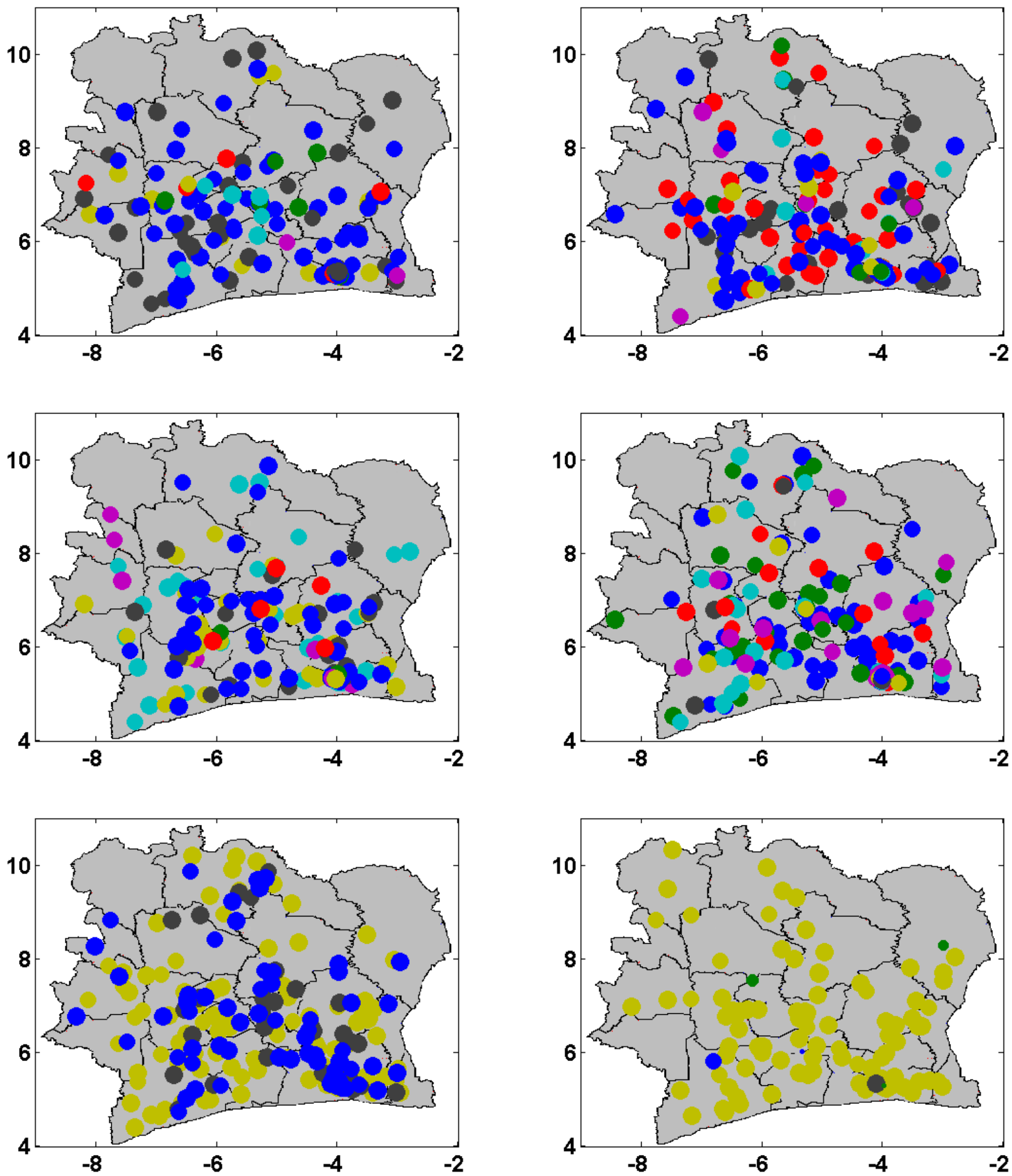


Figure 4–25: Visualization of antenna clusters (learned based on call numbers) for daytime, from 6pm to 11pm, respectively

the potential to be used by authorities to qualitatively study the mobility and movement of people across the country during the day. For instance, this information could be useful for planning new developments such as roads or neighborhoods. We intended to improve the results by exploring ways of regularization beyond thresholding, however, after the challenge period, the dataset was not available for further analysis.

4.6 Conclusion

One essential role of unsupervised learning in activity recognition problems is to discover routines and regularities in human daily behaviour with minimum requirements of annotation and prior knowledge about the structure of data. In this chapter, we proposed an HDP-based clustering method for analyzing location data obtained from mobile devices over extended periods of time, in order to glean information about the long-term behaviour patterns of different groups of users in multiple application scenarios.

The proposed method included two layers of learning steps; the hierarchical clustering layer that models the mobility behaviours, and, the regularization step on top of the HDP clustering, to keep the size of models in check. While a lot of the work on HDPs assumes that the strength of the prior will be sufficient to control the model size, our regularization approach provides a tighter control over this aspect. One interesting future work direction would be providing a theoretical explanation of the effect of the KL divergence used on the model from a Bayesian perspective. We anticipate that the proposed methodology would have a positive impact on other cases in which the possible model complexity is in fact bounded, but we still want it to grow as more data become available.

CHAPTER 5
Wifi-based Activity Recognition

In an effort to extend our academic knowledge of activity recognition and human data analysis beyond theory and in order to engage in solving real-world challenges, we participated in collaboration with ©TandemLaunch inc. for a project on "Activity recognition using physical layer information of wireless communications" in the context of home automation, in which we had to design and implement practical solutions for different aspects of the project.

In this chapter, we introduce the relatively new problem of wifi-based signal analysis for device-free activity recognition, which proposes using wireless transceivers in indoor environments as a sensing infrastructure to obtain information about users' behaviour. As seen in previous chapters, classical activity recognition methods often rely on wearable sensors, mobile devices or environmental sensors to infer human activities or behavioural models. Although these approaches are widely accepted and have been shown to be effective in traditional studies, they raise some practical concerns such as deployment cost, privacy issues and energy consumption, when it comes to long-term monitoring of real world conditions or public applications with many users and over a long period of time.

Alternatively, there is a very recent research area that focuses on activity recognition by employing off-the-shelf wifi-enabled devices, e.g., access points, laptops, smart TVs. This is mainly motivated by wireless technology improvements and the fact that wifi signals are pervasive in daily life at home, work and even public places. The key idea is to monitor the influence of human body movements and gestures on the changes in the strength and pattern of wireless communications between the transmitter and receiver [153]. Studies suggest that information gleaned from the physical layer in wireless infrastructures (e.g. wifi signals), such as channel state information (CSI) and received signal strength indicator (RSSI) values has the potential to characterize the environment, which includes both ambient objects and human movements and gestures.

The current design and implementation of this novel technology exhibits some limitations due to the complexity of the wireless signal propagation in indoor environments and due to the challenging nature of human behaviour itself. Therefore, in this chapter, we present an extended study on human activity recognition, using multiple supervised and unsupervised techniques, for the design and implementation of an smart indoor space applications. The evaluation and experimental results are all performed on real-data collected and processed at ©TandemLaunch.

First, in Section 5.1, a brief review of existing studies in radio-based activity recognition is provided, and then necessary background material is presented. Then, we describe different aspects of the subjective problem of designing and modeling device-free smart indoor environments using wifi signals in Section 5.2. In Section 5.3 we evaluate our learning strategies and approaches on real data collected in several application scenarios, and finally in Section 5.4, we conclude the experimental results with a discussion on future directions for this work.

5.1 Background

Although this research topic is still in its initial development stage, several research groups have developed activity recognition frameworks based on the analysis of different types of radio data. In this section we first introduce and review some related work on radio data mining and technical and practical issues that arise when working with wireless signals. This is followed by a review of wireless physical layer infrastructures, namely Channel State Information (CSI), as the foundation of our proposed recognition models in this chapter. Finally, we briefly review the Latent Dirichlet Allocation algorithm that is used in our design for location identification.

5.1.1 Radio-frequency Data Mining

In recent years, an interest in evaluating technologies for activity identification through device-free approaches has emerged, since they do not require people to carry any devices, which is an attractive property for the industry. There are various technologies and approaches for obtaining radio data (including Zigbee, wifi, RFID, microwave, FM signals, etc.) with different characteristics and processing steps. The common idea among all these techniques is that they try to use radio-frequency sensing infrastructure to study the influence of different human activities on the covered sensing environment.

One class of approaches focuses on monitoring the Doppler shifts and multi-path distortions of wifi signals originated by human physical activities or capturing radio reflections bounced off the human body, in order to detect and classify different movements and gestures in the environment [5, 6, 7, 130]. For example, *WiSee* presented in [130] exploits the Doppler shift in narrow bands extracted from wideband OFDM (orthogonal frequency division modulation) transmissions to recognize nine different human gestures. Also, technologies introduced in [6, 7] (*WiTrack*, *WiTrack2.0*) measure the time it takes for a customized signal to travel from its transmitter to the reflecting body, called Time of flight (TOF), using frequency-modulated continuous wave transmission (FMCW) and then use this information to track and recognize user movements. However, all these approaches need specific hardware for the transmission radar.

Another class of approaches employ low-power, low-cost wireless network standards, such as Zigbee and RFID tags, to sense the user's environment and observe the effect of different human activities on the wireless communication patterns between sensor nodes and the base station. For example, *RF-IDraw* [152] introduces a virtual touch screen system that use multi-resolution positioning to trace the trajectory shape of RFID tag placed on a user's finger in order to let them input characters by drawing in the air. These techniques can provide high recognition accuracy without the burden of high system deployment costs. However, they are not considered a device-free system, because some specific hardware is still needed.

In contrast, we are interested in a research area that solely utilizes off-the-shelf wifi-enabled devices for sensing the environment. These approaches are mainly motivated by the wireless technology improvements introduced in the IEEE 802.11n and IEEE 802.11ac standards, and also the market penetration expected for chipsets in compliance with these standards, powered by the Internet-of-Things (IoT). Intuitively, these new techniques not only assume that there is no need for the user to carrying any devices, but also intend to perform sensing entirely by using off-the-shelf devices. For instance, *WiGest*, a gesture detection technology introduced in [4] performs in-air hand gesture recognition similar to *WiSee*, but captures the hand movements using off-the-shelf wifi devices [4].

Another inspiring example is the technology used in *E-eyes* [155], which offers an activity recognition system based on physical layer measurements of the communication links between transmitter-receiver, provided by IEEE 802.11n and 802.11ac standards respectively. E-eyes collects fine-grained channel state information (CSI), which allows location-oriented identification of in-place activities (such as watching TV on a sofa, washing dishes, etc.) and tracking of walking activities by detecting different paths and people passing through doorways. The results suggest that CSI, which is the channel response at the receiver in the frequency domain, is sensitive to environment influences, and there is a consistent relationship between CSI variations and the movement of people. A detailed discussion of CSI is given in Section 5.1.2. Many researchers have recently began to leverage this new technology in different application areas, such as indoor localization [154, 160], motion recognition [166] and crowd-counting [159]. However, all of the current studies on this novel technology have been performed under controlled conditions and often exhibit some limitations, due to the complexity of the wireless signal propagation in indoor environment and their high sensitivity towards surrounding variations. We aim to follow this trend and build an activity recognition system in the context of intelligent indoor environments, and propose practical solutions to address these limitations.

On the other hand, most of these radio-based activity recognition studies [130, 5, 6, 159, 4] leveraged a two-step heuristic approach to infer activities and gestures from the data, where the raw data is simply segmented into smaller pieces and then these segments are matched based on their similarities. The segmentation phase manually or automatically determines the boundaries of performed activities and then the identification or matching step maps the bounded segments to different activity families. This naive approach seems to work well in practice under the assumption that for all of the intended activities, explicit template models exist. However, in real-world application scenarios the template matching solution becomes very challenging, because users are free to perform new, unknown activities or different users may perform the same activity differently. In this case, more sophisticated techniques, namely, machine learning algorithms, are needed to train a general model for various activities.

Only a few previous works have been employed learning algorithms (supervised or unsupervised) to train human activity models from Radio frequency data or wifi signals. For example, authors in [157] have used sparse classification to build a device-free location-oriented activity recognition system for four different activities from CSI measures. In [156], 13 different activities are learned from CSI measurements of wifi Multi-input Multi-output (MIMO) radio data using a Kernel SVM-based classification algorithm. Finally, *E-eye* [155] applies a template matching algorithm to build activity profiles for clustering 9 in-place activities and 8 walking activities. They apply adaptive strategies to address the problem of unknown activities, in which the user is only prompted to label an activity when significant differences are detected from the existing profiles. However, the wifi signals can be affected by signal interference, other ambient movements and any changes in the environment, therefore all of the existing designs and implementations need to be initially considered in very naive experimental setups e.g., a single occupant in a one-bedroom apartment. We believe that extensive research is still required to demonstrate the importance of machine learning techniques for overcoming the limitations of activity recognition from wireless measurements.

5.1.2 Channel State Information

In wireless communications, radio signals propagate between transmitter (Tx) and receiver (Rx) through several transmission channels. The channel properties of a communication link can be mathematically modeled based on the transmitted and received signals, as well as the disturbance effect of every object in the environment, such as reflections, fading, diffraction and scattering effects. In fact, the received signal is the result of the interference of several multi-path signals transmitted through the surrounding objects and all other disturbances events. Therefore, the characteristics of these communication channels are highly correlated to environmental variations.

This motivates a quantitative study of signal propagation behavior within a wifi-covered area to measure and evaluate different types of disturbances within the environment. The major challenge here is to statistically formulate the correlation between environmental events and communication channel properties. One important example of the collected measurement regarding channel properties, which forms the basis of our study, is the Channel State Information (CSI) values.

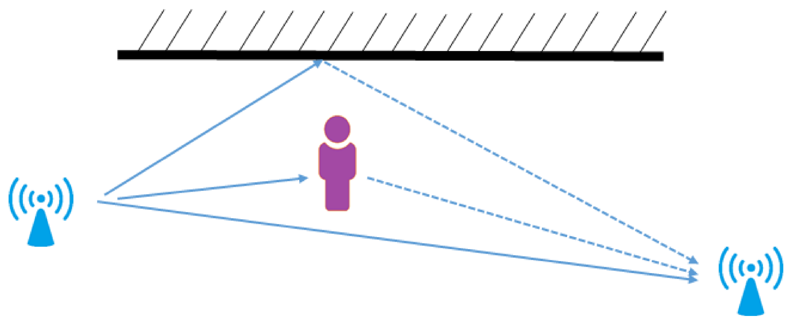


Figure 5–1: Schematic representation of multipath effect

In the frequency domain, consider a narrow-band and flat-fading channel with potentially multiple streams between transmitter and receiver antennae (MIMO), which can be modeled as,

$$\boldsymbol{y}_{(s,l)} = \boldsymbol{x}_{(s,l)}\boldsymbol{H}_{(s,l)} + \boldsymbol{n}_{(s,l)} \tag{5.1}$$

where, $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{n}$ and $\boldsymbol{H}$ denote the transmitted and received vectors, noise and the channel matrix for all streams, respectively.

The CSI measures are the estimated values of matrix $\boldsymbol{H}$, which is the channel response and the transformation of subcarrier $\boldsymbol{s} \in \{1, \cdots, S\}$ in stream $\boldsymbol{l} \in \{1, \cdots, L\}$ at time t. At each time-stamp, the pilot signal $\boldsymbol{x(t)}$ is transmitted, through L streams on S subcarriers in the frequency domain, and the signal $\boldsymbol{y(t)}$ is received. The CSI feedback received at the access point (AP) is a complex-number matrix, including magnitude and phase information, which shows the channel frequency response (CFR) of each individual subcarrier for all spatial streams. Figure 5–1 illustrates how the objects and people's presence influence the signal propagation. The CSI takes advantage of multipath effects and captures detailed changes on different subcarriers.

Beside application areas such as indoor localization and motion detection (mentioned earlier in 5.1.1), CSI measurements have been used as the sensing technology in many other human-centric systems for high-level and low-level activity recognition. For example in [165] the authors have leveraged the CSI information to analyze shopper behaviour and browsing patterns. The study in [10] claims to perform keystroke extraction on a keyboard from CSI measurements.

However, the raw CSI values are very noisy signals, since they can be drastically affected by many parameters such as signal interference, user movements, ambient movement and changing environments. Therefore, a system that aims to extract general patterns from the CSI measures has to differentiate the desired activities and movements from unwanted interferences in the measurements. Another challenge of analyzing this type of data is that the CSI readings are dependent on the location of the wifi devices, as well as the distance between the target user or object and the directional antenna. This property can be considered as an advantage when it comes to the localization problem and tracking the position of moving objects, or when distinguishing between stationary or in-place (e.g. washing dishes, cooking, brushing teeth) activities and displacing movements (e.g. walking and running). However, it creates an issue for the recognition of activities, for which the location of users changes during the activity, such as *walking* and *running*. This means, for example, the wifi devices capture different patterns when the same person walks for different distances or to different locations with respect to the access points. Although these challenges impose some restrictions and limitations on the user activity recognition from these type of data, we aim to propose practical solutions to address these problems. These solutions include noise filtering, multiple feature extraction techniques including both location-dependent and independent features and machine learning techniques for building a robust device-free system for smart indoor spaces.

5.1.3 Latent Dirichlet Allocation

In this section we briefly present latent Dirichlet allocation (LDA), [24], which is an unsupervised learning approach for clustering large collections of grouped data such as text documents. Similarly to the HDP (see section 4.1.2 for more details), LDA was originally proposed for topic modeling in text corpora by discovering latent variables or semantic topics in text documents. In fact, the HDP can be seen as a non-parametric extension of LDA, where the number of mixture component or the *topics* in document-modeling terms is unknown a priori.

Let's assume we have a collection of M *document*s denoted by $\boldsymbol{D} = \{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_M\}$, where each document is a sequence of N *word*s denoted by $\boldsymbol{w} = \{w_1, \cdots, w_N\}$, with w_n being the nth word in the sequence. A word w is the basic unit of discrete data, which is an item from a vocabulary indexed by $1, \cdots, V$, and a *topic* $z \in \{1, \cdots, K\}$ is a probability distribution over the vocabulary of V words. If w takes on the ith element in the vocabulary, then $w^i = 1$ and $w^j = 0$ for all $i \neq j$. The generative process of topic modeling assumes that each word within a document is generated by its own topic, and hence $\boldsymbol{z} = \{z_1, \cdots, z_N\}$ denotes the sequence of topics across all words in a document. The process for generating each document indexed by $m \in \{1, \cdots, M\}$ is as follows:

- Choose a K-dimensional topic weight vector θ_m from the Dirichlet distribution for parameter α, $p(\theta|\alpha) = Dir(\alpha)$.
- For each word indexed by $n \in \{1, \cdots, N\}$ in the document:
  - Choose a topic $z_n \in \{1, \cdots, K\}$ from the multinomial distribution $p(z_n = k|\theta_m) = \theta_m^k$.
  - Choose a word w_n from $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

In this process, the dimensionality k of the Dirichlet distribution is assumed known and fixed over all of the documents. The LDA model allows documents to contain multiple topics with different proportions. The word probabilities are parametrized by a $k \times V$ matrix β, where $\beta_{ij} = p(w^j = 1|z^i = 1)$, which is also assumed to be estimated from data and encodes each of the K topics as a distribution over V words.

A k-dimensional Dirichlet random variable θ can take values in the $(k-1)$-simplex (a k-vector θ lies in the $(k-1)$-simplex if $\theta > 0$, $\sum_{i=1}^{k} \theta_i = 1$) and has the probability density on this simplex as:

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1}, \tag{5.2}$$

where $\Gamma(x)$ is the Gamma function and the parameter α is a k-vector with components $\alpha_i > 0$.

The Dirichlet distribution has a number of nice properties that facilitate the development of inference and parameter estimation algorithms for LDA [24, 77]; it is in the exponential family, has finite dimensional sufficient statistics, and is a conjugate prior to the multinomial distribution.

Given the parameters α and β, the generative process given above defines the joint distribution of a topic mixture θ and a set of N words $\boldsymbol{w}$ as:

$$p(\theta, \boldsymbol{z}, \boldsymbol{w}|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta), \tag{5.3}$$

where the term $p(z_n|\theta)$ is simply $p(\theta_i)$ for the unique i such that $z_n^i = 1$. The central task for LDA is to determine the posterior distribution of the latent topic variables conditioned on the words that we observe in each document. From Bayes rule we have:

$$p(\theta, \boldsymbol{z}|\boldsymbol{w}, \alpha, \beta) = \frac{p(\theta, \boldsymbol{z}, \boldsymbol{w}|\alpha, \beta)}{p(\boldsymbol{w}|\alpha, \beta)}. \tag{5.4}$$

Integrating over θ and summing over $\boldsymbol{z}$, the marginal distribution, or likelihood, of a document is given by:

$$p(\boldsymbol{w}|\alpha, \beta) = \int p(\theta|\alpha)(\prod_{n=1}^{N} \sum_{z_n} p(z_n|\theta)p(w_n|z_n, \beta))d\theta. \tag{5.5}$$

Taking the product of the marginal probabilities of single documents, we obtain the probability of a corpus as:

$$p(\boldsymbol{D}|\alpha, \beta) = \sum_{d=1}^{M} \int p(\theta_d|\alpha)(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d)p(w_{dn}|z_{dn}, \beta))d\theta_d. \tag{5.6}$$

Figure 5–2 shows the graphical model representation of LDA.

Although LDA was originally proposed for text processing [68], in recent years the model has been extended to a wide range of applications in other domains such as, video analysis [120] and object recognition [57]. Another example is the study in [77], where LDA is employed to symbolic music files for automatic harmonic analysis.

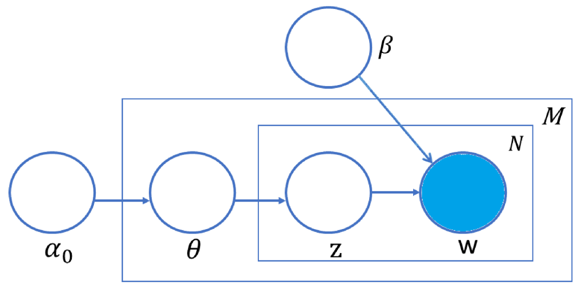


Figure 5–2: Graphical model of the LDA (reproduced from [24])

Or, the study in [133], in which the multinomial distribution in LDA is replaced by a set of Gaussian distributions and a set of Poisson distributions to better handle counts of stochastic events in the problem of behavior analysis of elderly people for assisted living applications.

5.2 Intelligent Indoor Spaces

Long-term automated monitoring of residential or small industrial properties is an interesting topic in the context of activity recognition. In this section we propose a device-free wifi-based activity recognition system developed for smart indoor spaces. First, we try to give some intuition for using CSI measurements as a sensing technology in the design of the smart indoor spaces. The main motivation is that these measures can be collected from any existing wifi-enabled devices, which are already available in most indoor areas. Moreover, this sensing infrastructure can operate through walls, non-intrusively, in privacy preserving manner and at a low cost.

The design of a smart indoor space usually consists of three major steps; data collection from sensing technology, data analysis and decision making. Figure 5–3 outlines the architecture of the proposed activity recognition system. The sensing

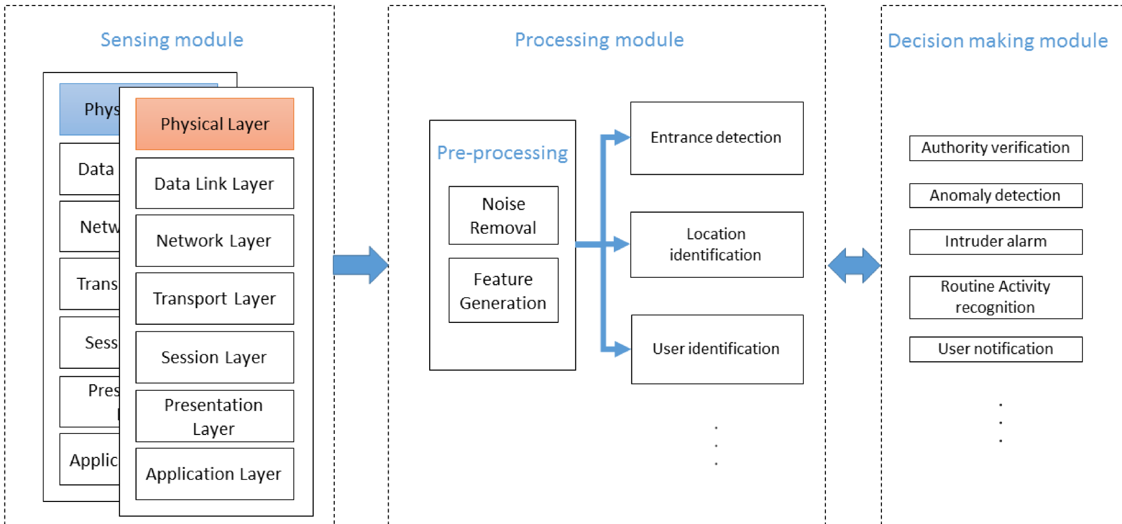


Figure 5–3: General overview of the proposed device-free smart indoor system

module in our application can be any (pair of) existing off-the-shelf wifi devices such as smartTV, wifi access points, laptops and routers, in an indoor area such as apartments, offices, and large open areas. This pair of wireless transmitter/receiver are continuously communicating through multiple communication channels, which creates wireless network coverage within the sensing area. The OSI (Open Systems Interconnection) model [49] characterizes and standardizes the internal functions of a communication system by partitioning it into abstraction layers such as physical layer (e.g CSI and RSSI information) and data link layer (e.g. MAC addresses). The left panel of Figure 5–3 depicts the OSI model.

The CSI information is constantly transferred to the data analysis module to understand and learn the correlation between wifi signal dynamics and human activities and moving patterns. Depending on the overall goal of the system and the data type, a variety of learning tasks can be performed in the data analysis step, including but not limited to: daily activity recognition, anomaly detection, behaviour modeling, user identification, moving object tracking, presence detection and location identification. There are various applications that can be considered in the process of designing a smart space, including

- Security and surveillance [102], e.g. intruder alarms
- Real-time, long-term health monitoring [144], e.g. assisted living for elderly people or patients with disabilities
- Control of building equipment [72], e.g lighting, energy management, heating and ventilation
- Smart human-machine interaction systems [46], e.g. entertainment, smart kitchen and smart TV room

We are particularly interested in creating task-specific learning approaches that can be utilized by multiple problem domains in smart spaces or even other applications.

The major problems addressed in our work are as follows:

1. Entrance detection: Once a person walks into a sensing area, we are able to observe an abrupt change in the CSI values. We desire to capture the moment when and where (if more than one doorway exits) this entrance happens. For this task, we need to know the number of existing doorways and also, some training samples to learn characteristics of this event, since the dynamics of *entrance* action can differ from place to place. Therefore, we propose to extract a set of features that describe the action of *entrance* properly, and then employ

a classifier to detect the event and distinguish between entering from multiple doorways.

2. User identification: We assume that each individual walks in their own specific manner, which yields identifiable changes in the CSI measures. We propose to build a user identification classifier to discriminate between a fixed and known group of users residing in a sensing area. For this purpose, we need to extract specific features that capture the diversity of human walking precisely, before applying the classifier.

3. Localization: We introduce an unsupervised localization technique that only needs to know the number of distinct locations (e.g. rooms or distinguishable positions) within an indoor space a priori. We map the CSI measurements into a bag-of-words data structure and then employ a topic modeling approach to discover clusters in the location of a moving user within the wifi-covered sensing area.

After data analysis, the last step of the system is to incorporate the outputs of the preceding processes into an intelligent decision making unit. This module might take actions, or query the processing module for further information. Depending on the user preferences and specific needs, these strategies can be selected from a predetermined set of actions, automatically learned or heuristically adapted from data. For example, if we are designing a smart space equipped with security and surveillance systems, the decision making block could be as simple as an anomaly detection process that flags suspicious events such as

- entrance from unusual entry points other than the known doorways, such as windows or emergency exits
- entrance of a person whose identity cannot be recognized by the user identification process
- unusual fast movement among different rooms right after entrance

In this example, if the decision making module concludes *normal* state, a routine activity recognition system may be activated, where depending on the position of the subject in the area, the heating or cooling system can start functioning, or the lighting can be automatically adjusted according to user preferences.

There are many other applications that could benefit from the combination of these processing units, as well as many other intelligent strategies to make decision for a smart space system, which we will discuss on in Chapter 6. In this chapter, we will focus on the technical description of the contributions listed above.

5.2.1 Interpreting from CSI Readings

In this section we describe the common challenges and limitations of our wifi-based smart space system. As briefly mentioned earlier, there are three major challenges in data mining with CSI measurements; the noise from multiple sources (e.g. signal interferences and power level adaptation in access points), the undesired changes in the environment (e.g. movement of non-target person or object within sensing area), and the location-dependency of activities and movements. Here, we suggest a few assumptions and pre-processing steps that yield the removal of unwanted disturbances without compromising the resolution of the information and computing relevant statistics that represent the desired variations.

5.2.1.1 Assumptions and Limitations

One practical fact that we needed to consider is that most of the off-the-shelf wifi access points have an internal power adaptation technology that automatically recognizes and adjusts transmission power levels to minimize interference in zones covered by multiple access points. In general, this provides an overall higher quality wireless connection, but the CSI estimations of a transmitted packets may vary with different power levels. This can cause a problem for our data modeling tasks, since some unwanted variation might occurs in the data stream. Therefore, for our data collection we used access points with modified firmware to obtain a fixed range of power settings, to minimize this effect. Another solution would be to normalize the magnitude of the CSI values of different subcarriers in each stream, such that $C_{l-Norm} = \frac{C_l - min(C_l)}{max(C_l) - min(C_l)}$, which reduces the effect of power variation. However, in some tasks the actual magnitude range of the CSI values conveys information about the events and using the normalized values is not as effective.

Due to the location-dependency property of the CSI measurements, the location and distance between the transmitter and receiver affect the observed communication. However, since the locations of wifi endpoints do not often change once they are placed, we train and test our models while the location of these endpoints are fixed. If the location of these devices changes, the parameters of the models would need to be retrained to preserve the accuracy.

Additionally, in this stage of the technology, our system is designed to track variations of the CSI measurements from a single user at a time. To the best of our knowledge, most of the studies working with the same type of data are in their

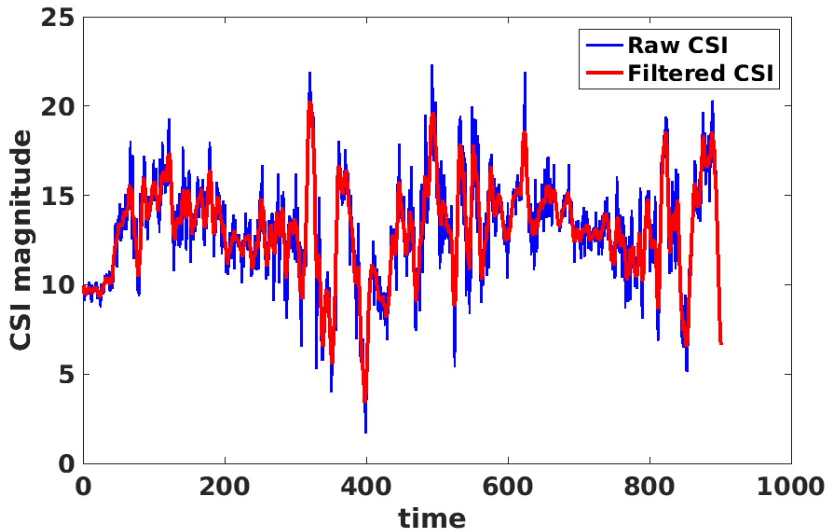


Figure 5–4: An example segment of CSI magnitude of one subcarrier, before and after applying the noise-removal filter. This segment is collected while one person is *walking* in the sensing area.

preliminary stages of development and their techniques are implemented in a single-user setting. We followed the same setting and assumed that at each time stamp there is only one main target user, which we track.

5.2.1.2 Noise Removal

When a person walks into the sensing area, the variations of CSI data can be used to infer information about the characteristics of this event. However, the raw data contain noise introduced by surrounding and high-frequency movements. On the other hand, the duration of typical human activities and gestures is greater than hundreds of milliseconds and happens at low frequencies (no more than $2Hz$) [117]. Therefore, we apply a low-pass filter with cut-off frequency of $2Hz$ in order to remove the high-frequency noise as well as the static component. From now on when we mention the CSI data, we mean the filtered data. Figure 5–4 exhibits an example of the raw CSI magnitude of one specific subcarrier before and after applying the low-pass filter.

5.2.1.3 Feature Generation

Several problem-dependent feature extraction techniques are proposed in the literature for dealing with radio signals, including RSSI and CSI measurements, that are in the form of trajectory data or time-series. The most commonly used techniques include statistical features (such as mean, variance, min, max, entropy and histogram), discrete wavelet transform (DWT) and fast Fourier transform (FFT) [130, 6, 159, 4]. Mapping the raw data into a proper feature space reduces the computational complexity of the system, improves the performance of the processing units and helps overcome the location-dependency problem in detecting and recognizing activities with displacement movements.

Let $\boldsymbol{C(t)} = \{C_1, \cdots, C_l\}$ denote the CSI samples of L streaming links, where $C_l(t) = \{c_{1,l}, \cdots, c_{s,l}\}$ represent the magnitude of samples on subcarrier $s \in \{1, \cdots, S\}$ in stream $l \in \{1, \cdots, L\}$ at time t. Human motions and environmental changes affect the L stream independently, but they affect the S subcarriers of one stream in a similar manner. Figure 5–5 illustrates an example of measured CSI magnitudes from all subcarriers in all streams in the same area.

The feature generation step begins by sliding a moving window with length w to create a frame of consecutive temporal samples,

$$\boldsymbol{W}_{s,l}(t) = \{\boldsymbol{C(t-w+1)}, \cdots, \boldsymbol{C(t-1), C(t)}\} \tag{5.7}$$

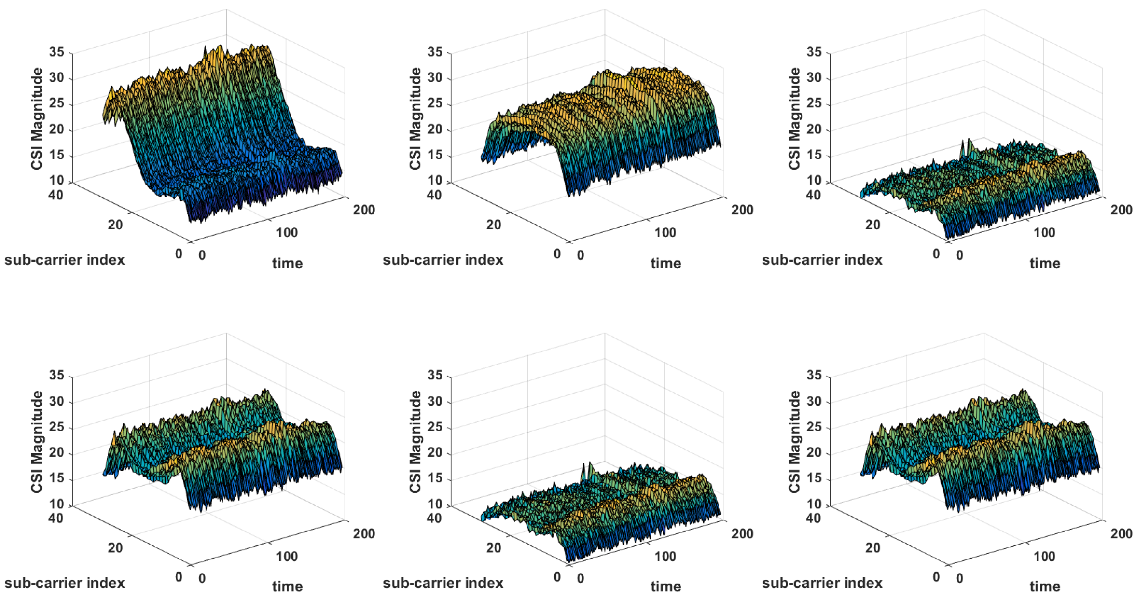


Figure 5–5: An example of CSI magnitudes of all subcarriers. In this example, for each packet the CSI values were measured from 6 data streams and 30 subcarriers per stream. These plots are exhibiting CSI readings of all streams in the same sensing area and the same time-interval.

wherein we attempt to infer the intended event.

Variance. A very intuitive feature that helps to quantify how far the measurements are spread out during the time frame $\boldsymbol{W}$ is the moving variance, which we denote by $V_{s,l}(\boldsymbol{W}) = \{v_{1,1}, \cdots, v_{s,l}\}$ where

$$v_{s,l} = \frac{1}{w-1} \sum_{i=1}^{w} |\boldsymbol{W}_{s,l}(i) - \mu_{s,l}|^2, \tag{5.8}$$

$$\mu_{s,l} = \frac{1}{w} \sum_{i=1}^{w} \boldsymbol{W}_{s,l}(i), \tag{5.9}$$

is individually computed for subcarrier s in stream l.

Shannon entropy. Entropy of the magnitudes $H_{s,l}(\boldsymbol{W}) = \{h_{1,1}, \cdots, h_{s,l}\}$ is a measure of unpredictability of information content, and is individually calculated for subcarrier s, stream l during time frame $\boldsymbol{W}$ as

$$h_{s,l} = -\sum_{i=1}^{w} \boldsymbol{W}_{s,l}(i) log(\boldsymbol{W}_{s,l}(i)). \tag{5.10}$$

Peak analysis. Another simple but informative type of feature that can be extracted from CSI magnitudes relates to signal peaks and valleys including peak positions, counts and heights. Peak information emphasizes the local minimum and maximum of the signals and is correlated with frequency of changes in the environment. Therefore, peaks they can be used to identify the occurrence of different events in the environment.

Histogram. The histogram of the CSI magnitudes within the time frame $\boldsymbol{W}$, represents the distribution of magnitudes over some pre-defined intervals (bins), and reflects patterns and/or locations of the movements or events within captured frame. Figure 5–6 shows example of histograms of CSI magnitudes for different activities/location scenarios.

So far, we have found these features to be the best representations of the activities and events that we aim to learn. Figure 5–7 shows examples of the raw signal and the extracted features while a user is performing two different activities in the same sensing area. We observe that the features have different *detection delays*, i.e. time that each feature takes to detect the start of new activities, when switching between activities. To evaluate the impact of different features on detection delay, one can measure the time or number of packets before a feature responds to changing

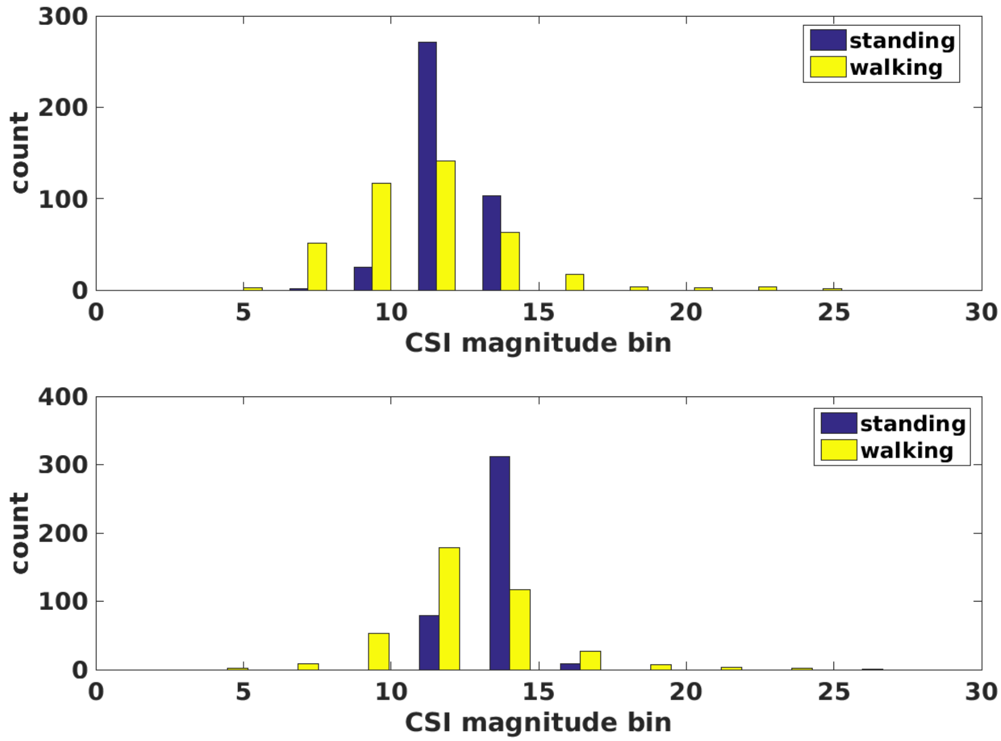


Figure 5–6: Histogram of CSI magnitudes of particular subcarriers for two different users performing *standing* and *walking* at the same locations.

activities. Table 5–1 depicts the detection delays of the features shown in the Figure 5–7. In the following sections, we will introduce other informative features that are

Detection delay	Variance	Entropy	Peak counts
Standing-Walking	0.25 ± 0.1	1.24 ± 0.3	0.23 ± 0.1
Walking-Standing	1.04 ± 0.2	0.53 ± 0.15	0.87 ± 0.2

Table 5–1: Detection delays of different features, in seconds, when a users switch between *standing* and *walking* (averaged over 10 rounds)

specifically suitable for each problem.

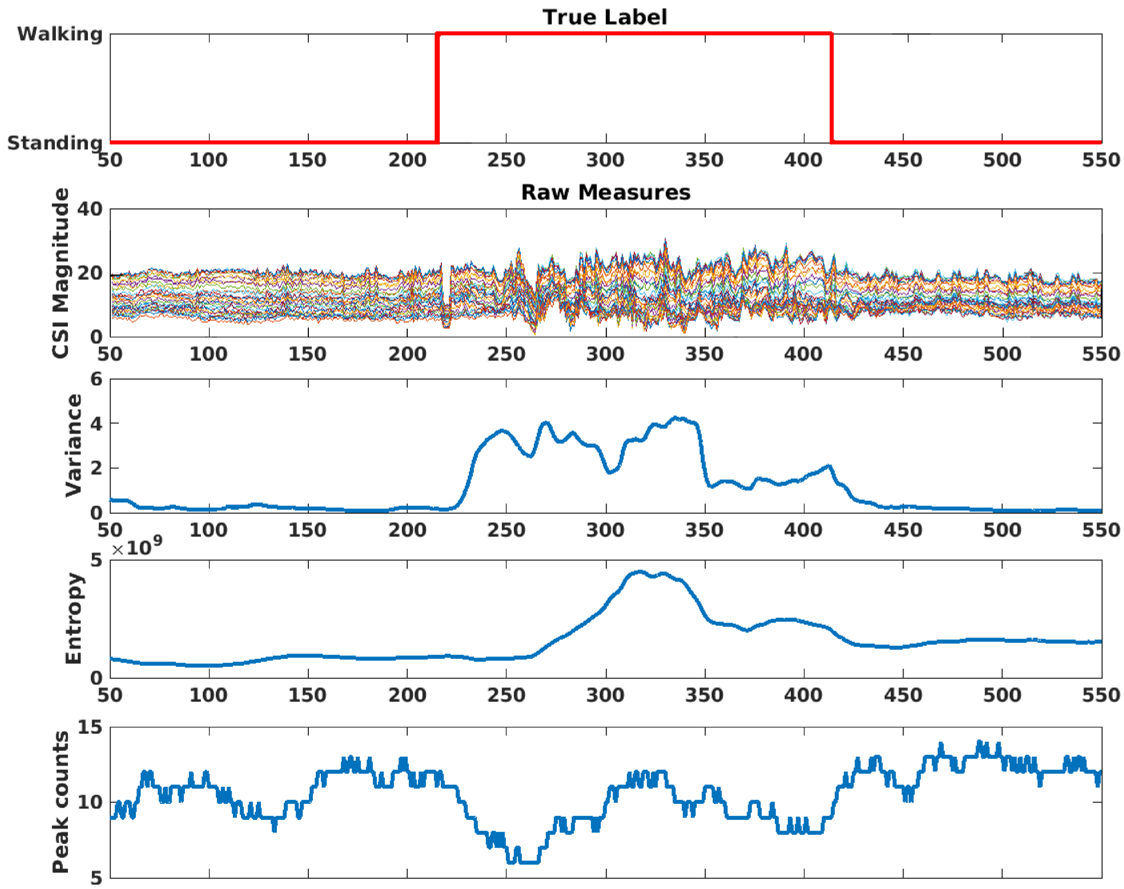


Figure 5–7: CSI magnitudes of all subcarriers of one stream and the extracted features using moving window for *standing* and *walking*.

5.2.2 Entrance Detection

A person entering into an empty sensing area will cause an abrupt change in the characteristics of the CSI values and detecting this event may seem obvious and easy. An intuitive solution would be monitoring the variance of the data streams and set a threshold to detect any abrupt change after a steady condition. However, there are many other activities or environmental changes that may produce the same or very a similar effect, such as standing up or sitting down and falling. Therefore,

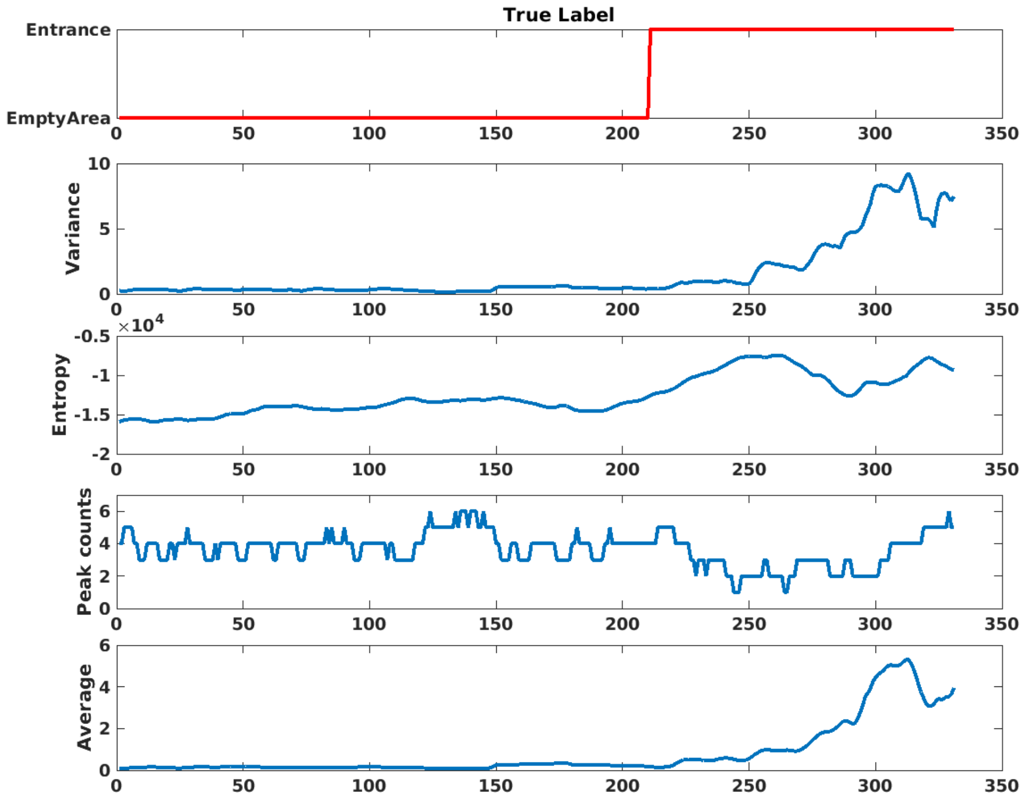


Figure 5–8: An example of extracted features for a segment of data including the state of empty spaces and an *entrance* event.

we need to learn the specific characteristics of the *entrance* action, which basically is a sporadic activity that happens when a person walks from a position out of the sensing area into the area through a doorway. An accurate temporal analysis of the data stream is required in order to detect an isolated, short event of this type.

We begin by applying the moving window method, as explained in the previous section, to create an overlapping sequence of data. We choose the following features to characterize anomaly patterns: 1) maximum moving variance, 2) average of CSI magnitudes, 3) maximum signal entropy, 4) peak counts. Figure 5–8 shows an example of extracted features from CSI magnitudes when a user enters an empty sensing area. This example illustrates the performance of each extracted feature in detecting the moment of *entrance*. Table 5–2 shows the detection delays of these features.

Detection delay	Variance	Entropy	Peak counts	Average
Empty-Enter	1.16 ± 0.15	0.63 ± 0.5	0.51 ± 0.1	1.75 ± 0.25

Table 5–2: Detection delays of different features, in seconds, when a users *enters* an empty sensing area (averaged over 20 rounds)

After feature extraction, we applied k-nearest neighbor (k-NN) classification. k-NN is an instance-based or lazy learning algorithm, which does not attempt to construct a general model but simply stores instances of the training data. The classification output is computed from a simple majority vote of the k nearest neighbors of a query point. In order to find the nearest point to a given point, we must define a distance function. A common choice for similarity measure between two data points $\boldsymbol{x}$ and $\boldsymbol{y}$ in N-dimensional space, is Euclidean distance:

$$\sqrt{\sum_{i=1}^{N}(x_i - y_i)^2}. \tag{5.11}$$

An optimal choice of k is highly data-dependent and usually is made based on cross-validation. In general, large values of k reduce the effect of noise on the classification but increase the bias of the classification, by only allowing simpler decision boundaries between classes.

The accuracy of k-NN degrades if feature scales are not consistent with their importance. Here, in our problem all feature contribute evenly to the classification performance and therefore, we need to normalize the values of the features to have them in a similar range.

There are various anomaly detection strategies, which are suitable for discovering sudden changes in the data. For instance, the moment of user *entrance* can be learned by a one-class classification algorithm, such as one-class Support Vector Machine (SVM). However, we considered the general case, where a sensing area can have multiple doorways and the smart space needs to be prepared to discover entrance from any doorway. Therefore, we have applied multi-class classification, which is capable of distinguishing between an empty space and the events of *entrance* through either of the possible doorways. In Section 5.3 we will describe the details of the experimental setup for this problem.

5.2.3 User Identification

User identification can be considered a prerequisite for any smart space application. Apart from the security aspect, if the smart home detects which member has entered the space, it can activate user-specific customization such as recommendations on TV programs and adjusting room temperature and lighting. Current identity recognition systems mostly rely on biometric attributes such as fingerprint, retina and face recognition or behavioural attributes such as voice and signature [149]. These technologies are usually accurate and reliable, but most of them suffer from practical issues such as privacy concerns, physical contact with sensors, high implementation and maintenance cost, and cooperation from the subjects. Therefore, a device-free method can be a great alternative for user identification, at least in screening applications such as surveillance of smart spaces. We assume that a

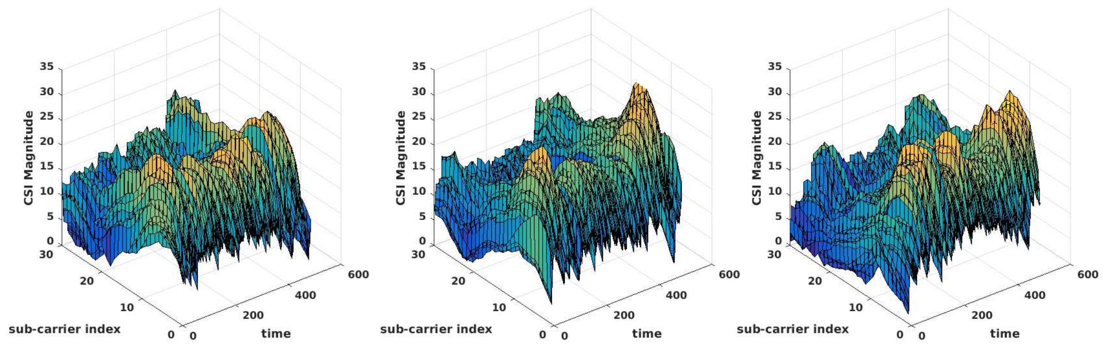


Figure 5–9: CSI magnitudes of *walking* on the same path performed by 3 different users.

typical residential or small industrial indoor space such as home or office is usually shared between some known number of residents. We also consider the fact that, generally speaking, a person passes through a doorway by *walking* into the sensing area. Therefore, the authorized users of an smart space can be recognized and distinguished by the specific way or pattern of their walk, also known as *gait* recognition. The variation of CSI measurements reflect the person's gait as well as unique body part movements or postures. Quantifying the exact body characteristics such as body mass, height and specific body part movements (e.g., legs, arms and hands) from wifi signals is a difficult task, requires specific antenna assembly or software radio, and is not feasible using current off-the-shelf wifi hardwares [5, 130].

We suggest to create a feature space that captures both walking patterns and general body shape of different users. For data collection, we considered CSI samples gathered during both *walking* and in-place activities such as *sitting* and *standing*. Figure 5–9 presents an example of CSI magnitudes of all $S = 30$ subcarriers in one particular stream (link) for *walking* activity performed by 3 different users. The feature set is calculated separately for each data stream by averaging over all samples $t \in W$, which includes: 1) variance, 2) maximum, 3) minimum, 4) peak counts, 5) skewness, and 6) kurtosis. In addition to these features for each stream, we aggregate the CSI magnitudes of all subcarriers into one single value by getting the average of 5 successive subcarriers as suggested in [71]. These time-domain statistics capture the shape of instantaneous distortion of channel frequency response (CFR) of all subcarriers.

The other aspect of user identification is to choose a proper classification technique. We expect our system to recognize different users from a predefined set and be able to classify any other person in a *stranger* class. We suggest to use Random Forest classifier, which provides multi-class classification as well as a confidence level of the classification. The idea is to use the confidence level to indicate the presence of a stranger in the sensing area. When someone other than the predefined members walks into the sensing area, the classifier would still try to find the closest user profile match and assigns a class to them. However, since the walking pattern of the stranger is completely different from the existing profiles we expect the classifier to produce a very low confidence level. In this case, we classify the samples with very low confidence level in the stranger class. Depending on the sensitivity of the measurements and the experimental setup of the space, the threshold for indicating the *stranger* class can be heuristically set to minimize the false alarm (false positive) rate.

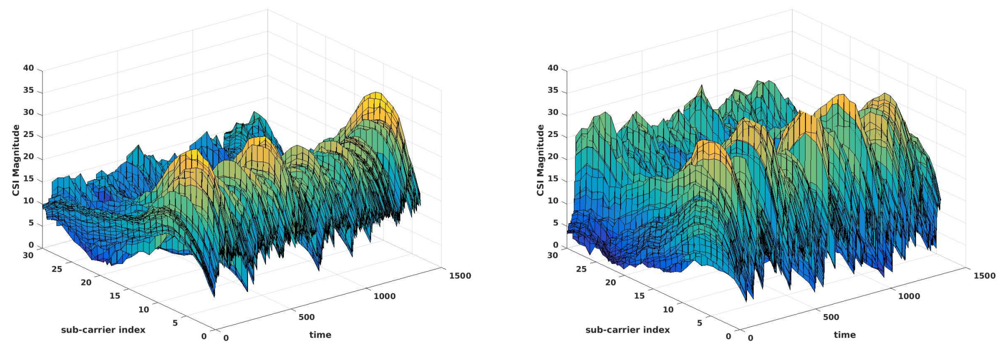


Figure 5–10: CSI magnitudes of *walking* of a user on 2 different locations within the same sensing area.

5.2.4 Localization in Smart Spaces

Indoor localization refers to the problem of positioning or determining the location of a device or moving object within a space using radio waves, magnetic fields, acoustic signals or other sensory information. We propose an unsupervised learning approach for clustering CSI measurements in order to build fingerprints for locations within a sensing area. We intend to design a smart space system that automatically adapts to new places without great re-training and annotation effort. Tracking the location of a user across the smart space can be useful in many application scenarios, such as assisted living, security and human-machine interaction. For example, the location of an elderly person who lives alone is a very important component of assisted living application that helps offsite caregivers to observe daily routines in order to detect hazardous events such as long stays in bed or in an unusual area (e.g., shower or bathroom).

In this work we aim to take advantages of the location-dependency behaviour of CSI measurements and quantify the variations in these values that occur due to the displacement of a moving user. Figure 5–10 depicts the CSI magnitudes of *walking* on a straight path at two different locations of a sensing area. Although we are aware of the fact that the CSI data, depends on the location of the disturbance, it is not clear how these variations can result in location identification. To investigate this, we suggest to observe the CSI readings of the sensing area while users perform an specific activity, namely *walking*, across all locations within the place. At each time stamp, the values of the CSI matrix elements reflects diversity of the disturbances in frequency and space.

In order to discover the frequency-space correlation, we suggest to extract histogram-based features from the frequency component of the CSI values. The distribution of CSI values over the frequency domain is highly correlated with the location information of the disturbances. Inspired by topic modeling in text documents, we aim to employ LDA to discover location clusters (i.e., topics) from the distribution of CSI magnitudes. Before employing LDA clustering, we need to construct bag-of-word data from raw CSI magnitudes, where each observation (i.e., document) is a distribution over a fixed sequence of discrete units. We suggest to compute the histogram of CSI magnitudes of all subcarriers during time frame W. In this way, at each time frame we obtain the CSI distributions (i.e., *word counts*) over a fixed set of consecutive, non-overlapping magnitude intervals (i.e., *words*). Figure 5–11 shows a representation of histogram-based features extracted from the same signals

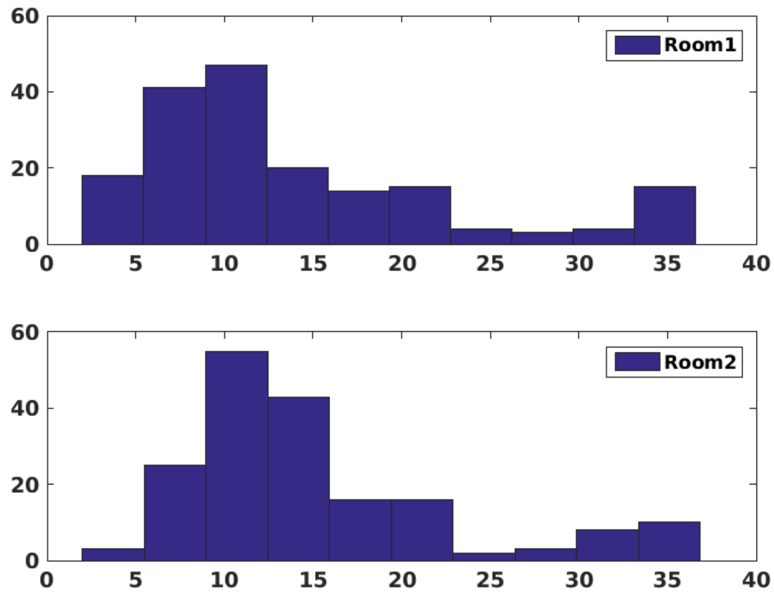


Figure 5–11: Histogram of CSI magnitudes of *walking* of a user on 2 different locations within the same sensing area.

presented in Figure 5–10. We can observe that the distribution of magnitudes varies from one location to another.

The size of the topic space or the number of location clusters is a fixed parameter which given a priori. The more location clusters we define within a sensing area, the higher localization resolution we obtain. From our experiments, we find that the CSI values changes gradually when the source of disturbance changes its position; hence, abrupt changes only happen when a significant displacement has occurred, for example, when the user walks from one room into another. Therefore, the number of topics that LDA can successfully discover from the CSI data is roughly the number of separate rooms in the sensing area. The main reason behind choosing the LDA

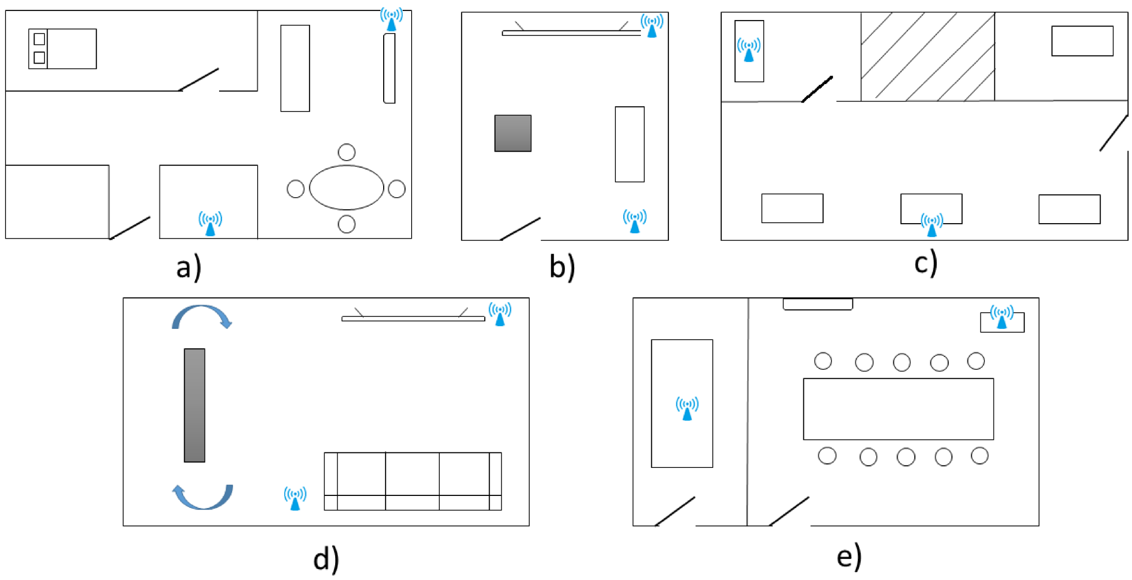


Figure 5–12: Location layouts and device setup: a) apartment (about $650ft^2$), b) office 1(about $200ft^2$), c) lab (about $600ft^2$), d) TV room (about $250ft^2$) e) conference rooms (about $300ft^2$)

algorithm for this problem is that for indoor localization we have location clusters that may not have very clear boundaries and LDA is capable of handling shared regions.

5.3 Evaluation

In this section we illustrate the implementation and various experimental setups under different naturalistic conditions to evaluate the performance of each proposed method. Our approaches are implemented and evaluated on a data set collected between November 2015 and April 2016, at TandemLaunch.

5.3.1 Experimental Setup

The experiments were conducted in real world situations and the CSI data was collected using 2 commercial off-the-shelf wifi devices. The access point is a Linksys

EA6500 V2 802.11n wifi router with 3 omnidirectional antennae and the client is an APU (access point unit) 2B4 wifi card with 2 external antennae, both operating at $2.4GHz$ band. The packet transmission rate was set to 40 packets per second, and for each packet a CSI sample with CRF of 30 subcarriers were extracted. Therefore, the CSI is collected in 30 subcarriers and 6 streams per packet, which reflects the signal diversity in frequency and space. For temporal analysis of the signals, an overlapping sliding window with a length of $W = 20$ packets for entrance detection and user identification, and $W = 100$ for the localization experiment, were chosen. The length of the moving window can affect the performance of each processing task. Ideally, we would like to have a long enough window to capture the activities, but short enough to preserve the resolution of temporal variations in the signals and allow quick detection.

Since our primary focus is smart spaces such as homes and offices, we choose multiple indoor areas for our data collection and evaluation. These locations include a one-bedroom apartment and different sections of one big office, i.e., office 1, lab , TV room and conference rooms. The sensing area layouts and their corresponding approximate sizes, as well as the setup of devices are depicted in Figure 5–12. A total of 8 volunteers were selected to collect CSI data while they were asked to perform some simple activities such as *walking* and *standing* in different locations depending on the experiment scenarios.

The algorithms and techniques were implemented in ©MATLAB, and for classification we used the machine learning toolkit WEKA [70].

Classified / Actual class	Lingering	Entrance	Walking
Lingering	0.96	0.02	0.02
Entrance	0.19	0.79	0.02
Walking	0.15	0.02	0.83

Classified / Actual class	Lingering	Entrance	Walking
Lingering	0.96	0.04	0
Entrance	0.08	0.91	0.01
Walking	0	0.02	0.98

Classified / Actual class	Lingering	Entrance1	Entrance2	Walking
Lingering	0.88	0.02	0.09	0.01
Entrance1	0.02	0.79	0.04	0.15
Entrance2	0.07	0.04	0.77	0.12
Walking	0.01	0.14	0.05	0.80

Figure 5–13: Normalized confusion matrices of entrance detection at; Top left: office 1, Top right: conference room, bottom: TV room.

5.3.2 Entrance Detection Validation

We considered two typical scenarios for entrance detection; a space with only one doorway and an area with two entry doorways. Therefore, the experiments for entrance detection were conducted in 3 different locations; office 1 (one doorway), right-side conference room (one doorway) and TV room (two doorways). For collecting the training and test data, the entrance activity was repeated for 20, 20 and 40 rounds by 2 to 4 participants in office 1, conference room and TV room, respectively. Each round of the experiments took 20-35 seconds to complete. Since the activity of *entrance* typically consists of 3 consecutive states of *lingering* outside of, *entering* into and *walking* inside the sensing area, the recognition tasks were performed on 3 classes (for one doorway) and 4 classes (for two doorways) to extract the *entrance* moments. The normalized confusion matrices of the k-NN classifiers are depicted in

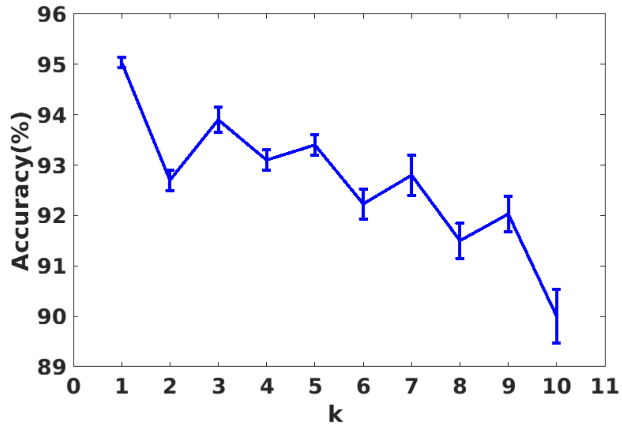


Figure 5–14: Cross-validation over conference room data for choosing the best number of neighbors for k-NN.

the Figure 5–13, independently for each location, due to the specific device setup of each place. We observe that *entrance* is often misclassified as *lingering* in the conference room location and office 1, whereas in the TV room it is more misclassified as *walking*. This is expected since in the first two places the doorways are actual doors that users have to open and then walk into the room, which creates a temporal delay in detecting the *entrance* activity. From the experiments that we have conducted, this detection delay is approximately about 20 packets or between 0.4-0.6 seconds. In the TV room, we do not have actual doors and the doorways, indicated by arrows in Figure 5–12 (d), are open gateways which users can walk through right after *lingering*, and therefore the *entrance* activity can be confused with *walking* more often.

In order to find the optimum number of neighbors, 10-fold cross-validation was employed and for this dataset $k = 1$ was always the best number of neighbors. Figure 5–14 illustrates how the accuracy changes with different values of k used for classification. Overall, the system was able to detect the correct activities with an average accuracy of $90.51\% \pm 1$, $95.04\% \pm 1$ and $84.04\% \pm 2$ for office 1, conference room and TV room, respectively. The results indicate that the entrance detection system is robust in identifying the presence of a subject in the sensing area and has the potential to provide an alternative to current device-oriented presence detection systems.

5.3.3 Person Identification

For this experiment we considered different numbers of participants at the lab, conference room and TV room. For gathering walking patterns, the participants were asked to *walk* in the sensing area from 30 to 60 seconds (3 rounds each), on the same path. Most of the research on gait recognition imposes restrictions on the walking activity, e.g. that it should be performed on a straight line path, but we aim to investigate a framework where the users can explore more naturalistic *walking* paths. Therefore, in addition to walking on a straight line in the TV room, we also conducted an experiment in the conference room where the participants were asked to walk counter-clockwise around the table (see Figure 5–12 (e)). In the lab location, the participants were asked to freely walk in the area and even explore small rooms inside the lab, as long as they followed predefined paths consisting of straight lines, turns and circular paths.

Beside *walking*, we also asked the participants to perform 2 in-place activities, *standing* and *sitting*, to capture the characteristics of their body while no location dependent displacement is occurring. At each location, we built 2 types of classifiers using Random Forest, based on the type of training data; 1) only from *walking* data, and 2) from combination a of *walking*, *sitting* and *standing*. Although the pattern of *walking* is a very strong discriminative feature, which can be used for identifying different people, experimental results show that including in-place activities improves the recognition results.

Note that in a realistic problem setup we need to take the temporal dependency of the predicted labels into consideration. This means that the user identification can not be performed independently on consecutive samples of the data stream. Therefore, in order to have a consistent system, it is necessary to decide the user's identity within a time frame, instead of per instance. This time frame guarantees that a user is detected within the sensing area only if the classifier has been predicting the same label for a reasonable time frame. We empirically set this delay to 40 packets or 1 second, which is a reasonable duration considering the frequency of human activities.

Another factor that affects the robustness of the system is the users group size. This system is targeted towards smart homes or other small indoor spaces, where the area is typically shared between 2-5 family members or 4-6 colleagues. Based on this assumption, we assessed the performance of the user identification with group sizes between 3 to 6 people. The recognition results are presented in Figure 5–15 for different locations, classifier types and numbers of participants.

In order to detect a person outside the group, i.e. *stranger*, we used the confidence level score provided by Random Forest to reveal any significant uncertainty in the prediction models. For each location, a group of 3 *strangers*, whose data was not included in the training process, were asked to perform similar activities. When we

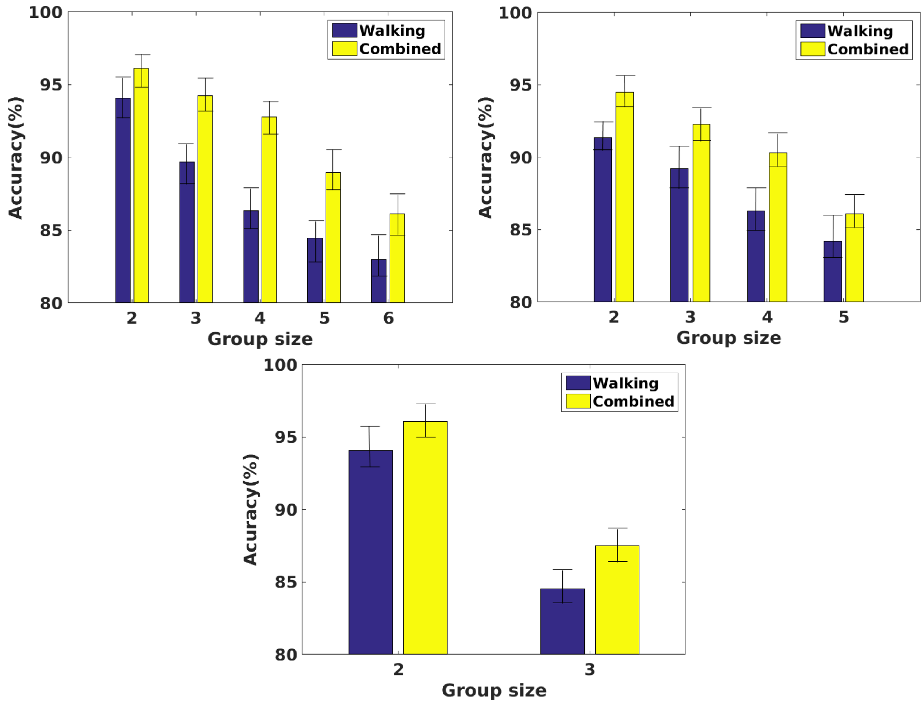


Figure 5–15: Performance of the user identification system with different group size of participants and types of activities at; Left: conference room, Middle: office 1, Right: lab. In all locations, the classifiers trained on combined activities including *walking*, *standing* and *sitting* outperform classifiers trained only on *walking* data

test this new data against the trained models, they get classified into the most similar user profile classes. However, the confidence level of these predictions is usually very low. We needed to set a threshold to decide if a person's walking fits into the existing models, which can be found by applying a heuristic search. The optimum threshold for different locations was selected between $0.75 - 0.84$, and based on the best threshold we were able to detect a *stranger*'s walking with $82.30\% \pm 1$, $85.34\% \pm 2$ and $88.06\% \pm 1$ accuracy for the conference room, office 1 and lab, respectively.

In general, the results of these experiments illustrate promising properties of CSI data for person identification. We observe that the combined analysis of human activities improves the accuracy over simply using walking analysis. This means that the CSI data and especially the time-domain statistics that we extracted from the data capture the shape of instantaneous distortions caused by not only walking, but also the body shape and volume of the human subject. This device-free user identification processing unit can be viewed as an added feature to any intelligent system by promoting personalized services.

5.3.4 Localization Identification using LDA

In this part we evaluate the proposed localization technique, which clusters a sensing area into a fixed number of regions based on the CSI data. This experiment was conducted at two different locations, the apartment and the lab, by 2 to 4 participants (10 rounds each). For each round, the participant was asked to walk inside all the regions of each location (for 30 seconds), following a predefined path, while we gathered CSI measurements. As explained earlier in Section 5.2.3, we mapped the CSI data into a histogram-based feature space and then clustered these features using the LDA algorithm. Both of these locations have 4 distinguishable rooms, ideally the LDA should cluster the unlabelled data into the 4 correct categories.

For the feature extraction step, we computed the counts over histogram bins for each stream, for the moving window $W = 100$. The reason for choosing a larger window in this experiment is that we are interested in spatial displacements of the events in the environment, and due to the low frequency of human mobility we required to capture at least 1-2 steps cycles of their walk for accurate localization.

	Unsupervised		Supervised	
	LDA-Hist	k-means	SVM	Random Forest
apartment	83 ± 2	65 ± 3	86 ± 1	89 ± 2
lab	82 ± 2	67 ± 3	86 ± 1	83 ± 1

Table 5–3: Accuracy (%) of different learning techniques for location identification from histogram-based features.

As baselines, we applied two supervised learning techniques, support vector machine (SVM) and Random Forest, and an unsupervised learning algorithm, k-means clustering, to evaluate the performance of the proposed algorithm. The baseline algorithms were applied on the CSI magnitudes of all subcarriers and all streams.

Table 5–3 summarizes the clustering and classification results for both locations. As we expected, the supervised learning techniques have better performance in identifying the region of users' movements. However, for localization it is often time-consuming and infeasible to obtain labelled data with good resolution. Our technique is outperforming the classic k-means and only reduces the accuracy of identification by 4-6%, which is negligible for the types of tasks we have in mind.

One parameter that affects the performance of the system is the number of histogram bins, where we mapped the CSI magnitudes of each stream into. In fact, the number of these bins (per stream) gives the vocabulary size of the topic modeling approach. Figure 5–16 shows the comparison of results for different number of features. The best accuracy is achieved with vocabulary size of 150 words.

Misclassification mostly occurs because regions from different categories share a lot of common characteristics. The LDA allows the embedded location information

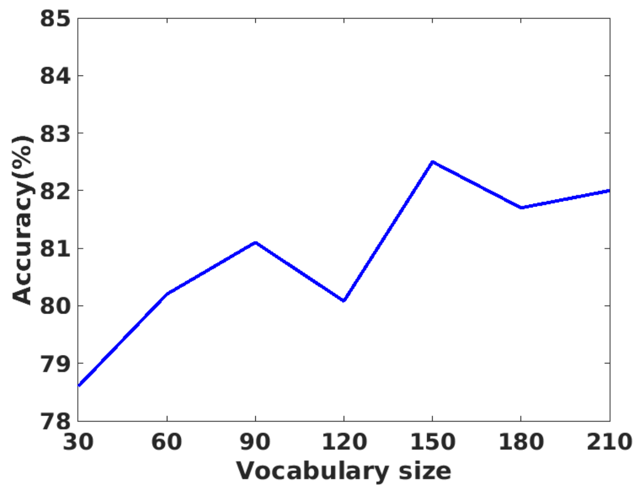


Figure 5–16: Clustering accuracy for different vocabulary size.

in the CSI data to be represented by mixtures of various location clusters. The evaluation results suggest that the proposed framework, including the histogram-based features and LDA mixture modeling, is able to discover the location-dependency of CSI variations.

5.4 Conclusion

This chapter describes a novel system for device-free smart environments that incorporates off-the-shelf wifi devices as the sensing infrastructure and leverages the wireless signals to learn about human activities and behaviour within an indoor area. The central idea to use physical layer information from wireless devices to quantify the disturbance and variations that occur in the signals due to environmental changes, including human movements and activities. We presented practical solutions for three important design challenges of smart space leveraging this technology: entrance detection, user identification and localization. All of these methods work with the same stream of data, only differ in data mining strategies, and can be employed individually or in parallel with the others. While existing methods addressing these tasks need to be trained on a large set of data, we demonstrated that even with the fairly small amount of data (about 5-10 minutes per activity class) our proposed approach is able to accurately predict labels for user presence, identity and location. The experimental results for the proposed wifi-based system show that device-free activity recognition is a promising line of research both for academia and industry.

We focused on customized approaches for each individual location, because in real-world scenarios some calibration and parameter tuning will be needed depending on the conditions of the sensing area. CSI measurements are very sensitive to the location of the transmitter and receiver devices, as well as ambient unwanted variations, which needs to be considered while training the models. Designing a more general systems that could adaptively learn the difference between locations and tune parameters automatically, is an interesting future direction for this application.

CHAPTER 6
Conclusion

Smart mobile devices have become an essential part of our society and they are capable of sensing a wealth of information about human behaviour and mobility. User activity recognition is a well studied and challenging research area because of the diversity and complexity of human behaviour. In this spirit, we proposed and evaluated various machine learning techniques for learning human behaviour patterns from mobility data obtained from multiple sensing technologies, on both individual and group level. Motivated by technical challenges of inferring semantic information from a large volume of human data, we studied three distinct research topics in the field: energy efficiency in wearable computing, high-level mobility analysis and smart indoor spaces. When considered together, these solutions can be at the foundation of an intelligent system that discovers meaningful patterns and routines from human daily life, indoors and outdoors. We conclude this thesis by summarizing the main contributions presented in the preceding chapters and by providing directions for future work.

6.1 Contributions

In Chapter 3 we presented a sensor selection strategy for mobile wearable systems to efficiently reduce the power consumption of the activity recognition applications on small portable devices, while maintaining accuracy. We proposed an approach that actively selects a smaller subset of sensors that are the most informative yet energy-effective for each time frame, and evaluated the performance of the algorithm on real data contained a number of sensor modalities and multiple activities. The empirical results confirm that the proposed online classifier selection method provides good power efficiency without significant loss in prediction accuracy.

In Chapter 4, we applied a fully unsupervised method, to discover routines and preferences in human daily movements and mobility behaviour, with minimum requirements of annotation and prior knowledge about the structure of data. We proposed to use an HDP(Hierarchical Dirichlet process)-based clustering method for analyzing location data obtained from mobile devices over extended periods of time, in order to glean high-level information about the long-term behaviour patterns of different groups of users. In general, the proposed method included two layers of learning steps: the hierarchical clustering layer that models the mobility behaviour, and a regularization step on top of the HDP clustering to keep the size of models in check. We evaluated our method on three different real data sets including both fine-grained indoor trajectories and coarse-grained outdoor mobility traces. Our results confirm that the regularized-HDP model is capable of interesting qualitative modeling, which offers a high-level perspective for learning the underlying structure of events, even in the presence of noisy and complex trajectories. While a lot of the works on HDPs assume that the strength of the prior will be sufficient to control the model size, our regularization approach provides tighter control over this aspect. The proposed regularization is general and should be useful for other applications of HDPs as well. The proposed approach is thoroughly conducted and evaluated on wide range of real-world data types, including user mobility traces (indoor and outdoor scales), interpersonal interaction levels and cellular antenna tower statistics. Therefor, the proposed methodology have the potential to be integrated with current technologies to provide opportunities for better understanding of human mobility behaviour on both individual and social levels.

In Chapter 5, we presented a device-free activity recognition system in the context of smart spaces, consisting of three main tasks: entrance detection, user identification and localization. These recognition modules leveraged very recent sensing technology created by the wifi network coverage of off-the-shelf wireless devices, in order to monitor the behaviour and movements of users within an indoor space. In particular, we were interested in analyzing the disturbance in the channel state information (CSI) values, which occurs due to human movements and activities within a sensing area. Each of the entrance detection and user identification blocks included a feature generation step that extracts statistical characteristics of the wireless signals and a classification step that distinguishes among different activity profiles and different user profiles, respectively. For location identification, we proposed to transform the CSI magnitudes into a histogram-based feature space to discover the frequency-space correlation between the raw CSI measurements and different locations of the area covered by the wifi. Then we applied Latent Dirichlet Allocation (LDA) to cluster distinct locations within the area. The experimental results on the proposed wifi-based system confirm that device-free activity recognition is a promising line of research in the field of context-aware computing.

6.2 Future Directions

In this section, we discuss interesting opportunities for future direction, suggested by the research presented in this thesis.

6.2.1 Efficient Wearable Computing

Human activity recognition from wearable devices provides great potential for context-aware computing for many applications. However, due to the computational and power capacity limitations on portable devices, the problem of efficiently processing large amount of modality data is not yet fully addressed. One possible improvement for the proposed framework in Chapter 3 is to take into account the precise energy consumption of each sensor, and optimize the sensor selection process based on the energy efficiency by introducing a cost factor, instead of making the decisions only based on the overall accuracy rate. Reinforcement learning may prove to be a useful methodology in this respect.

Another possible solution is to incorporate additional informative modalities such as microphone or camera to investigate the user's activities and environment context. These modalities are capable of capturing the surrounding events with high resolution. However, using visual and auditory information to improve the performance of wearable technologies brings extra computation cost, and is only applicable in specific domains, where privacy concerns are trivial or the users are willing to voluntarily provide such information in order to enhance the quality of their interaction with smart devices. Improving power consumption in the presence of such devices would be very useful.

6.2.2 Location-based Activity Recognition

In Chapter 4 we investigated some open questions in the field of modeling and analyzing human mobility trajectories from mobile sensing technologies.

Two alternative directions include 1) extending the regularized clustering methodology by evaluating on more diverse data types, 2) generalizing the methodology by providing a theoretical explanation of the effect of the regularization step used on the model from a Bayesian perspective. At the application level, we anticipate that the multi-layer clustering method would be helpful in discovering different semantic layers in structured data with. On the theory side, we expect that generalizations of the proposed regularization methodology would have a positive impact on other non-parametric frameworks, where the possible model complexity is in fact bounded, but we still want it to grow as more data become available.

6.2.3 Device-free Activity Recognition

As a new sensing strategy for activity recognition, there are many open research questions in the field of device-free smart systems that need to be addressed. We discuss a few existing limitations and issues of this technology and provide some possible directions for future studies.

As shown in Chapter 5, most of the modules in the smart spaces from data acquisition to decision making, contain several parameters that need to be tuned, such as transmission power level, sliding window size, packet sampling rate, classification parameters and thresholds. Each parameter may have a large impact on the accuracy and robustness of the smart system, hence, automatic optimization of all these parameters is an important research question. One possible solution is to use control theory, to analyze the behavior of dynamical system and offer hints of good settings.

In order to improve the quality of the recognition tasks in smart environments, a necessary step is to obtain feedback on the performance from the end-user of the technology. Therefore, developing a user interface that incorporates user commands and preferences in the loop of decision making would be an natural way to proceed. For instance, the interactive interface can prompt the user on their smart phone when the system is uncertain about the identity of the person who just walked into the smart home.

Another way to improve recognition accuracy is to use information from other layers of wireless connections, such as RSSI and MAC addresses. For example, as soon as a user arrives at home or office their mobile smart devices connect to the wifi network, and the MAC address of their device can be used to recognize the identity of the person. Also, RSSI information is usually correlated with the location of disturbances in the area and can be incorporated as an informative feature for the localization unit.

The majority of wifi-based activity recognition studies work with the assumption that only a single target user is moving within the sensing area at a time. One of the reasons for this limitation is that some activities or locations are out of reach of one pair of wifi devices. One possible solution is to combine multiple access point readings in order to obtain higher resolution information and improve the recognition results. Another way to increase the information gained from a sensing area is to incorporate other types of radio-frequency sensing modules such as Zigbee or RFID radio. However, the fusion of these different sources of data is still under exploration and needs further research.

So far, there is no mathematical theory that models and analyzes the correlation between human mobility and activities and corresponding wireless transmission features. The majority of research in this field illustrates the effectiveness of radio sensing through real world experiments. Therefore, one alternative direction would be to use information theoretic analysis to achieve essential insights that improve the design and performance of such systems.

References

[1] 2011 activity recognition challenge. http://www.opportunity-project.eu/challenge.

[2] Data 4 development challenge. http://www.d4d.orange.com.

[3] Reality mining - mit. http://realitycommons.media.mit.edu.

[4] Heba Abdelnasser, Moustafa Youssef, and Khaled A Harras. Wigest: A ubiquitous wifi-based gesture recognition system. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 1472–1480. IEEE, 2015.

[5] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C Miller. 3d tracking via body radio reflections. In *Usenix NSDI*, volume 14, 2014.

[6] Fadel Adib, Zachary Kabelac, and Dina Katabi. Multi-person localization via rf body reflections. In *Proceedings of the 12th USENIX Conference on Networked Systems Design and Implementation*, pages 279–292. USENIX Association, 2015.

[7] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 837–846. ACM, 2015.

[8] Michaelk Alandros, Lucy Y Pao, and Yu-Chi Ho. Randomization and super-heuristics in choosing sensor sets for target tracking applications. In *Decision and Control, 1999. Proceedings of the 38th IEEE Conference on*, volume 2, pages 1803–1808. IEEE, 1999.

[9] David J Aldous. *Exchangeability and related topics*. Springer, 1985.

[10] Kamran Ali, Alex Xiao Liu, Wei Wang, and Muhammad Shahzad. Keystroke recognition using wifi signals. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 90–102. ACM, 2015.

[11] Kerem Altun and Billur Barshan. Human activity recognition using inertial/magnetic sensor units. In *Human behavior understanding*, pages 38–51. Springer, 2010.

[12] Kerem Altun, Billur Barshan, and Orkun Tunçel. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognition*, 43(10):3605–3620, 2010.

[13] K Aminian, Ph Robert, EE Buchser, B Rutschmann, D Hayoz, and M Depairon. Physical activity monitoring based on accelerometry: validation and comparison with video observation. *Medical & biological engineering & computing*, 37(3):304–308, 1999.

[14] Daniel Ashbrook and Thad Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5):275–286, 2003.

[15] Louis Atallah, Benny Lo, Rachel King, and Guang-Zhong Yang. Sensor placement for activity detection using wearable accelerometers. In *Body Sensor Networks (BSN), 2010 International Conference on*, pages 24–29. IEEE, 2010.

[16] Akin Avci, Stephan Bosch, Mihai Marin-Perianu, Raluca Marin-Perianu, and Paul Havinga. Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey. In *Architecture of computing systems (ARCS), 2010 23rd international conference on*, pages 1–10. VDE, 2010.

[17] Ryan Aylward and Joseph A Paradiso. Sensemble: a wireless, compact, multiuser sensor system for interactive dance. In *Proceedings of the 2006 conference on New interfaces for musical expression*, pages 134–139. IRCAMCentre Pompidou, 2006.

[18] Jie Bao, Yu Zheng, David Wilkie, and Mohamed F Mokbel. A survey on recommendations in location-based social networks. *ACM Transaction on Intelligent Systems and Technology*, 2013.

[19] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data. In *Pervasive computing*, pages 1–17. Springer, 2004.

[20] Eric Becker, Zhengyi Le, Kyungseo Park, Yong Lin, and Fillia Makedon. Event-based experiments in an assistive environment using wireless sensor networks and voice recognition. In *Proceedings of the 2nd International Conference*

on PErvasive Technologies Related to Assistive Environments, page 17. ACM, 2009.

[21] Gerald Bieber and Christian Peter. Using physical activity for user behavior analysis. In *Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments*, page 94. ACM, 2008.

[22] Kanad K Biswas and Saurav Kumar Basu. Gesture recognition using microsoft kinect®. In *Automation, Robotics and Applications (ICARA), 2011 5th International Conference on*, pages 100–103. IEEE, 2011.

[23] David Blackwell and James B MacQueen. Ferguson distributions via pólya urn schemes. *The annals of statistics*, pages 353–355, 1973.

[24] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[25] Vincent D Blondel, Markus Esch, Connie Chan, Fabrice Clérot, Pierre Deville, Etienne Huens, Frédéric Morlot, Zbigniew Smoreda, and Cezary Ziemlicki. Data for development: the d4d challenge on mobile phone data. *arXiv preprint arXiv:1210.0137*, 2012.

[26] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[27] Leo Breiman and Adele Cutler. Random forests-classification description. *Department of Statistics, Berkeley*, 2007.

[28] Michael Buettner, Richa Prasad, Matthai Philipose, and David Wetherall. Recognizing daily activities with rfid-based sensors. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 51–60. ACM, 2009.

[29] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3):33, 2014.

[30] Andreas Bulling, Jamie Ward, Hans Gellersen, Gerhard Tröster, et al. Eye movement analysis for activity recognition using electrooculography. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(4):741–753, 2011.

[31] Nirupama Bulusu, John Heidemann, and Deborah Estrin. Gps-less low-cost outdoor localization for very small devices. *Personal Communications, IEEE*, 7(5):28–34, 2000.

[32] JBJ Bussmann, WLJ Martens, JHM Tulen, FC Schasfoort, HJG Van Den Berg-Emons, and HJ Stam. Measuring daily behavior using ambulatory accelerometry: the activity monitor. *Behavior Research Methods, Instruments, & Computers*, 33(3):349–356, 2001.

[33] Pierluigi Casale, Oriol Pujol, and Petia Radeva. Human activity recognition from accelerometer data using a wearable device. In *Pattern Recognition and Image Analysis*, pages 289–296. Springer, 2011.

[34] David Castanon. Approximate dynamic programming for sensor management. In *Decision and Control, 1997., Proceedings of the 36th IEEE Conference on*, volume 2, pages 1202–1207. IEEE, 1997.

[35] Claudette Cedras and Mubarak Shah. Motion-based recognition a survey. *Image and Vision Computing*, 13(2):129–155, 1995.

[36] Kin-Pong Chan and Ada Wai-Chee Fu. Efficient time series matching by wavelets. In *Data Engineering, 1999. Proceedings., 15th International Conference on*, pages 126–133. IEEE, 1999.

[37] Olivier Chapelle, Bernhard Schölkopf, Alexander Zien, et al. Semi-supervised learning. 2006.

[38] Lei Chen and Raymond Ng. On the marriage of lp-norms and edit distance. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, pages 792–803. VLDB Endowment, 2004.

[39] Liming Chen and Ismail Khalil. Activity recognition: Approaches, practices and trends. In *Activity Recognition in Pervasive Intelligent Environments*, pages 1–31. Springer, 2011.

[40] Yen-Ping Chen, Jhun-Ying Yang, Shun-Nan Liou, Gwo-Yun Lee, and Jeen-Shing Wang. Online classifier construction algorithm for human activity detection using a tri-axial accelerometer. *Applied Mathematics and Computation*, 205(2):849–860, 2008.

[41] Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM*

SIGKDD international conference on Knowledge discovery and data mining, pages 1082–1090. ACM, 2011.

[42] Tonmoy Choudhury, Sunny Consolvo, Brent Harrison, Jeffrey Hightower, Antonio Lamarca, Louis LeGrand, Azar Rahimi, Adam Rea, G Bordello, Bruce Hemingway, et al. The mobile sensing platform: An embedded activity recognition system. *Pervasive Computing, IEEE*, 7(2):32–41, 2008.

[43] Brian Clarkson and Alex Pentland. Unsupervised clustering of ambulatory audio and video. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 6, pages 3037–3040. IEEE, 1999.

[44] Brian Clarkson, Alex Pentland, and Kenji Mase. Recognizing user context via wearable sensors. In *iswc*, page 69. IEEE, 2000.

[45] Diane Cook, M Schmitter-Edgecombe, Aaron Crandall, Chad Sanders, and Brian Thomas. Collecting and disseminating smart home sensor data in the casas project. In *Proceedings of the CHI Workshop on Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research*, 2009.

[46] Diane J Cook. Learning setting-generalized activity models for smart spaces. *IEEE intelligent systems*, 2010(99):1, 2010.

[47] Diane J Cook, Michael Youngblood, Edwin O Heierman III, Karthik Gopalratnam, Sira Rao, Andrey Litvin, and Farhan Khawaja. Mavhome: An agent-based smart home. In *null*, page 521. IEEE, 2003.

[48] Tuan Hung Dao, Seung Ryul Jeong, and Hyunchul Ahn. A novel recommendation model of location-based advertising: Context-aware collaborative filtering using ga approach. *Expert Systems with Applications*, 39(3):3731–3739, 2012.

[49] Ray Denenberg. Open systems interconnection. *Library hi tech*, 3(1):15–26, 1985.

[50] Wen Dong, Daniel Olguin-Olguin, Benjamin Waber, Taemie Kim, and Alex Pentland. Mapping organizational dynamics with body sensor networks. In *Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference on*, pages 130–135. IEEE, 2012.

[51] Nathan Eagle and Alex Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.

[52] Nathan Eagle and Alex Sandy Pentland. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066, 2009.

[53] Nathan Eagle, Alex Sandy Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, 2009.

[54] Charles Elkan. The foundations of cost-sensitive learning. In *International joint conference on artificial intelligence*, volume 17, pages 973–978. Citeseer, 2001.

[55] Miikka Ermes, Juha Parkka, Jani Mantyjarvi, and Ilkka Korhonen. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *Information Technology in Biomedicine, IEEE Transactions on*, 12(1):20–26, 2008.

[56] Bienvenu Fassinut-Mombot and Jean-Bernard Choquel. A new probabilistic and entropy fusion approach for management of information sources. *Information Fusion*, 5(1):35–47, 2004.

[57] Li Fei-Fei and Pietro Perona. A bayesian hierarchical model for learning natural scene categories. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 524–531. IEEE, 2005.

[58] Thomas S Ferguson. A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230, 1973.

[59] F Foerster, M Smeja, and J Fahrenberg. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in Human Behavior*, 15(5):571–583, 1999.

[60] James Fogarty, Carolyn Au, and Scott E Hudson. Sensing from the basement: a feasibility study of unobtrusive and low-cost home activity recognition. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, pages 91–100. ACM, 2006.

[61] Emily B Fox. *Bayesian nonparametric learning of complex dynamical phenomena.* PhD thesis, Massachusetts Institute of Technology, 2009.

[62] Emily B Fox, Erik B Sudderth, Michael I Jordan, and Alan S Willsky. An hdp-hmm for systems with state persistence. In *Proceedings of the 25th international conference on Machine learning*, pages 312–319. ACM, 2008.

[63] Jason Frank, Shie Mannor, Joelle Pineau, and Doina Precup. Time series analysis using geometric template matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(3):740–754, 2013.

[64] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*, page 3. ACM, 2012.

[65] Enrique Garcia-Ceja, Ramon F Brena, Jose C Carrasco-Jimenez, and Leonardo Garrido. Long-term activity recognition from wristwatch accelerometer data. *Sensors*, 14(12):22500–22524, 2014.

[66] Hristijan Gjoreski, Mitja Luštrek, and Matjaž Gams. Accelerometer placement for posture recognition and fall detection. In *Intelligent Environments (IE), 2011 7th International Conference on*, pages 47–54. IEEE, 2011.

[67] David Graumann, Walter Lara, Jeffrey Hightower, and Gaetano Borriello. Real-world implementation of the location stack: The universal location framework. In *Mobile Computing Systems and Applications, 2003. Proceedings. Fifth IEEE Workshop on*, pages 122–128. IEEE, 2003.

[68] Thomas L Griffiths, Mark Steyvers, David M Blei, and Joshua B Tenenbaum. Integrating topics and syntax. In *NIPS*, volume 4, pages 537–544, 2004.

[69] Vijay Gupta, Timothy H Chung, Babak Hassibi, and Richard M Murray. On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica*, 42(2):251–260, 2006.

[70] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[71] Chunmei Han, Kaishun Wu, Yuxi Wang, and Lionel M Ni. Wifall: Device-free fall detection by wireless networks. In *INFOCOM, 2014 Proceedings IEEE*, pages 271–279. IEEE, 2014.

[72] Dae-Man Han and Jae-Hyun Lim. Design and implementation of smart home energy management systems based on zigbee. *Consumer Electronics, IEEE Transactions on*, 56(3):1417–1425, 2010.

[73] Zhenyu He and Lianwen Jin. Activity recognition from acceleration data based on discrete consine transform and svm. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, pages 5041–5044. IEEE, 2009.

[74] John R Hershey and Peder A Olsen. Approximating the kullback leibler divergence between gaussian mixture models. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 4, pages IV–317. IEEE, 2007.

[75] Peter Hevesi, Sebastian Wille, Gerald Pirkl, Norbert Wehn, and Paul Lukowicz. Monitoring household activities and user location with a cheap, unobtrusive thermal sensor array. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 141–145. ACM, 2014.

[76] Jeffrey Hightower, Sunny Consolvo, Anthony LaMarca, Ian Smith, and Jeff Hughes. Learning and recognizing the places we go. In *UbiComp 2005: Ubiquitous Computing*, pages 159–176. Springer, 2005.

[77] Diane J Hu. Latent dirichlet allocation for text, images, and music. *University of California, San Diego. Retrieved April*, 26:2013, 2009.

[78] Ya-Xuan Hung, Chih-Yen Chiang, Steen J Hsu, and Chia-Tai Chan. Abnormality detection for improving elders daily life independent. In *Aging Friendly Technology for Health and Independence*, pages 186–194. Springer, 2010.

[79] Tam Hunh, Ulf Blanke, and Bernt Schiele. Scalable recognition of daily activities with wearable sensors. In *Location-and context-awareness*, pages 50–67. Springer, 2007.

[80] Tam Huynh, Mario Fritz, and Bernt Schiele. Discovery of activity patterns using topic models. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 10–19. ACM, 2008.

[81] Tam Huynh and Bernt Schiele. Analyzing features for activity recognition. In *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies*, pages 159–163. ACM, 2005.

[82] Luciana C Jatoba, Ulrich Grossmann, Chistophe Kunze, Jörg Ottenbacher, and Wilhelm Stork. Context-aware mobile health monitoring: Evaluation of different pattern recognition methods for classification of physical activity. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 5250–5253. IEEE, 2008.

[83] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65. ACM, 2007.

[84] Ming Jin, Han Zou, Kevin Weekly, Ruoxi Jia, Alexandre M Bayen, and Costas J Spanos. Environmental sensing by wearable device for indoor activity and location estimation. In *Industrial Electronics Society, IECON 2014-40th Annual Conference of the IEEE*, pages 5369–5375. IEEE, 2014.

[85] Emil Jovanov, Amanda O'Donnell Lords, Dejan Raskovic, Paul G Cox, Reza Adhami, and Frank Andrasik. Stress monitoring using a distributed wireless intelligent sensor system. *Engineering in Medicine and Biology Magazine, IEEE*, 22(3):49–55, 2003.

[86] Emil Jovanov, Aleksandar Milenkovic, Chris Otto, and Piet C De Groen. A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. *Journal of NeuroEngineering and rehabilitation*, 2(1):6, 2005.

[87] Adil Mehmood Khan, Young-Koo Lee, Sungyoung Y Lee, and Tae-Seong Kim. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *Information Technology in Biomedicine, IEEE Transactions on*, 14(5):1166–1172, 2010.

[88] Eunju Kim, Sumi Helal, and Diane Cook. Human activity recognition and pattern discovery. *Pervasive Computing, IEEE*, 9(1):48–53, 2010.

[89] Andreas Krause, Daniel P Siewiorek, Asim Smailagic, and Jonny Farringdon. Unsupervised, dynamic identification of physiological and activity context in wearable computing. In *null*, page 88. IEEE, 2003.

[90] Narayanan C Krishnan and Diane J Cook. Activity recognition on streaming sensor data. *Pervasive and mobile computing*, 10:138–154, 2014.

[91] John Krumm and Eric Horvitz. Locadio: Inferring motion and location from wi-fi signal strengths. In *MobiQuitous*, pages 4–13, 2004.

[92] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.

[93] Anthony LaMarca, Yatin Chawathe, Sunny Consolvo, Jeffrey Hightower, Ian Smith, James Scott, Timothy Sohn, James Howard, Jeff Hughes, Fred Potter, et al. Place lab: Device positioning using radio beacons in the wild. In *Pervasive computing*, pages 116–133. Springer, 2005.

[94] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150, 2010.

[95] Oscar D Lara and Miguel A Labrador. A survey on human activity recognition using wearable sensors. *Communications Surveys & Tutorials, IEEE*, 15(3):1192–1209, 2013.

[96] Óscar D Lara, Alfredo J Pérez, Miguel A Labrador, and José D Posada. Centinela: A human activity recognition system based on acceleration and vital sign data. *Pervasive and mobile computing*, 8(5):717–729, 2012.

[97] André de Assis Lauria, Tony Meireles dos Santos, Paulo Roberto dos Santos Amorim, Fábio Antonio Damasceno Marques, and Jorge Roberto Perrout de Lima. Prediction of baseline heart rate of individuals with different levels of habitual physical activity. *Revista Brasileira de Medicina do Esporte*, 19(1):22–26, 2013.

[98] Juha K Laurila, Daniel Gatica-Perez, Imad Aad, Olivier Bornet, Trinh-Minh-Tri Do, Olivier Dousse, Julien Eberle, Markus Miettinen, et al. The mobile data challenge: Big data for mobile computing research. In *Pervasive Computing*, number EPFL-CONF-192489, 2012.

[99] Jonathan Lester, Tanzeem Choudhury, and Gaetano Borriello. A practical approach to recognizing physical activities. In *Pervasive Computing*, pages 1–16. Springer, 2006.

[100] Jonathan Lester, Tanzeem Choudhury, Nicky Kern, Gaetano Borriello, and Blake Hannaford. A hybrid discriminative/generative approach for modeling human activities. In *IJCAI*, volume 5, pages 766–772. Citeseer, 2005.

[101] Quannan Li, Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, and Wei-Ying Ma. Mining user similarity based on location history. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, page 34. ACM, 2008.

[102] Hong-Yuan Mark Liao, Duan-Yu Chen, Chih-Wen Sua, and Hsiao-Rang Tyan. Real-time event detection and its application to surveillance systems. In *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, pages 4–pp. IEEE, 2006.

[103] Lin Liao, Donald J Patterson, Dieter Fox, and Henry Kautz. Learning and inferring transportation routines. *Artificial Intelligence*, 171(5):311–331, 2007.

[104] Bin Liu, Yanjie Fu, Zijun Yao, and Hui Xiong. Learning geographical preferences for point-of-interest recommendation. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1043–1051. ACM, 2013.

[105] Yunhao Liu, Yiyang Zhao, Lei Chen, Jian Pei, and Jinsong Han. Mining frequent trajectory patterns for activity monitoring using radio frequency tag arrays. *Parallel and Distributed Systems, IEEE Transactions on*, 23(11):2138–2149, 2012.

[106] Clemens Lombriser, Nagendra B Bharatula, Daniel Roggen, and Gerhard Tröster. On-body activity recognition in a dynamic sensor network. In *Proceedings of the ICST 2nd international conference on Body area networks*, page 17. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007.

[107] Hong Lu, Wei Pan, Nicholas D Lane, Tanzeem Choudhury, and Andrew T Campbell. Soundsense: scalable sound sensing for people-centric applications on mobile phones. In *Proceedings of the 7th international conference on Mobile systems, applications, and services*, pages 165–178. ACM, 2009.

[108] Mitja Luštrek and Boštjan Kaluža. Fall detection and activity recognition with machine learning. *Informatica*, 33(2), 2009.

[109] Andrea Mannini and Angelo Maria Sabatini. Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, 10(2):1154–1175, 2010.

[110] Jani Mäntyjärvi, Johan Himberg, and Tapio Seppänen. Recognizing human motion with multiple acceleration sensors. In *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, volume 2, pages 747–752. IEEE, 2001.

[111] Uwe Maurer, Asim Smailagic, Daniel P Siewiorek, and Michael Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In *Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. International Workshop on*, pages 4–pp. IEEE, 2006.

[112] Georgios Meditskos, Efstratios Kontopoulos, and Ioannis Kompatsiaris. Knowledge-driven activity recognition and segmentation using context connections. In *The Semantic Web–ISWC 2014*, pages 260–275. Springer, 2014.

[113] Mario Medvedovic and Siva Sivaganesan. Bayesian infinite mixture model based clustering of gene expression profiles. *Bioinformatics*, 18(9):1194–1206, 2002.

[114] Emiliano Miluzzo, Nicholas D Lane, Kristóf Fodor, Ronald Peterson, Hong Lu, Mirco Musolesi, Shane B Eisenman, Xiao Zheng, and Andrew T Campbell. Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 337–350. ACM, 2008.

[115] Michael D Morse and Jignesh M Patel. An efficient and accurate method for evaluating time series similarity. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 569–580. ACM, 2007.

[116] Peter Müller and Fernando A Quintana. Nonparametric bayesian data analysis. *Statistical science*, pages 95–110, 2004.

[117] Emmanuel Munguia Tapia. *Using machine learning for real-time activity recognition and estimation of energy expenditure*. PhD thesis, Massachusetts Institute of Technology, 2008.

[118] Bijan Najafi, Kamiar Aminian, Anisoara Paraschiv-Ionescu, François Loew, Christophe J Büla, and Philippe Robert. Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. *Biomedical Engineering, IEEE Transactions on*, 50(6):711–723, 2003.

[119] Le T Nguyen, Patrick Tague, Ming Zeng, and Joy Zhang. Superad: Supervised activity discovery. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 1463–1472. ACM, 2015.

[120] Juan Carlos Niebles, Hongcheng Wang, and Li Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. *International journal of computer vision*, 79(3):299–318, 2008.

[121] Daniel Olguin Olguin, Benjamin N Waber, Taemie Kim, Akshay Mohan, Koji Ara, and Alex Pentland. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(1):43–55, 2009.

[122] Nuria Oliver and Eric Horvitz. A comparison of hmms and dynamic bayesian networks for recognizing office activities. In *User Modeling 2005*, pages 199–209. Springer, 2005.

[123] Peter Orbanz and Yee Whye Teh. Bayesian nonparametric models. In *Encyclopedia of Machine Learning*, pages 81–89. Springer, 2011.

[124] Julien Pansiot, Danail Stoyanov, Douglas McIlwraith, Benny PL Lo, and Guang-Zhong Yang. Ambient and wearable sensor fusion for activity recognition in healthcare monitoring systems. In *4th international workshop on wearable and implantable body sensor networks (BSN 2007)*, pages 208–212. Springer, 2007.

[125] Juha Pärkkä, Miikka Ermes, Panu Korpipää, Jani Mäntyjärvi, Johannes Peltola, and Ilkka Korhonen. Activity classification using realistic data from wearable sensors. *Information Technology in Biomedicine, IEEE Transactions on*, 10(1):119–128, 2006.

[126] Donald J Patterson, Lin Liao, Dieter Fox, and Henry Kautz. Inferring high-level behavior from low-level sensors. In *UbiComp 2003: Ubiquitous Computing*, pages 73–89. Springer, 2003.

[127] Jan Petzold, Faruk Bagci, Wolfgang Trumler, and Theo Ungerer. Next location prediction within a smart office building. *Cognitive Science Research Paper-University of Sussex CSRP*, 577:69, 2005.

[128] David Pinto, José-Miguel Benedí, and Paolo Rosso. Clustering narrow-domain short texts by using the kullback-leibler distance. In *Computational Linguistics and Intelligent Text Processing*, pages 611–622. Springer, 2007.

[129] Susanna Pirttikangas, Kaori Fujinami, and Tatsuo Nakajima. Feature selection and activity recognition from wearable sensors. In *Ubiquitous Computing Systems*, pages 516–527. Springer, 2006.

[130] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking*, pages 27–38. ACM, 2013.

[131] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. Activity recognition from accelerometer data. In *AAAI*, volume 5, pages 1541–1546, 2005.

[132] Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark Hansen, and Mani Srivastava. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)*, 6(2):13, 2010.

[133] Kristin Rieping, Gwenn Englebienne, and Ben Kröse. Behavior analysis of elderly using topic models. *Pervasive and Mobile Computing*, 15:181–199, 2014.

[134] Daniel Roetenberg, Henk Luinge, and Per Slycke. Xsens mvn: full 6dof human motion tracking using miniature inertial sensors. *Xsens Motion Technologies BV, Tech. Rep*, 2009.

[135] Albrecht Schmidt, Hans W Gellersen, and Michael Beigl. A wearable context-awareness component. In *iswc*, page 176. IEEE, 1999.

[136] Jayaram Sethuraman. A constructive definition of dirichlet priors. *Statistica sinica*, pages 639–650, 1994.

[137] Burr Settles. Active learning literature survey. *University of Wisconsin, Madison*, 52(55-66):11, 2010.

[138] Takaaki Shiratori and Jessica K Hodgins. Accelerometer-based user interfaces for the control of a physically simulated character. In *ACM Transactions on Graphics (TOG)*, volume 27, page 123. ACM, 2008.

[139] Maja Stikic, Diane Larlus, Sandra Ebert, and Bernt Schiele. Weakly supervised recognition of daily life activities with wearable sensors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(12):2521–2537, 2011.

[140] Maja Stikic, Diane Larlus, and Bernt Schiele. Multi-graph based semi-supervised learning for activity recognition. In *Wearable Computers, 2009. ISWC'09. International Symposium on*, pages 85–92. IEEE, 2009.

[141] Xing Su, Hanghang Tong, and Ping Ji. Activity recognition with smartphone sensors. *Tsinghua Science and Technology*, 19(3):235–249, 2014.

[142] Amarnag Subramanya, Alvin Raj, Jeff A Bilmes, and Dieter Fox. Recognizing activities and spatial context using wearable sensors. *arXiv preprint arXiv:1206.6869*, 2012.

[143] Emmanuel Munguia Tapia, Stephen S Intille, William Haskell, Kent Larson, Julie Wright, Abby King, and Robert Friedman. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *Wearable Computers, 2007 11th IEEE International Symposium on*, pages 37–40. IEEE, 2007.

[144] Emmanuel Munguia Tapia, Stephen S Intille, and Kent Larson. *Activity recognition in the home using simple and ubiquitous sensors*. Springer, 2004.

[145] Konstantinos Tarabanis, Peter K Allen, Roger Y Tsai, et al. A survey of sensor planning in computer vision. *Robotics and Automation, IEEE Transactions on*, 11(1):86–104, 1995.

[146] Yee Whye Teh and Michael I Jordan. Hierarchical bayesian nonparametric models with applications. *Bayesian nonparametrics*, 1, 2010.

[147] Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical dirichlet processes. *Journal of the american statistical association*, 101(476), 2006.

[148] Can Tunca, Hande Alemdar, Halil Ertan, Ozlem Durmaz Incel, and Cem Ersoy. Multimodal wireless sensor network-based ambient assisted living in real homes with multiple residents. *Sensors*, 14(6):9692–9719, 2014.

[149] JA Unar, Woo Chaw Seng, and Almas Abbasi. A review of biometric technology along with trends and prospects. *Pattern recognition*, 47(8):2673–2688, 2014.

[150] Tim Van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. Accurate activity recognition in a home setting. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 1–9. ACM, 2008.

[151] Michail Vlachos, George Kollios, and Dimitrios Gunopulos. Discovering similar multidimensional trajectories. In *Data Engineering, 2002. Proceedings. 18th International Conference on*, pages 673–684. IEEE, 2002.

[152] Jue Wang, Deepak Vasisht, and Dina Katabi. Rf-idraw: virtual touch screen in the air using rf signals. In *ACM SIGCOMM Computer Communication Review*, volume 44, pages 235–246. ACM, 2014.

[153] Shuangquan Wang and Gang Zhou. A review on radio based activity recognition. *Digital Communications and Networks*, 1(1):20–29, 2015.

[154] Xuyu Wang, Lingjun Gao, Shiwen Mao, and Santosh Pandey. Deepfi: Deep learning for indoor fingerprinting using channel state information. In *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*, pages 1666–1671. IEEE, 2015.

[155] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 617–628. ACM, 2014.

[156] Yi Wang, Xinli Jiang, Rongyu Cao, and Xiyang Wang. Robust indoor human activity recognition using wireless signals. *Sensors*, 15(7):17195–17208, 2015.

[157] Bo Wei, Wen Hu, Mingrui Yang, and Chun Tung Chou. Radio-based device-free activity recognition with radio frequency interference. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, pages 154–165. ACM, 2015.

[158] Anthony D Wood, John Stankovic, Gilles Virone, Leo Selavo, Zhimin He, Qiuhua Cao, Thao Doan, Yafeng Wu, Lei Fang, Radu Stoleru, et al. Context-aware wireless sensor networks for assisted living and residential monitoring. *Network, IEEE*, 22(4):26–33, 2008.

[159] Wei Xi, Jizhong Zhao, Xiang-Yang Li, Kun Zhao, Shaojie Tang, Xue Liu, and Zhiping Jiang. Electronic frog eye: Counting crowd using wifi. In *INFOCOM, 2014 Proceedings IEEE*, pages 361–369. IEEE, 2014.

[160] Jiang Xiao, Kaishun Wu, Youwen Yi, Lu Wang, and Lionel M Ni. Pilot: Passive device-free indoor localization using channel state information. In *Distributed computing systems (ICDCS), 2013 IEEE 33rd international conference on*, pages 236–245. IEEE, 2013.

[161] Xiangye Xiao, Yu Zheng, Qiong Luo, and Xing Xie. Finding similar users using category-based location history. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 442–445. ACM, 2010.

[162] Xiangye Xiao, Yu Zheng, Qiong Luo, and Xing Xie. Inferring social ties between users with human location history. *Journal of Ambient Intelligence and Humanized Computing*, 5(1):3–19, 2014.

[163] Jhun-Ying Yang, Jeen-Shing Wang, and Yen-Ping Chen. Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers. *Pattern recognition letters*, 29(16):2213–2220, 2008.

[164] Jie Yin, Qiang Yang, and Jeffrey Junfeng Pan. Sensor-based abnormal human-activity detection. *Knowledge and Data Engineering, IEEE Transactions on*, 20(8):1082–1090, 2008.

[165] Yunze Zeng, Parth H Pathak, and Prasant Mohapatra. Analyzing shopper's behavior through wifi signals. In *Proceedings of the 2nd workshop on Workshop on Physical Analytics*, pages 13–18. ACM, 2015.

[166] Yunze Zeng, Parth H Pathak, Chao Xu, and Prasant Mohapatra. Your ap knows how you move: fine-grained device motion recognition through wifi. In *Proceedings of the 1st ACM workshop on Hot topics in wireless*, pages 49–54. ACM, 2014.

[167] Yongmian Zhang and Qiang Ji. Sensor selection for active information fusion. 2005.

[168] Feng Zhao, Jaewon Shin, and James Reich. Information-driven dynamic sensor collaboration. *Signal Processing Magazine, IEEE*, 19(2):61–72, 2002.

[169] Vincent W Zheng, Yu Zheng, Xing Xie, and Qiang Yang. Towards mobile intelligence: Learning from gps history data for collaborative recommendation. *Artificial Intelligence*, 184:17–37, 2012.

[170] Yu Zheng. Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3):29, 2015.

[171] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. Understanding mobility based on gps data. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 312–321. ACM, 2008.

[172] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on World wide web*, pages 791–800. ACM, 2009.

[173] Chun Zhu and Weihua Sheng. Human daily activity recognition in robot-assisted living using multi-sensor fusion. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 2154–2159. IEEE, 2009.

[174] Wiebren Zijlstra and Kamiar Aminian. Mobility assessment in older people: new possibilities and challenges. *European Journal of Ageing*, 4(1):3–12, 2007.