Carnegie Mellon University

CARNEGIE INSTITUTE OF TECHNOLOGY

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF Doctor of Philosophy

TITLE **Nonparametric Discovery of Human Behavior Patterns from Multimodal Data**

PRESENTED BY **Feng-Tso Sun**

ACCEPTED BY THE DEPARTMENT OF

Electrical and Computer Engineering

Martin Griss — ADVISOR, MAJOR PROFESSOR — 5/1/14 — DATE

Jelena Kovacevic — DEPARTMENT HEAD — 5/1/14 — DATE

APPROVED BY THE COLLEGE COUNCIL

Vijayakumar Bhagavatula — DEAN — 5/1/14 — DATE

Nonparametric Discovery of Human Behavior Patterns from Multimodal Data

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

Feng-Tso Sun

B.S., Electrical Engineering, National Tsing Hua University
M.S., Electrical Engineering, University of California, Los Angeles

Carnegie Mellon University
Pittsburgh, PA

May, 2014

Keywords: Activity recognition, machine learning, topic modeling, nonparametric Bayesian, probabilistic graphical models, context-aware systems, ubiquitous computing, mobile computing, user behavior modeling, sensor data analysis, artificial intelligence.

For my parents and family

Abstract

Recent advances in sensor technologies and the growing interest in context-aware applications, such as targeted advertising and location-based services, have led to a demand for understanding human behavior patterns from sensor data. People engage in routine behaviors. Automatic routine discovery goes beyond low-level activity recognition such as sitting or standing and analyzes human behaviors at a higher level (e.g., commuting to work). The goal of the research presented in this thesis is to automatically discover high-level semantic human routines from low-level sensor streams.

One recent line of research is to mine human routines from sensor data using parametric topic models. The main shortcoming of parametric models is that they assume a fixed, pre-specified parameter regardless of the data. Choosing an appropriate parameter usually requires an inefficient trial-and-error model selection process. Furthermore, it is even more difficult to find optimal parameter values in advance for personalized applications.

The research presented in this thesis offers a novel nonparametric framework for human routine discovery that can infer high-level routines without knowing the number of latent low-level activities beforehand. More specifically, the framework automatically finds the size of the low-level feature vocabulary from sensor feature vectors at the vocabulary extraction phase. At the routine discovery phase, the framework further automatically selects the appropriate number of latent low-level activities and discovers latent routines. Moreover, we propose a new generative graphical model to incorporate multimodal sensor streams for the human activity discovery task.

The hypothesis and approaches presented in this thesis are evaluated on public datasets in two routine domains: two daily-activity datasets and a transportation mode dataset. Experimental results show that our nonparametric framework can automatically learn the appropriate model parameters from multimodal sensor data without any form of manual model selection procedure and can outperform traditional parametric approaches for human routine discovery tasks.

Acknowledgments

Completing my PhD degree is not only a milestone but also the most challenging activity of my life-long journey as a student. I would like to thank many people because this dissertation would not have been possible to complete without their help.

First of all, I would like to express my greatest gratitude to my advisor, Martin Griss, for giving me this opportunity to start the journey of research. Academically and emotionally, he has always attempted to understand our concerns from a student point of view and given his constructive feedback. I really appreciated his support, guidance, and kindness over the past five years.

Next, I would like to thank my thesis committee - Professor Bob Iannucci, Professor Ole Mengshoel, and Dr. Cynthia Kuo. I am grateful for their advice for helping me during the process from the thesis proposal to the defense. I also want to thank my undergraduate advisor in National Tsing Hua University, Professor Yarsun Hsu, without whom I would not be able to come to Carnegie Mellon.

Carnegie Mellon has been a fantastic place for pursuing PhD. Martin has created an amazing group. I would like to thank all the group members. They are not only smart, but also funny and warm. Special thanks to Heng-Tze Cheng, Senaka Buthpitiya, Faisal Lupman, David Huang, Zheng Sun, and Lu Zheng. I would also like to thank my friends in Taiwan for their endless support. Finally, I would like to acknowledge the financial support from Carnegie Mellon University and Nokia Research Center.

I would especially like to acknowledge my parents - Jui-Mau Sun and Jui-Mei Chou. At the first time I brought up the idea to study abroad, they showed their support without a doubt. During the journey of studying abroad, they always asked me not worry about everything back home and focus on my study. I feel lucky to have such selfless parents to support me. My big brother, Feng-Yu Sun, has been always telling me that I should follow my heart whenever I make decisions. I would like to thank him for giving me this valuable advice and it helps me overcome all difficult times without regrets.

Lastly, I would like to thank my dear wife and soulmate, Yi-Ting Yeh. Thank you for always being there for me in incredible ways. With your accompany, I become more brave to try anything new because I know you are the one who always gives me unconditional support. I enjoy every moment with you and look forward to the next chapter of our lives.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Recent advances in mobile phone technologies and the significant growth of sensor deployments make continuous and large-scale sensory data more accessible [64]. For example, in our daily lives, we make phone calls, go for a run, or pay a parking ticket. In such a digital and networked world, these tiny transactional events leave a digital breadcrumb trail captured by sensors in mobile phones, wearable devices, or security cameras [63]. With the massive amount of raw sensor data, we have the opportunity to learn the basic activity units that compose our daily routines and mine the regularities of these routines. The more precisely we can model or reason about our daily routines from raw sensor data, the more proactive context-aware or user-centric mobile applications can be provided. In order to add value to these digital crumbs collected from individuals, we need to understand them. As a result, discovering representations from rich sets of sensory data and capturing high-level abstractions of human behavior patterns play a crucial role beyond data collection.

1.1 Motivation

The idea of recognizing basic human activities or uncovering regular patterns in a human's behaviors from observed sensor data has been explored in several prior works [17, 21]. Many techniques have been proposed to perform low-level activity recognition such as walking, sitting, or opening a door [1, 7, 49, 83]. While low-level activity recognition provides us with building blocks to understand human behaviors, they are not enough to convey higher-level semantic meaning. For example, people usually describe what they did during the day using high-level routines (e.g., "commuting to work" or "having lunch") rather than a sequence of low-level activities such as "walk-walk-stand-sit-sit-run-walk-stand".

Discovering high-level semantic meanings of human activities not only helps understand how people behave and communicate in our daily lives, but also provides the core component for potential context-aware applications, such as detecting changes in activities for elderly care, targeted advertising, and location-based services [32, 57]. For example, human location patterns can be learned from GPS data. These human location patterns can further be used for anomalous activity detection in the remote elder-care monitoring systems [10].

In this thesis, we focus on recognizing high-level human behaviors using statistical learning techniques. We refer to the term "human behavior" as a relatively short, measurable pattern of activity instead of the definition in psychology related to a person's intention or ability. More specifically, human routine discovery is about extracting temporal regularities in people's daily lives [22]. A routine can be seen as a composition of multiple low-level activities. Multiple low-level activities can occur within the same routine. Different routines may contain the same kinds of low-level activities, but with different proportions. For example, the "`Grocery Shopping`" routine may involve more "standing" and "walking" activities compared to the "`Office Work`" routine, in which "sitting" is the major part. The composition of multiple low-level activities in a routine is similiar to a document can contain

Text	Sensor data
topic proportions	low-level activity proportions
topics	low-level activities
documents	sensor data segments (temporal windows)
bag-of-words	bag-of-features
vocabulary	a set of discrete labels of data features
text words	discretized data features

Table 1.1: List of the analogy of terms between text and sensor data.

multiple topics. Based on the common properties of text mining and human activity modeling, we can make an analogy between text and sensor data in the topic modeling paradigm. Table 6.1 lists the analogy of terms between sensor data and text used in this thesis.

Figure 1.1 shows the analogy of sensor data and text modeling using topic models. In the context of mining a sequence of sensor data, sensor data features are first mapped into a set of discrete *labels* (vocabulary). Each mapped data feature becomes a word. Then, the bag-of-features in each temporal window is used to train the topic model. Sensor data segments belong to the same routine if they have similar topic proportions. Most existing approaches for automatic high-level activity discovery are built on the bag-of-words concept using parametric topic models, such as latent Dirichlet allocation (LDA) [22, 35]. The bag of features in each temporal window (document) is used to train the topic model.

1.2 Challenges

There are at least two issues that have not solved in previous parametric topic model-based approaches: First, in the parameteric setting, the aforementioned topic proportion learning procedure requires two types of parameters to be predefined: the size of the vocabulary and the number of latent topics. Typically, they are chosen in a trial-and-error fashion. Figure 1.2 shows the overview of the comparison between the parametric and the nonparametric topic model framework for human routine discovery. For parametric methods, the size of

Figure 1.1: Comparison between topic model-based approach to human routine discovery and text modeling.

vocabulary and the number of topics must be specified in advance. Since the appropriate parameters are difficult to specify initially, searching through the parameter space to achieve the best performance is necessary. Our proposed framework uses nonparametric methods to automatically select the proper size of vocabulary and the number of topics from the data.

For routine discovery, such parameter specification poses several challenges. The best parameter values for personalized models may be different for different users. For example, due to the fact that different people usually have very distinct behavior patterns based on their lifestyles, gender, marital status, job types, or ages, their routine patterns may require a significantly different number of latent topics to model appropriately. Moreover, even for a single user, it is possible that his/her behavior patterns change over time. The best parameter values must also be adjusted accordingly. Hence, we need the model to automatically select parameter values based on individual users' behavior patterns.

Second, these previous topic model-based approaches incorporate only a single sensor modality (e.g., accelerometer or GPS) or use simple concatenation from different modalities [22]. They do not take into account informative relationships among different modalities of sensor data. For example, accelerometer data can help recognize different activities at a single location shown by closely related GPS readings. Conversely, GPS readings can tell if you are driving or sitting in front of the computer that a motion sensor alone cannot distinguish because sitting occurs most of time in two different activities. Given the limitations of the existing approaches, this thesis aims to solve two research problems:

Problem 1: How to discover human behavior patterns and automatically learn model parameters from data?

Problem 2: Given a sequence of multimodal data, how to model the complex relationships among different modalities of data with nonparametric settings for human behavior discovery?

In this thesis, we propose a novel human routine discovery framework using nonparametric

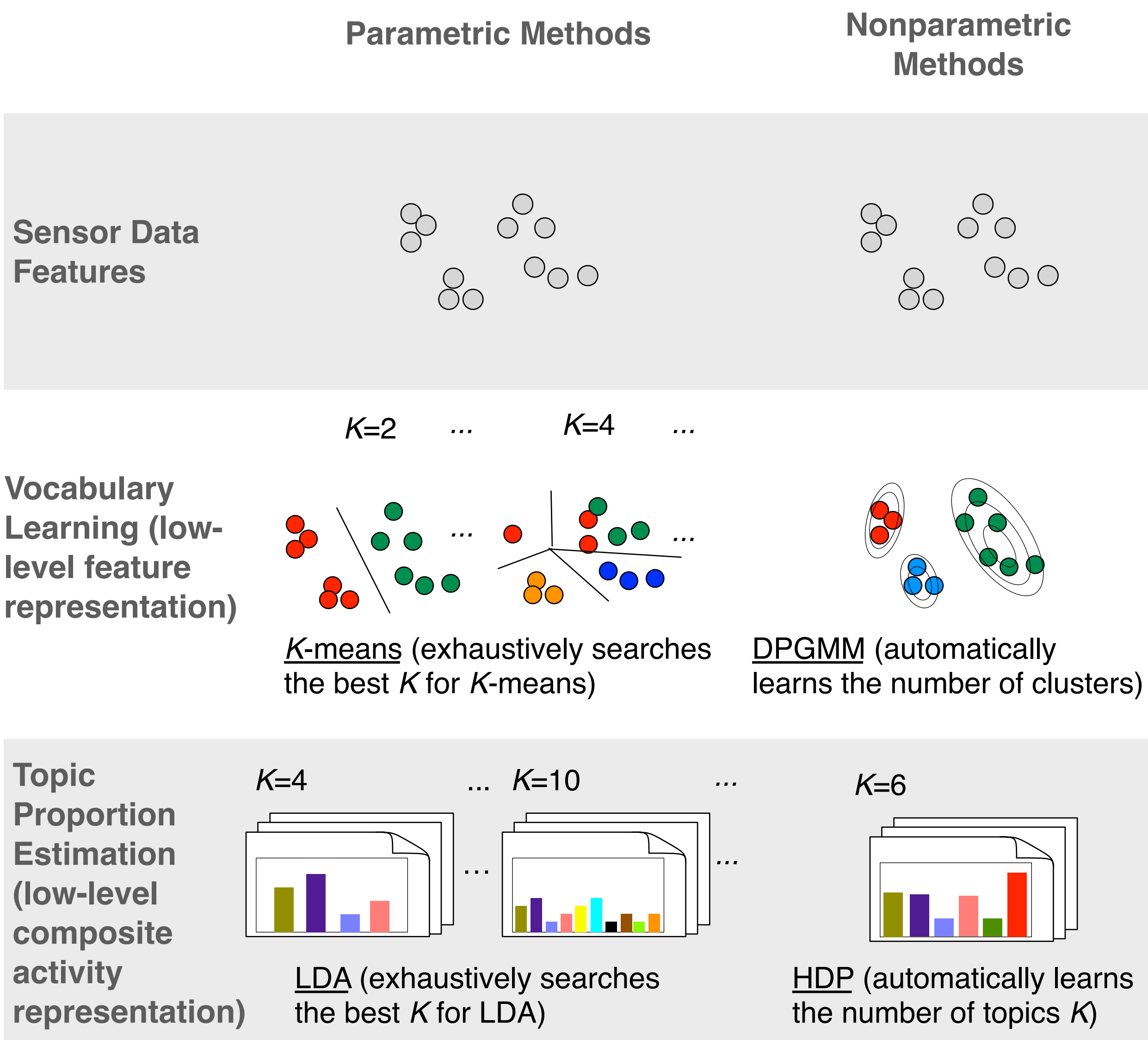


Figure 1.2: Comparison of parametric (e.g., K-means or latent Dirichlet allocation [LDA]) and nonparametric methods (e.g., Dirichlet process Gaussian mixture model [DPGMM] and hierarchical Dirichlet process [HDP]) for routine discovery.

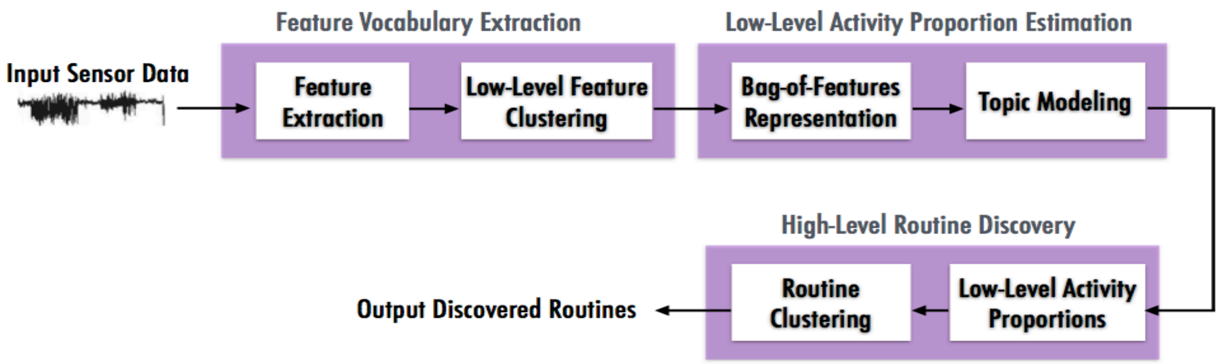


Figure 1.3: Overview of our two-phase nonparametric routine discovery framework.

Bayesian methods. The goal is to address two aforementioned research questions.

Our framework consists of two major phases: vocabulary extraction and routine extraction shown in Figure 1.3. During the first phase, we build up the feature vocabulary and automatically determine the size of vocabulary (low-level feature representation) from raw sensor data to create a vocabulary and provide the bag-of-features representation for each temporal window. In the second phase, we infer low-level activity proportions for each sensor data segment with the automatically determined number of latent low-level activities and extract latent routines. Moreover, in order to take advantage of the informative multimodal data, we propose a graphical model to capture the relationships among multimodal data sources for high-level routine discovery. The details of the proposed framework will be presented in Chapter 4 and Chapter 5.

1.3 Contributions

This thesis makes contributions to human behavior discovery in the field of mobile computing and to Bayesian nonparametric topic modeling in the field of statistical machine learning. The main contributions of this thesis are summarized as follows:

- The design and implementation of a nonparametric framework for high-level human behavior discovery without the need of a trial-and-error model selection process.
- The novel representation and automatic generation of sensor data vocabulary for topic modeling using nonparametric clustering methods.
- The design and implementation of a probabilistic graphical model that models the relationships among a sequence of multimodal sensor data for human behavior discovery.
- The evaluation of the proposed framework on three real-world public datasets in human activity and transportation mode domains.

1.4 Thesis Organization

This thesis is organized as follows:

Chapter 2 - Related Work provides the overview of related work in the areas of human activity recognition and discovery and topic modeling for text mining. The main design considerations such as activity granularity, sensing modality, and learning paradigm in the previous work of human behavior modeling are reviewed.

Chapter 3 - Background of Nonparametric Bayesian Modeling introduces the probabilistic graphical representation of topic models. The background of the Dirichlet process, one of the most widely used nonparametric Bayesian distributions, and its realization algorithms are discussed. The details of basic probabilistic topic models such as latent Dirichlet allocation (LDA) and hierarchical Dirichlet process (HDP) are also introduced.

Chapter 4 - Nonparametric Discovery of Human Behavior presents a novel nonparametric framework for human behavior discovery from sensor data. We describe two major phases in the framework for automatic low-level data feature vocabulary creation and high-level routine discovery. Moreover, experimental ressults on two real-world datasets of daily life activity and transportation mode domains are also described.

Chapter 5 - Towards Multimodal Discovery of Human Behavior presents the design, implementation, and evaluation of a probabilistic graphical model (i.e., multimodal hierarchical Dirichlet process [MHDP]) for modeling the complex relationships among multiple modalities of sensor data for human behavior discovery.

Chapter 6 - Conclusion and Future Work summarizes the contributions and limitations of work in this thesis. Based on this research, future research directions and challenges on human routine discovery from sensor data are described.

Chapter 2

Related Work

The research reported in this thesis draws heavily on the intersection of two existing research areas. One is human activity discovery in the field of context-aware computing. The other is topic modeling in the field of text mining. In this chapter, we will review the most relevant previous work in these two research areas. More specifically, we will first introduce three common modules used in human activity recognition and discovery frameworks: 1) the sensing module, 2) the feature extraction module, and 3) the statistical inference module. Then, we will discuss three design choices commonly considered in activity discovery research: 1) activity granularity, 2) sensing modality, and 3) learning paradigm. Following that, we will review the concept of topic modeling and how it was previously applied in activity discovery research. Lastly, we will discuss limitations of existing approaches.

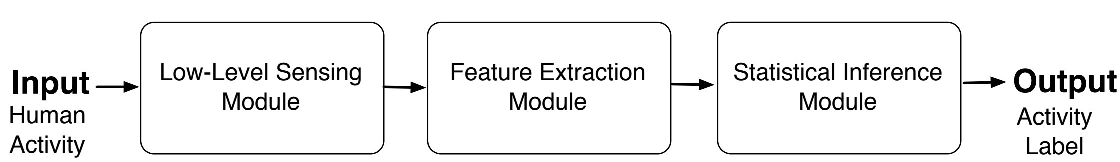


Figure 2.1: An overview of basic components for human activity recognition.

2.1 Activity Recognition and Discovery

The goal of activity recognition and discovery is to recognize human activity patterns in daily life settings [39]. Figure 2.1 shows an overview of a general human activity recognition framework. The framework consists of three major modules: the low-level sensing module, the feature extraction module, and the statistical inference module.

The sensing module measures signals that are associated with human activities. For example, different activities result in different acceleration signals. Running causes repetitive changes in acceleration values while sitting causes fixed acceleration values. Therefore, one common approach is to attach accelerometers on the human body and record the acceleration signals in x, y, and z axes.

Next, the feature extraction module converts raw sensor data to a set of features, represented as feature vectors. For example, we can calculate the mean and variance of acceleration signals every second along each axis as features. Then, we can concatenate the features within the same time window into a multidimensional feature vector.

Finally, the statistical inference module learns a mapping from feature vectors to an activity label. There are two types of learning tasks. For an activity recognition task, we have a predefined set of activity classes to map to. On the other hand, for an activity discovery task, we do not have a predefined set of activity classs. Instead, an inference module discovers patterns in observed feature vectors and segment them into clusters.

2.2 Design Choices in Activity Recognition and Discovery

Figure 2.2 shows a conceptual three dimensional design space for activity recognition research. Each dimension corresponds to a design choice. First, what is the granularity of

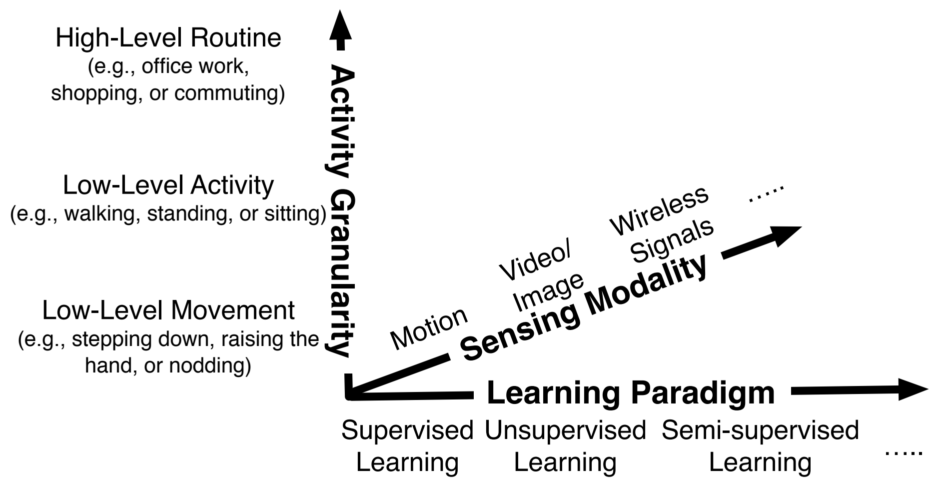


Figure 2.2: A conceptual three dimensional design space (i.e., activity granularity, sensing modality, and learning paradigm) for the activity recognition research.

activities we would like to recognize? Second, what signals are we capturing? Third, what learning paradigm to use for inference? In this section, we review each of the design choices explored in existing activity recognition and discovery research.

2.2.1 Activity Granularity

In the field of human activity recognition and discovery, the definition of "human activity" varies in the literature. In this thesis, we adopt the definition of physical activities proposed by Huynh [34], where physical activities are grouped in three different granularities: (1) low-level movements, (2) low-level activities, and (3) high-level routines. Next, we will describe how physical activities are categorized into these granularity levels based on the observed duration and complexity.

Low-Level Movement

Low-level movements usually happen on a timescale of seconds or less. Examples of low-level movements include raising the forearm and climbing a step. This kind of low-level movement recognition is used in the applications of fall detection in elder care [46] and hand gesture recognition in human computing interaction [47].

Low-Level Activity

For a sequence of low-level movements on a timescale of minutes, we consider it to be a low-level activity. For example, walking consists of a regular sequence of low-level movements that are stance and swing movements. Many techniques have been proposed to perform low-level activity recognition such as walking, sitting, or opening a door. This is performed using various kinds of sensor data such as motion data [1], GPS/Bluetooth/WiFi signals [83], ambient sound [49], and RFID-tagged objects [7].

High-Level Routine

Beyond low-level activity recognition, we consider a collection of low-level activities as a high-level routine. High-level routines typically last longer, with durations from several minutes to a few hours. Extracting routines (e.g., dining, office work, or taking a bus) has received attention because routine information provides high-level semantics for understanding human behaviors. Eagle et al. [21] use principal component analysis (PCA), a general-purpose dimensionality reduction method, to obtain main components that comprise human daily routines.

2.2.2 Sensing Modality

Many sensing techniques have been proposed to recognize different types of activities. Researchers have extensively explored two major types of approaches: One is to use on-body wearable sensors, and another is to use ambient sensors installed in the environment for activity recognition [43]. For the on-body sensing approach, the wearable sensors are usually positioned on the human body to recognize activities of interest. Ambient sensors such as cameras or microphones are used to monitor and analyze the activities of users in a space such as smart homes or offices.

Wearable Sensors

For basic body movement classification such as walking, running, standing, and climbing stairs, simple tri-axial accelerometers have been used and placed on the body to capture body motions and recognize activities of interest [53, 84]. With the advance of mobile phone technology, more and more researchers have leveraged the sensing power of a smart phone's on-board sensors (e.g., accelerometer, compass, and gyroscope) and used the phone as a wearable sensor to recognize gestures and motion patterns [26, 42]. As an increasing number of sensor types are available, researchers have gathered and fused multimodal sensor data streams for the activity recognition task. Choudhury et al. conducted pioneering work in activity recognition using a wearable multi-sensor board called Sociometer [17]. The Sociometer contains an accelerometer, a microphone, and IR sensor to capture human interaction activities within a group of people. As many physiological sensors become miniaturized, they also provide the opportunity for continuous physiological signal monitoring and recognition. Sun et al. combined motion data and electrocardiogram (ECG) data from a wearable sensor to develop a continuous stress monitoring system excluding the effects of physical activity [72]. Saponas et al. developed a muscle sensing technique to recognize gestures using an electromyogram (EMG) sensor in a form of an armband [67].

Ambient Sensors

Ambient sensors include video cameras, microphones, and wireless signal transceivers (e.g., WiFi, infrared, and radio-frequency signals). They are usually pre-installed in the environment where participants' activity patterns can be recorded and analyzed. There has been extensive research on activity recognition using cameras once a conventional 2D camera became inexpensive, with the video sequence containing rich information. For example, video collected from smart home residents can be analyzed to recognize activities [66, 77]. As depth cameras have become widely available, researchers have been able to recognize more complex activities using depth informantion [62]. The signature of wireless signals such as Bluetooth and WiFi was also used for indoor activity pattern recognition [83]. For other activities that are not be able to be classified with body movements alone, researchers have observed the pattern of an individual's interaction with objects in a space such as the kitchen and the living room. For example, objects such as refrigerators, doors, and cookwares are tagged with motion sensors or RFID tags and the objects' usage patterns are recorded [18, 80].

2.2.3 Learning Paradigm

Learning paradigm is a critical step to model observed data and provide activity inference results. In the field of activity recognition and discovery, researchers have developed and applied several machine learning paradigms such as supervised learning, semi-supervised learning, unsupervised learning, or zero-shot learning to recognize human activities given different problem settings and various types of sensor data. Cheng presented a table of learning paradigms in related work of human activity recognition [13], which serves as a complete summary and comparison of this area of research. In the following subsections, we will briefly introduce several different problem settings and how different statistical learning methods aim to recognize human activities. At the same time, we will review related work of different learning paradigms in the field of activity recognition and discovery.

Supervised Learning

In the supervised learning setting, the learning task is to infer a mapping function from labeled training data to predict the labels of unseen testing data. Supervised learning has been extensively used in human activity recognition research [1, 70]. More specifically, discriminative classifiers (e.g. Support Vector Machine) and generative models (e.g. Hidden Markov Model) are trained on a large set of labeled samples of every target activity [50, 54, 80]. However, one well-known drawback of supervised learning in human activity recognition is that the collection of labeled samples is usually time consuming and requires lots of effort in human annotation for ground truth labels [70, 71].

Semi-Supervised Learning

The design of semi-supervised learning is to eliminate the aforementioned limitation that occurs during labeled data collection in supervised learning approach. Semi-supervised learning makes use of the large amount of unlabeled data with the limited amount of labeled data in the training stage. More specifically, semi-supervised learning improves recognition performance by assigning high-confidence predicted labels to unlabeled samples. Previous work has shown the promise of applying semi-supervised learning to the human activity recognition task [48, 51, 58, 71].

Zero-shot Learning

Zero-shot learning aims to learn a classifier that can predict new classes that do not appear in the training data [28, 37]. This situation occurs when we predict a new activity class with a classifier trained on data of an exclusive set of activity classes. In order to predict new activity classes, a semantic attribute representation is introduced to describe basic elements characterizing an activity class. Recently, zero-shot learning was successfully introduced and used in human activity recognition to model unseen new activity classes [14, 15].

Unsupervised Learning

Unsupervised learning focuses on performing pattern discovery and clustering based on the similarity of observed samples without providing any ground truth labels [35, 56]. Sometimes, the ground truth labels are not available or the ground truth classes are hard to defined, but we are still interested in discovering similar activity samples. In contrast to activity recognition which identifies the activity class given a new observed sample, activity discovery finds similiar activity segments in terms of raw sensor data signatures in an unsupervised learning fashion. Based on the similarity of sensor data samples, an automatic activity discovery system can detect the boundary of sensor data sequences, segment sensor data samples, and cluster similiar samples together [16, 35].

2.3 Topic Modeling for Text Mining

Topic modeling is a method to model, analyze, and manage a large amount of unlabeled data. A topic model is a generative model describing the relationship between a group of observed and latent random variables. Moreover, it specifies a probabilistic procedure to generate topics [65]. Topic models were initially developed to model a collection of documents using a mixture of a number of topics where a topic is a multinomial distribution over the vocabulary [5]. In a topic model for text mining, a document is a mixture of a number of hidden topics which can be represented by a multinomial topic proportion [5].

2.4 Topic Modeling for Routine Discovery

The same idea of topic modeling can be applied to mine high-level human routine from sensor data. We can make an analogy between text and sensor data. In the context of mining a sequence of sensor data, sensor data features are first mapped into a set of discrete *labels* (vocabulary). Each mapped data feature is equivalent to a text word. Then, the bag-of-features in each temporal window (document) is used to train the topic model. Sensor data segments belong to the same routine if they have similar topic proportions. More specically, routine discovery is about extracting temporal regularities in peoples' daily lives [22]. A routine can be seen as a composition of multiple low-level activities. Multiple low-level activities can occur within the same routine. Different routines may contain the same kinds of low-level activities, but with different proportions. For example, the "`Grocery Shopping`" routine may invovle more "standing" and "walking" activities as compared to the "`Office Work`" routine.

Most existing approaches for automatic routine discovery are built on parametric topic models such as latent Dirichlet allocation (LDA) [22, 35]. In the parametric setting, the above procedure requires two types of parameters to be predefined: the size of vocabulary and the number of latent topics. Typically, they are chosen in a trial-and-error fashion [35][22]. However, for routine discovery, such parameter specification poses several challenges. First of all, the best parameter values for personalized models may be different for different users. For example, due to the fact that different people usually have very distinct behavior patterns based on their lifestyles, job types, or ages, their routine patterns may require different number of latent topics to model appropriately. Moreover, even for a single user, it is possible that her behavior patterns change over time. The best parameter values must also be adjusted accordingly. Hence, we need the model to automatically select parameter values based on individual users' behavior patterns.

Recently, the concept of nonparametric methods has shown promise in the field of mobile computing. For example, Hu et al. [33] solve low-level abnormal activity recognition problem by using hierarchical Dirichlet process hidden Markov model (HDP-HMM) to automatically decide the right number of states for HMM. Similarly, Zhu et al. [87] segment a small number of activities using HDP-HMM. Nguyeh et al. [61] apply HDP model to extract users' proximity patterns from sociometric badge data.

2.5 Conclusion

In this chapter, we reviewed the related work of activity recognition and discovery. We first introduced the common modules in human activity recognition and discovery frameworks. More, we reviewed three design choices commonly considered in activity discovery research: 1) activity granularity, 2) sensing modality, and 3) learning paradigm. Finally, we described how parametric topic models are used for human routine discovery in previous work and discuss the limitations of existing approaches.

In this thesis, we extend previous work by proposing a new nonparametric framework for human routine discovery that can construct low-level activity primitives and extract high-level routines from raw sensor data without the need of model selection procedures based on the topic modeling approach.

Chapter 3

Background of Nonparametric Bayesian Modeling

In the previous chapter, we reviewed the related work of activity recognition research and discussed the design choices commonly considered in the development of activity recognition and discovery frameworks. In this chapter, we will review the background of nonparametric Bayesian modeling by introducing the concept of Dirichlet distribution, Dirichlet process (DP), and Dirichlet process mixture model (DPMM). Then, we will cover parametric and nonparametric topic models such as the latent Dirichlet allocation (LDA) and hierarchical Dirichlet process (HDP) used in our human routine discovery framework.

3.1 Nonparametric Bayesian Modeling

Researchers regularly have the situation that they need to decide appropriate model parameters while exploring and modeling their data [25]. For example, the number of clusters should be chosen in the mixture model or the number of factors should be prespecified in the factor analysis. Traditionally, researchers tackle this problem by a trial-and-error model selection procedure. They first fit several candidate models and then select the most appropriate one based on model evaluation metrics [81]. In constrast to the traditional model selection procedure, nonparametric Bayesian methods provide a different approach to this problem [30]. Nonparametric Bayesian methods allow the model complexity to change as more data become available and automatically select parameters to fit a single model.

In the following subsections, we will first describe statistical foundations in order to better understand and develop nonparametric Bayesian models. Then, we will detail an application of nonparametric methods that is built on a mixture model by beginning with a finite mixture model and extending to its infinite version such as the Dirichlet process mixture model.

3.1.1 Dirichlet Distribution

A Dirichlet distribution is a distribution over multinomial distributions. It is parameterized by a vector of $\{\alpha_1, \ldots, \alpha_m\}$. A random vector $(\pi_1, \ldots \pi_m)$ $(\sum_{k=1}^{m} \pi_k = 1)$ is Dirichlet distributed if

$$P(\pi_1, \ldots \pi_m) = \frac{\Gamma(\Sigma_k \alpha_k)}{\prod_k (\Gamma(\alpha_k))} \prod_{k=1}^{m} \pi_k^{(\alpha_k - 1)} \tag{3.1}$$

where $\alpha_1, \ldots, \alpha_m > 0$. The support of an m-dimensional Dirichlet distribution is the $(m-1)$-dimensional probability simplex.

The random vector $(\pi_1, \ldots \pi_m) \sim Dirichlet(\alpha_1, \ldots, \alpha_m)$ defines the possible parameters for a multinomial distribution on the discrete space $\Theta = \theta_1, ..., \theta_m$ such that $P(\theta = \theta_i) = \pi_i$. From Eq (3.1), we see that, if $(\alpha_1, \ldots, \alpha_m) = (1, \ldots, 1)$, the distribution of $(\pi_1, \ldots \pi_m)$ is uniform.

3.1.2 Dirichlet Process

The Dirichlet process was first developed by Ferguson [23]. The Dirichlet process is an infinite-dimensional generalization of the Dirichlet distribution. Like the Dirichlet distribution, the Dirichlet process is also a distribution over distributions. A Dirichlet process can be denoted as $G \sim DP(\alpha, G_0)$, where $\alpha > 0$ is the concentration parameter, G_0 is the base distribution, and G is a random distribution drawn from a Dirichlet process. Mathematically, G can be written as:

$$G = \sum_{k=1}^{\infty} \pi_k \delta(\theta = \theta_k) \tag{3.2}$$

In the equation 3.2, G is the random distribution drawn from the Dirichlet process and it places its probability mass on a countably infinite collection of points called "atoms" [23]. π_k is the probability assigned to the kth atom and $\delta(\theta = \theta_k)$ is a Dirac delta function at θ_k which is the location of the atom.

Given a finite set of partitions of Θ, $\theta_1 \cup \ldots \cup \theta_K = \Theta$, if a random distribution measure $G \sim DP(\alpha, G_0)$, then each partition of Θ is Dirichlet distributed. Mathematically, it can be written as

$$(G(\theta_1), ..., G(\theta_K)) \sim Dirichlet(\alpha G_0(\theta_1), ..., \alpha G_0(\theta_K)) \tag{3.3}$$

The two parameters of the Dirichlet process have the following interpretations [25, 85]:

- The base distribution G_0 is like the mean of the Dirichlet process, where $\mathbb{E}[G(\theta)] = G_0(\theta)$. If we draw a random distribution G from a Dirichlet process and sum up the probability mass in the partition $\theta \in \Theta$, the average mass in that partition will be $G_0(\theta)$.

- The concentration parameter α is like the inverse-variance of the Dirichlet process, where $Var[G(\theta)] = \frac{G_0(\theta)(1-G_0(\theta))}{\alpha+1}$. For a higher value of α, the probability mass $G(\theta)$ is more centered around $G_0(\theta)$.

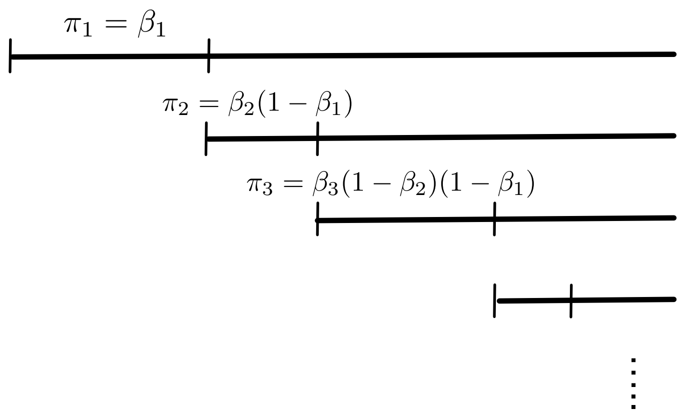


Figure 3.1: An illustration for generating π by breaking a stick of length 1 into segments.

There are several representations to describe and simulate a Dirichlet process. In the following subsections, we will introduce three representations of the Dirichlet process such as the stick-breaking construction, the Pólya urn scheme, and the Chinese restaurant process (CRP) [76].

Stick-Breaking Construction

Recall that a random distribution drawn from a Dirichlet process can be written as $G(\theta) = \sum_{k=1}^{\infty} \pi_k \delta(\theta = \theta_k)$. The infinite sequence of mixture weights π_k can be constructed by the stick-breaking construction [69]. With the stick-breaking construction, a stick of the unit length is divided into an infinite number of segments, with the kth segment's length being π_k. Figure 3.1 illustrates the process of generating π via the stick-breaking construction. More specifically, the construction process follows the steps:

- Consider a stick of unit length. Break the first segment out of the stick according to the proportion β_1, where $\beta_1 \sim Beta(1, \alpha)$, and set $\pi_1 = \beta_1$.
- For the remaining stick, we keep breaking off segments according to the proportion $\beta_k \sim Beta(1, \alpha)$, and set $\pi_k = \beta_k$.

The process of of generating π can be written as follows:

$$\beta_k \sim Beta(1, \alpha)$$
$$\pi_1 = \beta_1$$
$$\pi_k = \beta_k \prod_{i=1}^{k-1} (1 - \beta_i) \text{ for } k = 2, 3, \ldots . \tag{3.4}$$

Pólya Urn Scheme

The Pólya urn scheme is yet another way to produce a sequence of independent and identically distributed (i.i.d.) random variables $\phi_1, \cdots, \phi_n$ distributed according to G, where G is a random distribution drawn from a Dirichlet process [45]. It is equivalent to inferring the posterior predictive distribution of a Dirichlet process. Using the analogy of drawing colored balls from an urn, we can consider a Pólya urn sampling scheme with the following properties:

- Start with an urn with a black ball.
- Balls are drawn with probability proportional to their mass.
- The sole black ball has mass α.
- When the black ball is drawn, put it back and add a ball of a new color according to the base distribution G_0.
- When balls are drawn from the urn (not black), add a ball with the same color.

Mathematically, the procedure to generate the sequence of the i.i.d. random variables $\{\phi_i\}_{i=1}^{\infty}$ as follows:

$$\phi_1 \sim G_0$$
$$P(\phi_n | \phi_{1:n-1}) \sim \frac{\alpha G_0}{\alpha + n - 1} + \frac{\Sigma_{j=1}^{n-1} \delta(\phi_n - \phi_j)}{\alpha + n - 1} \tag{3.5}$$

Chinese Restaurant Process

The Dirichlet process can also be constructed by the Chinese Restaurant Process (CRP) [75]. In a CRP metaphor, the Dirichlet process can be described in the following way:

- A restaurant has an unlimited number of tables θ_k, $k = 1, ...\infty$.
- The first customer comes in and sits at the first empty table.
- Given K tables that are occupied before the i-th customer comes in, the customer can either sit at table k ($k \leq K$) with probability $\propto \frac{n_k}{\alpha+i-1}$ or at a new table $K+1$ with probability $\propto \frac{\alpha}{\alpha+i-1}$, where n_k is the number of customers already sitting at table k.

For a set of random variables $\phi_1, \cdots, \phi_n$ (the table assignment for each customer), the distribution of the random variable ϕ_i conditioned on the previous random variables can can be written as

$$P(\phi_i|\phi_1, ..., \phi_{i-1}, \alpha, G_0) \sim \sum_{k=1}^{K} \frac{n_k}{\alpha+i-1}\delta_{\theta_k} + \frac{\alpha}{\alpha+i-1}\delta_{\theta_{\overline{k}}} \tag{3.6}$$

where $\overline{k}$ represents a new table $K+1$.

3.1.3 Dirichlet Process Mixture Model

So far, we have reviewed the statistical foundations of Dirichlet process and its various representations. Let us now introduce the main application of the Dirichlet process in the context of data clustering using mixture models. In particular, we will provide an overview of the Dirichlet process mixture model (DPMM), a mixture model whose parameters of mixture components have Dirichlet process distributed random measure as a prior.

DPMM generalizes traditional finite mixture models by allowing the number of mixture components to be infinite. Let's start by considering the finite case. In traditional finite mixture models, it is assumed that the number of mixture components is given. For example, let $x_1, \ldots, x_N$ be N observation data points, and they are exchangeable. It means that any order of data points to be observed is equally likely. In a mixture model, a data point x_i is assumed to be drawn from the distribution $p(x) = \Sigma_{k=1}^{K} \pi_k f(x|\theta_k)$ where K is the number of mixture components, π_k is the mixture weight of component k, and $f(x|\theta_k)$ is the mixture component parameterized by θ_k. For example, one common choice is a Gaussian distribution of the mixture components, which is parameterized by mean and variance. The mixture weights sum to one.

In the mixture model problem, given the observation data points, it is convenient to introduce a latent discrete random variable, c_i, associated with each data point. It is often referred to as the indicator variable, whose domain is $\{1, \ldots, K\}$. It specifies which component the corresponding data point belongs to. Therefore, the generative process of the finite mixture model can alternatively be described by:

$$\begin{aligned} &p(c_i = k) = \pi_k \\ &x_i \sim f(x|\theta_{k=c_i}). \end{aligned} \tag{3.7}$$

This describes how each data point x_i has been generated by first sampling the component indicator c_i, and then sampling from the distribution of that mixture component.

Since the mixture weight $\boldsymbol{\pi} = \{\pi_1, ..., \pi_k\}$ is a multinomial distribution, it is convenient to use the Dirichlet distribution as the prior. We can thus use the Dirichlet distribution to construct the finite mixture model by the following steps:

$$\begin{aligned} \theta_{c_i} &\sim G_0 \text{ for } c_i = \{1, \ldots, K\} \\ (\pi_1, \ldots, \pi_K) &\sim Dirichlet(\alpha/K, \ldots \alpha/K) \\ c_i &\sim Multinomial(\pi_1, \ldots \pi_K) \\ x_i &\sim f(x|\theta_{c_i}). \end{aligned} \tag{3.8}$$

where G_0 is the base distribution encoding the prior beliefs about the parameters of the

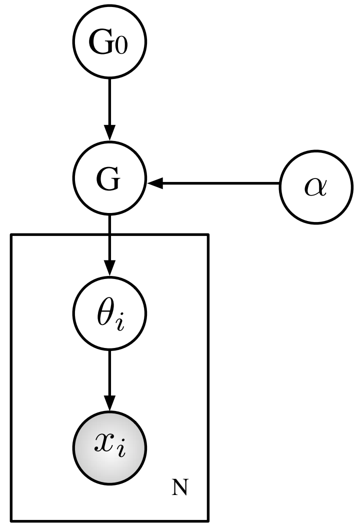


Figure 3.2: Graphical model of the Dirichlet process mixture model.

mixture components.

Next, consider the limiting case where $k \to \infty$, so that the mixture model becomes

$$p(x) = \sum_{k=1}^{\infty} \pi_k f(x|\theta_k). \tag{3.9}$$

The model described in Eq (3.8) then becomes

$$\begin{aligned} G &\sim DP(\alpha, G_0) \\ \theta_i &\sim G \\ x_i &\sim f(x|\theta_i). \end{aligned} \tag{3.10}$$

Here, the Dirichlet distribution becomes the Dirichlet process. Therefore, the infinite mixture model is also called the Dirichlet process mixture model. In DPMM, each component is still described by some set of parameters. These parameters come from the Dirichlet process.

Figure 3.2 illustrates the graphical model of DPMM. First, the prior distribution function G is drawn from a Dirichlet process $G \sim DP(\alpha, G_0)$ where α is the concentration parameter and G_0 is the base prior. Second, given G, we sample θ_i, the parameters for the component that x_i belongs to. Finally, given the parameters θ_i, we generate each data point x_i.

Learning via Gibbs Sampling

To perform the parameter estimation of DPMM, we can use the Gibbs sampling algorithm [60] based on the Chinese restaurant process. It starts by randomly initializing c_i's for all x_i's and then iterates the following steps:

1. Pick a data point x_i.
2. Sample its corresponding indicator variable c_i conditioned on fixing all other indicator variables $\{c_{-i}\}$ using the Chinese restaurant process [4]:

$$\begin{aligned} P(c_i = k, k \leq K | c_{-i}, \alpha) &= \frac{n_k}{\alpha + N - 1} f(x_i|\theta_k) \\ P(c_i = K+1 | c_{-i}, \alpha) &= \frac{\alpha}{\alpha + N - 1} f_{K+1}(x_i) \end{aligned} \tag{3.11}$$

   where N is the total number of data points, n_k is the number of data points being assigned to component k, and $f_{K+1}(x_i) = \int f(x_i|\theta) G_0(\theta) d\theta$. Note that with probability $\frac{\alpha}{(\alpha+N-1)}$, x_i is assigned to a new component $K+1$.

3. If we get a new component, we can draw its corresponding parameter values θ_{K+1} by

$$P(\theta_{K+1}|x_i) \propto f(x_i|\theta_{K+1}) G_0(\theta_{K+1}). \tag{3.12}$$

3.2 Probabilistic Topic Models

Topic modeling is a method to model, analyze, and manage a large amount of unlabeled data. A topic model is a generative model describing the relationship between a group of observed and latent random variables. Moreover, it specifies a probabilistic procedure to generate topics [65]. Topic models were initially developed to model a collection of documents

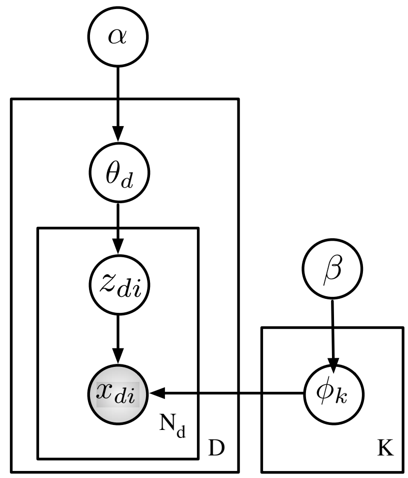


Figure 3.3: Graphical models of the latent Dirichlet allocation.

using a mixture of a number of topics where a topic is a multinomial distribution over the vocabulary [5]. Topic models also have been widely and successfully used for many different types of sensor data rather than text [22, 73].

Given the aforementioned details of the Dirichlet distribution and Dirichlet process, we are able to now introduce two specific topic models that are the latent Dirichlet allocation (LDA) and the hierarchical Dirichlet process (HDP). LDA is considered as a parametric topic model since the assumption about LDA is that the number of topics is assumed known and fixed [3]. However, HDP relaxes the assumption that the number of topics is known and automatically selects the appropriate number of topics based on nonparametric Bayesian methods [76].

3.2.1 Latent Dirichlet Allocation

The intuition of topic modeling is that documents exhibit multiple topics because words in a single document can be related to different topics [3]. For example, an article published at a computational biology conference might consists of words such as "`accuracy`" or "`prediction`" related to topic *data analysis* and words such as "`cell`" or "`gene`" related to topic *biology.* And LDA is one type of statistical mixture model to capture this intuition.

LDA is a generative model that specifies how documents in a large set of texts (text corpus) are generated. In the context of text modeling, we first define a *document* as a collection of words, usually represented as a bag-of-words (BoW) representation. A *topic* is a multinomial distribution over a *vocabulary* of the fixed size $|V|$. We assume we have K latent topics.

Informally, we can describe the generative process of LDA for each document as follows:

1. Randomly choose a distribution over topics (a multinomial distribution of size K)

2. For each word x_{di} in the document

   2.a Given the chosen multinomial distribution in step 1, randomly choose a topic k from the distribution.

   2.b Given the chosen topic in step 2.a, randomly choose a word x_{di} from the corresponding distribution over the vocabulary (a multinomial distribution of size $|V|$).

Figure 3.3 shows the graphical model of LDA, where α is the parameter of the Dirichlet prior on per-document topic distributions, β is the parameter of the Dirichlet prior on per-topic word distributions, θ_d is the topic distribution for document d, ϕ_k is the word distribution for topic k, z_{di} is the topic assignment for the ith word in document d, and x_{di} is the ith observed word in document d. The N_d plate represents the collection of words in documents. Similiarly, the D plate represents the collection of documents and the K plate denotes K latent topics (the multinomial distribution over the vocabulary). The shaded node is the observed words of the documents. The unshaded nodes are latent variables represented in the word generation process.

With the graphical model representation, we can describe the generative process of LDA in details. Two major steps to generate words in each document can be described as follows:

1. Draw a topic distribution $\theta_d \sim Dirichlet(\alpha)$, where $Dirichlet(\alpha)$ is a draw from a

uniform Dirichlet distribution with the scaling parameter α.

2. For each word x_{di} in the document d

   2.a Draw a topic assignment $z_{di} \sim Multinomial(\theta_d)$.

   2.b Given the chosen topic assignment z_{di} in step 2.a, draw a word $x_{di} \sim Multinomial(\phi_{k=z_{di}})$.

3.2.2 Hierarchical Dirichlet Process

In the previous subsection, we discussed the generative process of the topic model that is achieved by first selecting a topic from the topic distribution. Then, choose a word from the word distribution defined by the topic. In LDA, the topic distributions and the word distributions both have the Dirichlet prior. In particular, each document has its own topic distribution over a finite and fixed number of topics.

In hierarchical Dirichlet process (HDP), we relax the assumption that the number of topics is known. Thus, HDP can also be thought of as a nonparametric generalization of the LDA. Figure 3.4 illustrates the graphical model of HDP, which consists of two levels of Dirichlet processes.

Assume that each document is indexed by $d = 1, \ldots, D$ and that x_{di} denotes the ith word in document d. This generative model of HDP can be described as:

$$
\begin{aligned}
&G_0 \sim DP(\gamma, H) \\
&G_d \sim DP(\alpha, G_0) \text{ , for } d = 1, \ldots, D \\
&\theta_{di} \sim G_d \text{ , for } i = 1, \ldots, N_d \\
&x_{di} \sim f(x|\theta_{di}).
\end{aligned} \tag{3.13}
$$

In the upper level, G_0 is the distribution of an infinite mixture of topics for all documents. It is drawn from a Dirichlet process with the base distribution H and the concentration

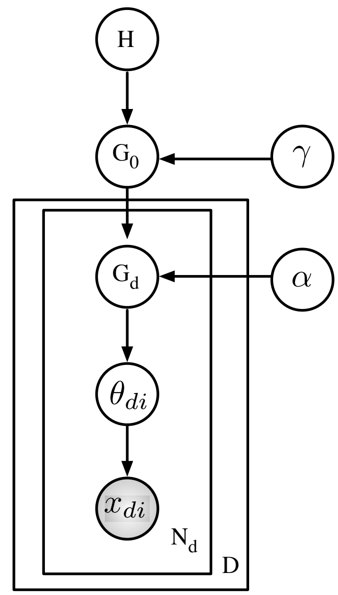


Figure 3.4: Graphical models of the hierarchical Dirichlet process.

parameter γ. In the lower level, G_d defines the mixture of topics of document d. The words are generated by repeatedly drawing from the corresponding topic distribution.

From the graphical model, we can see that the base distribution of G_d is G_0 which is also a random draw from another Dirichlet process. The reason that we need two levels of Dirichlet processes is because we want G_0 to be the common base distribution that is shared across all G_d. In this way, we achieve the goal that different documents share the same mixture components but with different mixture weights.

Recall that in DPMM, the Dirichlet process allows us to avoid pre-specifying the number of mixture components. In HDP, the number of topics activated in G_d also need not to be specified.

3.3 Conclusion

In this chapter, we review the background of nonparametric Bayesian modeling and topic modeling. We first introduce the statistical foundations of nonparametric Bayesian methods and their applications such as Dirichlet process and Dirichlet process mixture model. Then, we discuss the concept of topic modeling and the generative process of topic modeling. At the end, we review two important topic models such as LDA and HDP, that we use going forward. Also, we demonstrate how nonparametric Bayesian methods can be applied in HDP to relax the limitations of LDA.

In the next chapter, we will show how the core idea of a nonparametric Bayesian methods can be applied to human routine discovery. This idea is based on the intuition that human routines can be recognized in a fashion similar to how topics are uncovered from text corpus using a bag-of-words approach [39]. Then, we will detail the experimental results evaluated on two real-world datasets.

Chapter 4

Nonparametric Discovery of Human Behavior

In the previous chapter, we reviewed the background of nonparametric Bayesian modeling and two topic models such as the latent Dirichlet allocation (LDA) and the hierarchical Dirichlet process (HDP). In this chapter, we will emphasize on the discussion of our nonparametric human routine discovery framework. More specifically, we will talk about how topic modeling can be applied to discover human routines. The basic idea is similiar to how topics are uncovered from text corpus using a bag-of-words approach [39].

4.1 Human Routine Discovery

Human routine discovery is about extracting temporal regularities in people's daily lives [22]. Daily routines typically have a hierarchical structure such that higher-level activities (longer duration of time) can be decomposed into a set of lower-level activities (shorter duration of time) [2]. More specifically, a routine can be seen as a composition of multiple low-level activities. Multiple low-level activities can occur within the same routine. Different routines may share and contain the same kinds of low-level activities, but with different proportions. For example, the "`Grocery Shopping`" routine may invovle more "standing" and "walking" activities compared to the "`Office Work`" routine.

Based on the hierarchical structure of human activities, we can make an analogy between text and sensor data modeling using topic models. In the context of mining a sequence of sensor data, sensor data features are first mapped into a set of discrete *labels* (feature vocabulary). Each mapped data feature is treated as if it were a text word. Then, the bag-of-features in each temporal window, and is used to train the topic model. Sensor data segments belong to the same routine if they have similar topic proportions. Figure 4.1 visualizes the analogy of sensor data and text modeling using topic models.

According to the intuitive analogy of sensor data to text while applying topic models, most existing approaches for automatic routine discovery from sensor data are built on parametric topic models such as LDA [5]. In the parameteric setting, the generative procedure requires two types of parameters to be predefined: the size of feature vocabulary and the number of latent topics. Typically, they are chosen in a trial-and-error fashion [35][22].

However, for routine discovery, such parameter specification poses several challenges. First of all, the best parameter values for personalized models may be different for different users. For example, due to the fact that different people usually have very distinct behavior patterns based on their lifestyles, job types, or ages, their routine patterns may require different number of latent topics to model appropriately. Moreover, even for a single user, it is possible that her behavior patterns change over time. The best parameter values must also be adjusted accordingly. Hence, we need the model to automatically select parameter values based on individual users' behavior patterns.

The limitations of parametric settings motivate the development of the nonparametric human routine discovery framework presented in this chapter. Figure 4.2 shows each component of the routine discovery framework. The framework consists of two phases: vocabulary

Figure 4.1: Comparison between topic model-based approach to human routine discovery and text modeling.

extraction and routine extraction. During the first phase (vocabulary extraction), we build up the vocabulary and automatically determine the size of feature vocabulary (low-level activity representation) from raw sensor data using the Dirichlet process Gaussian mixture model (DPGMM). In the second phase (routine extraction), we infer topic proportions (high-level routine representation) for each data segment with the automatically determined number of latent low-level activities using hierarchical Dirichlet process (HDP) and extract latent routines.

4.2 Dataset Description

To evaluate our nonparametric Bayesian framework for routine discovery, we experimented with two public datasets including realistic daily life routines and transportation modes [35][86].

4.2.1 Daily Life Routine Dataset

The daily life routine dataset was released by Technische Universitat Darmstadt (TU Darmstadt)[1]. It contains 34 daily low-level activity classes (including the unlabeled class). In addition to the low-level activity class annotations, this dataset also provides 4 high-level routine class annotations (i.e., commuting, lunch, office work, and dinner). Figure 4.3 shows the ground truth labels of low-level activity and high-level routines in one day of data. The sensor data were collected from two wearable 3-axis accelerometers worn on the right hip pocket and the dominant wrist. The accelerometer data were sampled at a rate of 100Hz, and the features (i.e., mean and variance of acceleration of each axis) were computed over a window of 0.4 seconds (i.e., 2.5Hz).

[1] http://www.ess.tu-darmstadt.de/datasets/tud-ubicomp08

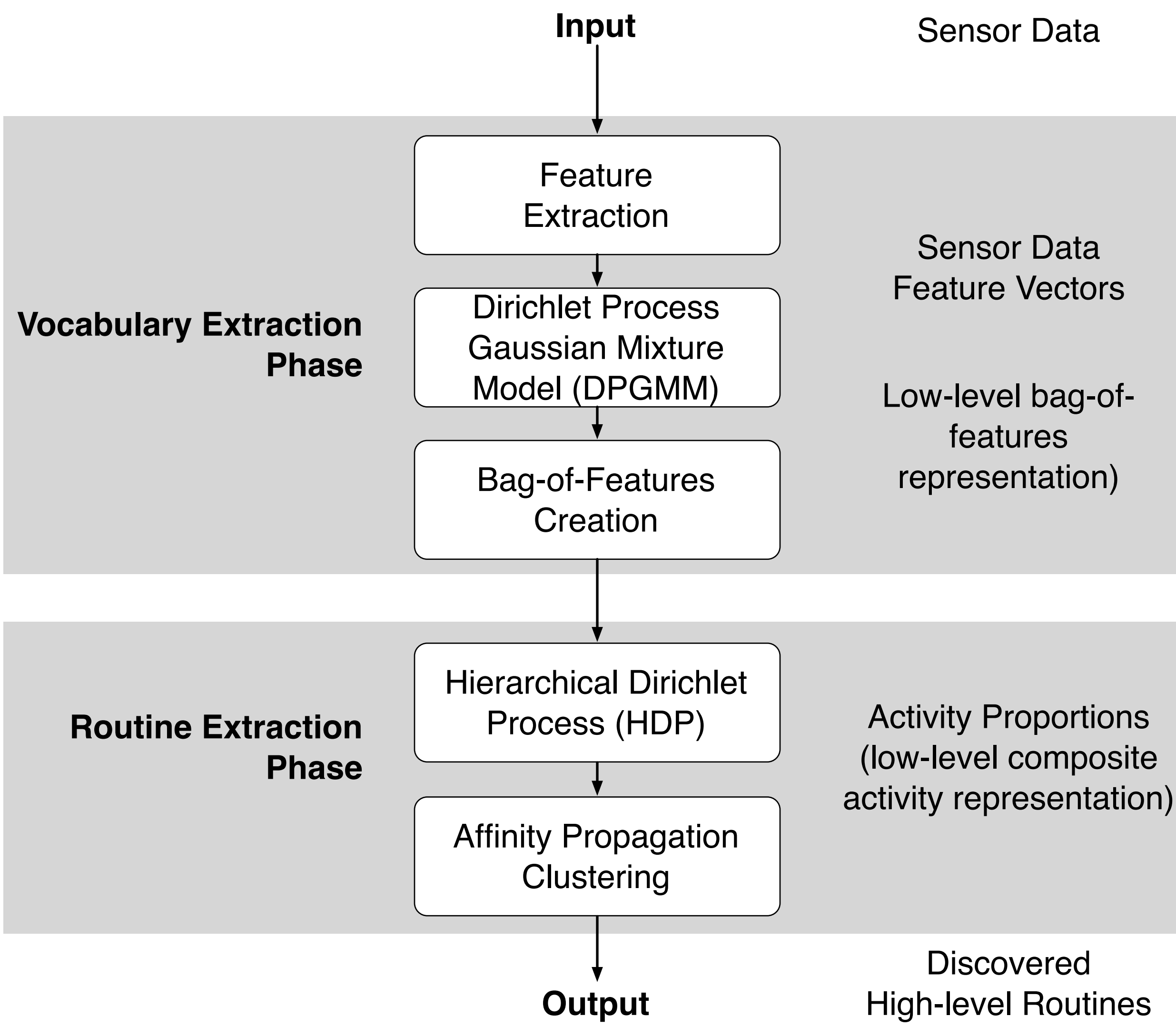


Figure 4.2: Overview of our two-phase nonparametric routine discovery framework.

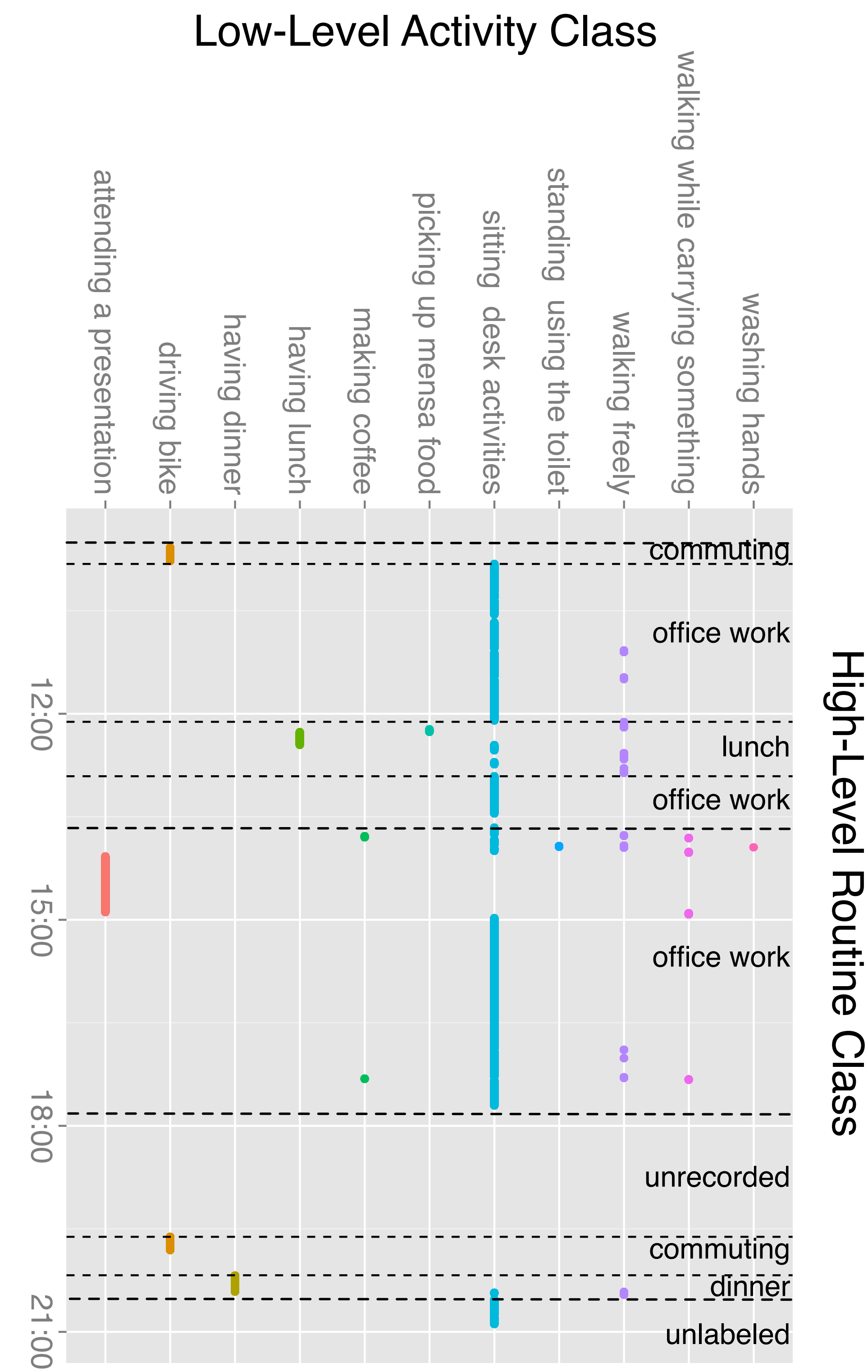


Figure 4.3: Illustration of the ground truth labels of daily activity dataset.

4.2.2 Transportation Mode Dataset

The transportation mode dataset was collected by Microsoft Research Asia (MSRA)[2]. This dataset contains GPS trajectory data from 182 users over a period of five years. A GPS trajectory is represented by a sequence of timestamped coordinates including longitude, latitude, and altitude. The GPS data were sampled approximately every two seconds (i.e., 0.5Hz). Each GPS trajectory was provided with a specific transportation mode annotation such as bike, walk, bus, and subway. In this experiment, we extracted a week of GPS trajectory data from one user for the routine discovery task.

4.3 Human Routine Discovery Framework

4.3.1 Feature Preprocessing

Our framework is agnostic to the type of input sensor data. Once the low-level features are extracted from raw sensor data, we input them to the activity vocabulary extraction module.

For the daily life routine dataset, we use 12 features consisting of *mean* and *standard deviation* of accelerometer data in dimension x, y, and z from wrist and pocket sensors. These features have been proven effective in previous work [35]. Figure 4.4 shows an example of feature distributions across four different daily routine classes. We see that the histograms of the same feature type behave quite differently across different routine classes.

The features for the transportation mode dataset include *velocity*, *heading direction change rate*, and *stop rate* derived from the GPS trajectory data. Previous work identified that this set of GPS features are robust to traffic conditions [86]. Figure 4.5 illustrates the distributions of one GPS feature (i.e., velocity) across four different transportation modes.

[2]http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/

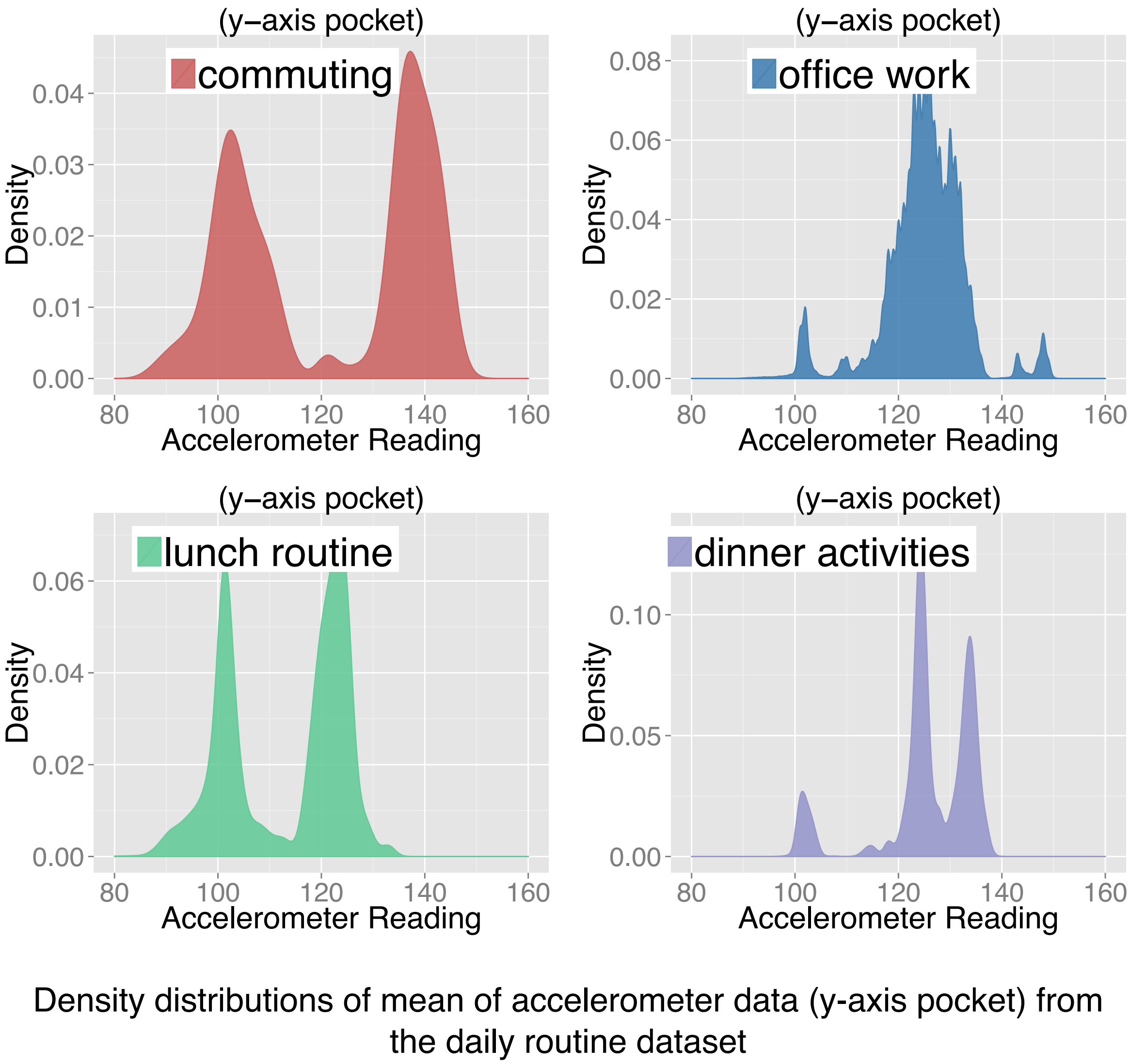


Figure 4.4: Examples of feature density distributions across different ground truth routines. Density distributions of the mean of the pocket accelerometer data in dimension y from the daily routine dataset.

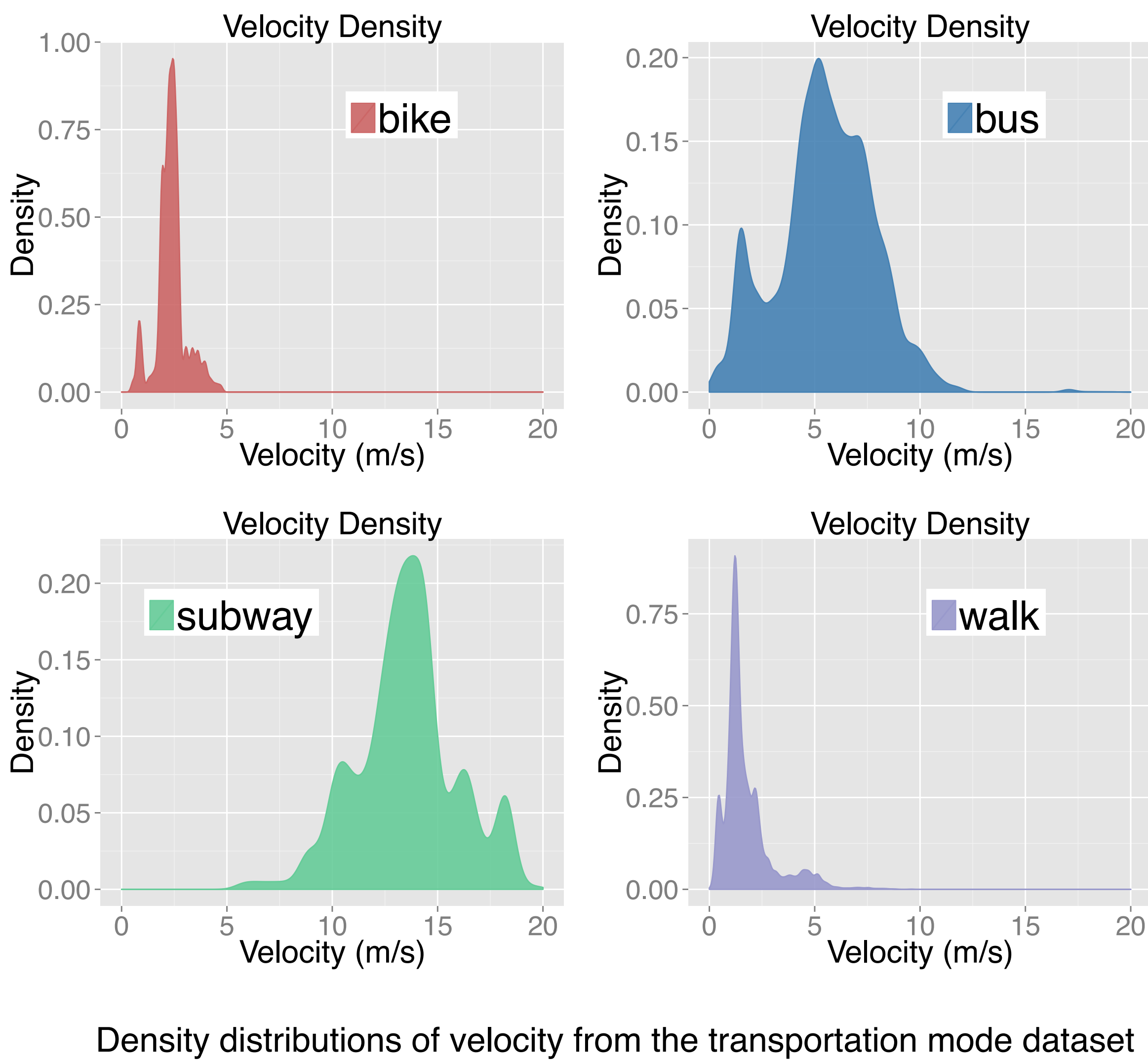


Figure 4.5: Examples of feature density distributions across different ground truth routines. Density distributions of the velocity feature from the GPS trajectory dataset. This suggests that feature density distributions of different routines can be modeled by Gaussian mixture models with different numbers of components.

4.3.2 Vocabulary Extraction

Figure 4.4 and 4.5 show that feature density distributions of different routines look like Gaussian distribution with multiple components. It suggests that feature distributions can be modeled by Gaussian mixture models with different numbers of components.

Hence, we describe how we use DPGMM to infer the set of discrete feature labels from the feature vectors in the context of the daily life routine dataset. Recall that in the daily life routine dataset, each data point is represented by a 12-dimensional feature vector ($\mu_{x-pocket}$, $\sigma_{x-pocket}$, ... , $\sigma_{z-wrist}$). Assume that there are N number of data points. One way to model these N data points with DPGMM is to use 12-dimensional Gaussian distributions as the component distributions. Each component is parameterized by a 12-dimensional mean vector and a 12×12 covariance matrix. However, a large number of parameters in the covariance matrix would create the problem of overfitting [41].

Therefore, we use the idea of dividing the feature space into lower-dimensional feature subspaces, fitting a different DPGMM for each subspace, and then combining the results together. More specifically, we organize the feature vectors into 6 subspaces by their sensor types and axis types. For example, one subspace corresponds to data points collected by the pocket sensor in the x-axis, ($\mu_{x-pocket}, \sigma_{x-pocket}$).

To simplify the notation, we use $\mathbf{t}_{ij}$ to denote the ith data point in subspace j where $i = 1, \ldots N$ and $j = 1, \ldots, 6$. The use of DPGMM to fit the data in subspace j can then be formulated as:

$$\begin{aligned}
G_\mu &\sim DP(\alpha, \mathcal{N}(\boldsymbol{\lambda}, \boldsymbol{\gamma}^{-1})) \\
G_S &\sim DP(\alpha, Gamma(\boldsymbol{\beta}, \boldsymbol{\omega}^{-1})) \\
\boldsymbol{\mu}_{c_{ij}} &\sim G_\mu \\
\boldsymbol{S}_{c_{ij}} &\sim G_S \\
\boldsymbol{t}_{ij}|\theta_{c_{ij}} &\sim \mathcal{N}(\boldsymbol{\mu}_{c_{ij}}, \boldsymbol{S}_{c_{ij}})
\end{aligned} \tag{4.1}$$

where c_{ij} is an indicator variable specifying the cluster associated with $\boldsymbol{t}_{ij}$, and $\{\boldsymbol{\mu}_{c_i}, \boldsymbol{S}_{c_i}\}$ is the set of parameters of the Gaussian component for cluster c_{ij}. $\boldsymbol{\mu}_{c_{ij}}$ and $\boldsymbol{S}_{c_{ij}}$ are generated by Dirichelet processes with base distributions Gaussian and Gammas, respectively. Four hyperparameters $\boldsymbol{\lambda}$, $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ specify the base distributions, expressing the strength of the prior belief in the distribution of the parameter space.

During inference, we run Gibbs sampling for 500 iterations with a burn-in period of 100 iterations to infer c_{ij} for each $\boldsymbol{t}_{ij}$ [60]. Finally, for each data point, we concatenate the corresponding cluster assignments from the 6 subspaces to form a discrete feature label $w_i = (c_{i1}, \ldots, c_{i6})$. Note that we do not need to specify the number of unique artificial words (vocabulary size) beforehand.

Similarly, for the transportation mode dataset, we applied DPGMM for each of the three feature vectors (i.e., *velocity*, *heading direction change rate*, and *stop rate*) to construct feature labels representing the data points.

4.3.3 Routine Extraction

Based on the artificial word representation for the sensor data, we now describe how to construct bag-of-features and extract routines using HDP.

To construct bag-of-features from a stream of feature labels, we represent each temporal window as a histogram of feature label occurrences in a sliding window with overlapping. For the daily routine dataset, the sliding windows are set to 30 minute duration with 2.5 minute overlapping, similar to previous work by Huynh et al. [35]. For the transportation mode dataset, the sliding window is 10 minutes duration with 1 minute overlap.

Using a Gibbs sampling scheme similar to the inference stage of DPGMM, we obtain the mixture proportion of latent topics for each data segment.

Finally, we cluster low-level activity proportions using the affinity propagation (AP) algorithm since it is also a nonparametric clustering algorithm and successfully applied in many applications [20, 24, 61]. Given the data points $\{x_i, ..., x_N\}$, the AP clustering algorithm takes the similarity of pairs of data points $\{s_{ij}\}_{i,j\in\{1,...,N\}}$ as input and forms cluster assignments $\{c_1, ..., c_N\}, c_i = \underset{i}{\operatorname{argmax}}[a(i,k) + r(i,k)]$ by finding data points that are examplars of clusters. $r(i,k)$ denotes an element of the responsibility matrix $\boldsymbol{R}$, and it quantifies how well x_k can be as the exemplar for x_i. $a(i,k)$ represents an element of the matrix $\boldsymbol{A}$, and how it is appropriate if x_i picks x_k as its exemplar. Then, the AP algorithm updates matrices R and A by executing two message passing steps iteratively: Mathematically, two update steps can be defined as follows [19]:

$$
\begin{aligned}
\forall i, k \in \{1, ..., N\}: \quad r(i,k) =& s(i,k) - \max_{k' \neq k}[s(i,k') + a(i,k')] \\
\forall i, k \in \{1, ..., N\}: \quad a(i,k) =& \{ \begin{array}{ll} \sum_{i' \neq i} \max[0, r(i,k')] & (k = i) \\ \min[0, r(k,k) + \sum_{i' \notin \{i,k\}} \max[0, r(i',k)]] & (k \neq i) \end{array}
\end{aligned}
\tag{4.2}
$$

We use Jensen-Shannon divergence as the distance function. Let π_i and π_j denote the topic proportions of document i and j. Their Jensen-Shannon divergence is defined by

$$
\begin{aligned}
JSD(\pi_i, \pi_j) =& \frac{1}{2} D_{KL}(\pi_i || M) + \frac{1}{2} D_{KL}(\pi_j || M) \\
D_{KL}(P||Q) =& \sum_i ln(\frac{P(i)}{Q(i)}) P(i)
\end{aligned}
\tag{4.3}
$$

where $M = \frac{1}{2}(\pi_i + \pi_j)$ and $D_{KL}(P||Q)$ is the Kullback-Leibler divergence of two distributions.

Then, the similarity between two topic proportions are computed by

$$
D(\pi_i, \pi_j) = e^{-JSD(\pi_i, \pi_j)} \tag{4.4}
$$

Each cluster corresponds to a discovered routine. Intuitively, two sliding windows (data segments) are assigned to the same gorund truth routine label if their low-level activity proportions are similar. Using the AP algorithm to perform routine clustering, the number of routines needs not be pre-specified.

4.4 Evaluation

In this section, we present qualitative and quantitative evaluation results of our nonparametric routine discovery framework. First, we show the discovered routines from both the daily life routine dataset and the transportation mode dataset. Second, we compare DPGMM and K-means, the baseline clustering method used in previous work [35], for vocabulary construction. Finally, the performance comparison of nonparametric (HDP) and parametric (LDA) topic models is presented.

4.4.1 Qualitative Analysis

We first show the output of our routine discovery method on the two datasets. Figure 4.6 illustrates the extracted routines for one day from the daily life routine dataset. The top chart shows how the learned topic proportion vectors change over time. The documents are constructed using a 30-minute sliding window with 2.5 minutes overlap. (Data between 14:00-16:30 are unlabeled.)

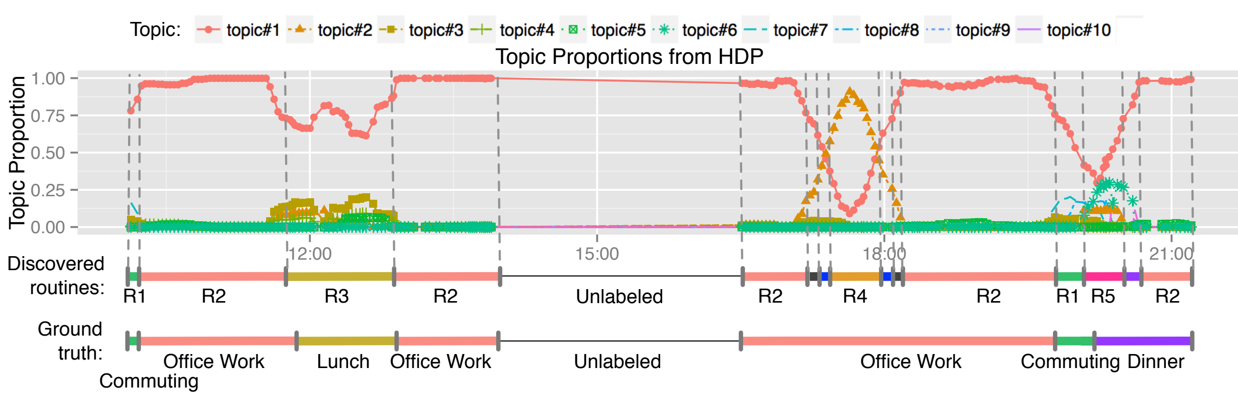


Figure 4.6: The chart (top) visualizes topic proportions inferred from HDP during the course of a day in the daily routine data. The chart (bottom) shows the comparison of the ground truth routine labels and the discovered routines from our framework. Note that the inferred topics reveal high correlation with annotated routine labels. For example, "`Office Work`" and "`Lunch`" correspond to higher proportion of Topic#1 and Topic#3 respectively. Hence, topic proportions allow us represent and discover high-level daily routines.

The middle chart shows the extracted routine classes by clustering these topic proportion vectors using affinity propagation clustering algorithm. The bottom chart shows the ground truth. The ground truth label of a specific sliding window is assigned with the most frequent routine class label in that window. Different colors indicate different routines. We see that the discovered routines “R1” (green), “R2” (pink), and “R3” (yellow) match ground truth labels “`Commuting`”, “`Office Work`”, and “`Lunch`”.

Our method extracted more routines than the ground truth labels. For example, 17:00-18:00 is labeled as “R4” while the ground truth label is still “`Office Work`”. We examined the ground truth activity labels in the dataset and found that “walking freely” occurs more frequently during this period of time compared to other parts of the “`Office Work`”. Thus, it is expected that our method would identify it as an additional routine.

Moreover, our method labels “R2” in the last part of the “`Dinner`” period. This is because the labeled “`Dinner`” period consists of “*cooking in the kitchen*” and “*sitting at the table to dine*”. From the accelerometer sensors’ point of view, “*sitting at the table to dine*” and “*sitting at the office table*” have similar feature distributions. Therefore, based on the sensor data, our method is not able to distinguish them. This scenario suggests that using other types of sensor data, such as location and time, might be useful.

Similarly, Figure 4.7 shows the inferred transportaion modes on a week of transportation mode data. The documents are constructed using a 10 minute sliding window with 1 minute overlap. We reorder and group documents with the same ground truth labels for better visualization. We see that the discovered routines are highly correlated to the ground truth transportation modes.

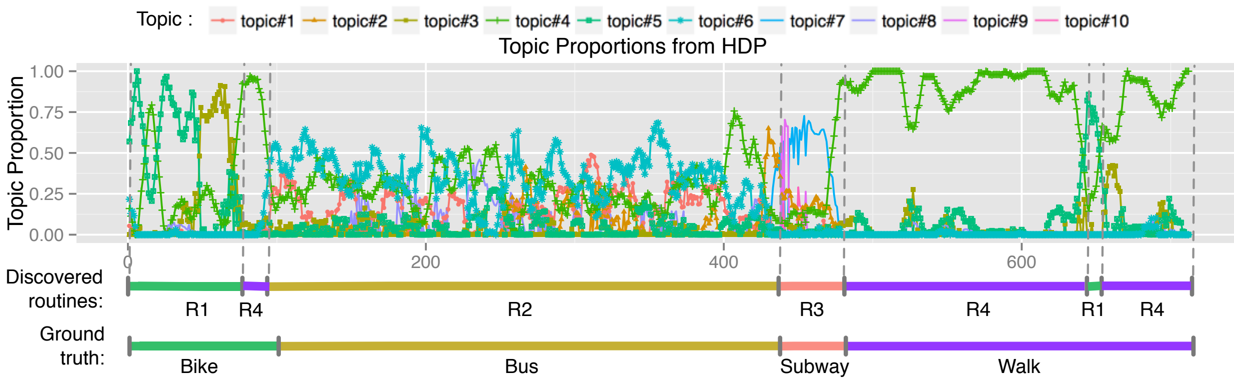


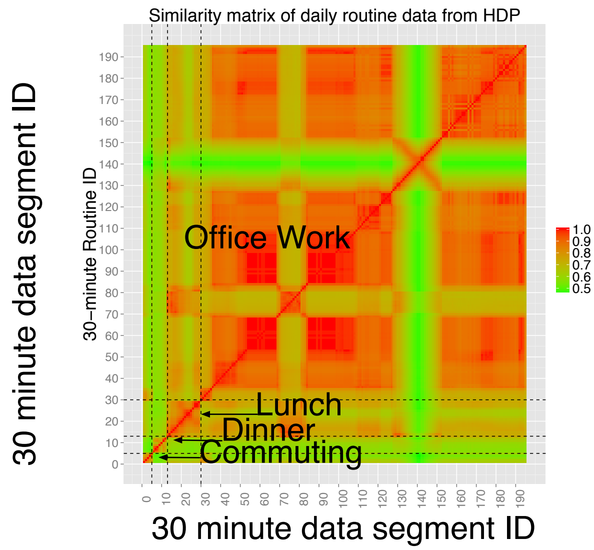
Figure 4.7: The chart (top) illustrates topic proportions inferred from HDP for a week of the transportation mode data. The chart (bottom) shows the comparison of the ground truth transportation mode labels and the discovered routines from our framework. Note that the inferred topics reveal high correlation with annotated transportation mode labels. For example, "`Bike`" and "`Walk`" correspond to higher proportion of Topic#5 and Topic#4 respectively. Hence, it suggests that we can use topic proportions to represent and discover transportation mode routines.

Figure 4.8 shows two similarity matrices of the extracted topic proportion vectors on the daily routine and the transportation mode datasets using Eq (4.4). The red color refers to higher similarity, and the green color refers to lower similarity. For better visualization, we group documents based on their ground truth routine labels. In both similarity matrices, we see that topic proportion vectors of documents with the same ground truth label are similar, forming red sub-blocks. Moreover, the green bands in the "`Office Work`" block in Figure 4.8(a) correspond to "`Office Work`" with more "walking freely" activity occurrences.

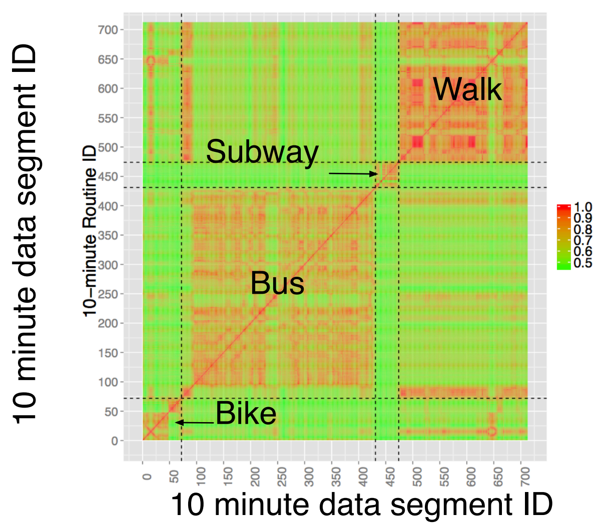
4.4.2 Quantitative Evaluation

Evaluation metrics

To measure the alignment between the discovered routines and the ground truth routines, we compute three widely used clustering evaluation metrics: the *cluster purity*, the *Rand index* and the *pair-counting F-measure* [52]. Given clustered documents, *cluster purity* assigns the cluster to the ground truth routine label which is most frequent in the cluster, and



(a) Daily routine data similarity matrix



(b) Transportation data similarity matrix

Figure 4.8: Similarity matrix for (a) the daily routine dataset and (b) the transportation mode dataset of the low-level activity proportion vectors learned from HDP. Red color refers to higher similarity and green color refers to lower similarity. Note that learned activity proportions of data segments with the same ground truth labels have higher similarity.

then computes the percentage of documents whose ground-truth label is the same as its cluster label. The *Rand index* looks at all pairs of documents and caculates the percentage of document pairs that are correctly classified, $\frac{TP+TN}{TP+FP+FN+TN}$, where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives respectively. More specifically, TP is the number of similar document pairs that are assigned to the same cluster and TN is the number of dissimilar document pairs that are assigned to different clusters. *F-measure* is defined as $F_\beta = \frac{(\beta^2+1)P \times R}{\beta^2 P+R}$ where P is precision and R is recall. In this experiment, we set $\beta = 1$ and *F-measure* becomes $\frac{2 \cdot TP}{2 \cdot TP+FP+FN}$ in which true positive are double counted. Intuitively, *F-measure* penalizes false negatives more than false positives.

DPGMM vs. K-means

We first study the effect of using nonparametric and parametric methods in the vocabulary construction phase. We consider K-means as the baseline parametric method for constructing vocabulary. Given a set of feature vectors $\{\boldsymbol{f}_1, ..., \boldsymbol{f}_n\}$, K-means clustering aims to separate the n feature vectors into K clusters, where K is less than n. The process of K-means is to minimize the within-cluster sum of squares. Mathematically, the process of minimizing the objective function (i.e., squared error function) can be written as follows:

$$\underset{\boldsymbol{C}}{\operatorname{argmin}} \sum_{i=1}^{K} \sum_{\boldsymbol{f}_j \in C_i} \|\boldsymbol{f}_j - \boldsymbol{\mu}_i\|^2 \tag{4.5}$$

where $\boldsymbol{\mu_i}$ is the mean of feature vectors belong to cluster C_i and $\|\boldsymbol{f}_j - \boldsymbol{\mu}_i\|^2$ is a chosen distance measure between a feature point $\boldsymbol{f}_j$ and the cluster center $\boldsymbol{\mu_i}$.

Figure 4.9 shows results on the daily routine dataset using K-means and DPGMM. The high level topic model is LDA. For the K-means baseline, we use $K = 60$ because it was reported as the optimal parameter value in previous routine discovery work on the same dataset [35]. We see that DPGMM performs consistently better than K-means in all three evaluation metrics as we change the number of topics. This demonstrates the advantage of nonparametric methods.

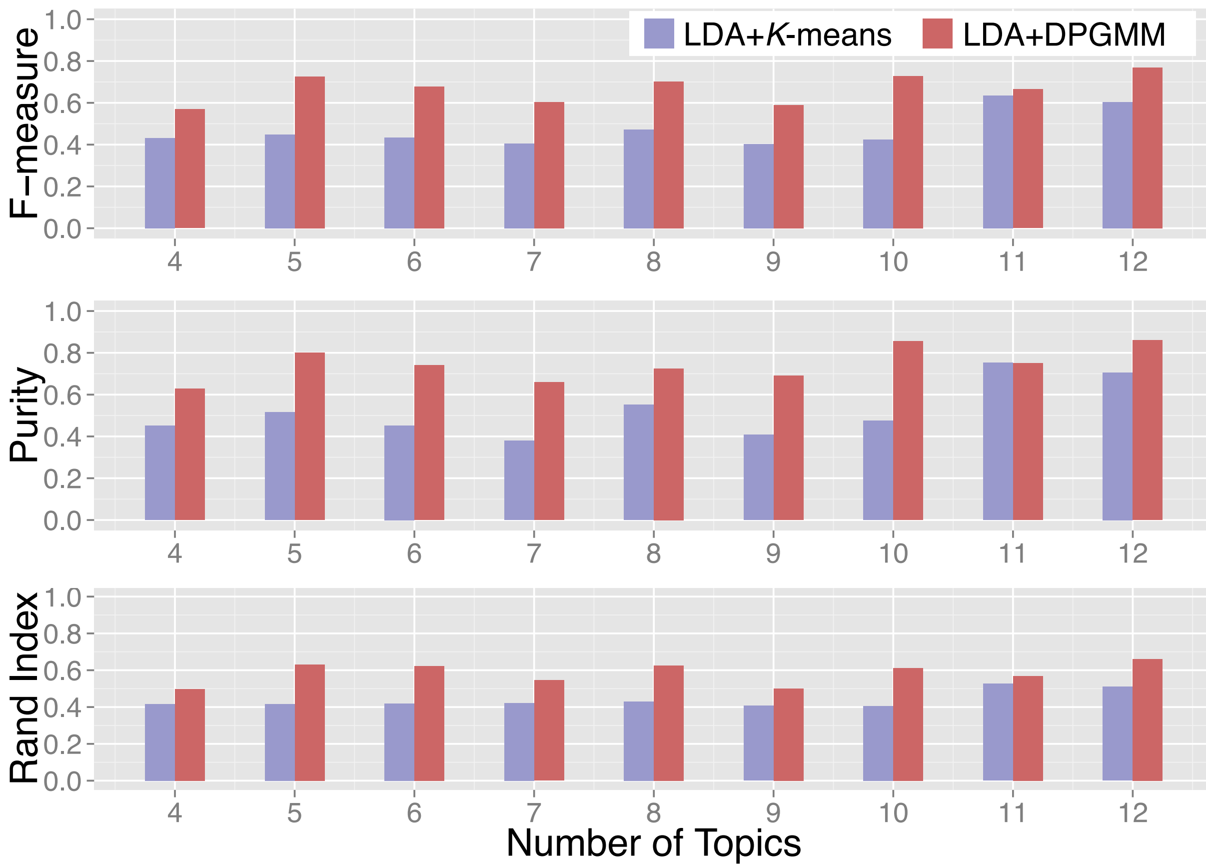


Figure 4.9: Performance comparison of DPGMM (nonparametric) against K-means (parametric) for vocabulary extraction on the daily routine dataset, in terms of the F-measure (top), the purity (middle), and Rand index (bottom) over different numbers of topics in LDA.

HDP vs. LDA

Next, we fix the vocabulary construction scheme (DPGMM) and compare the performance of HDP and LDA with various number of topics shown in Figure 4.10. The number of topics used in LDA varies from 4 to 12. Since the number of topics is not part of the problem formulation in HDP, they are horizontal lines in the charts. As we would expect, LDA is sensitive to the selection of the number of topics. Also, HDP and the optimal setting in LDA have comparable performance. Note that the number of topics does not correspond to the number of routines. Therefore, even if we know that we have 4 routines in advance, the optimal number of topics is not 4. The number of topics inferred in HDP on the daily routine and the transportation datasets is 12 and 10, respectively.

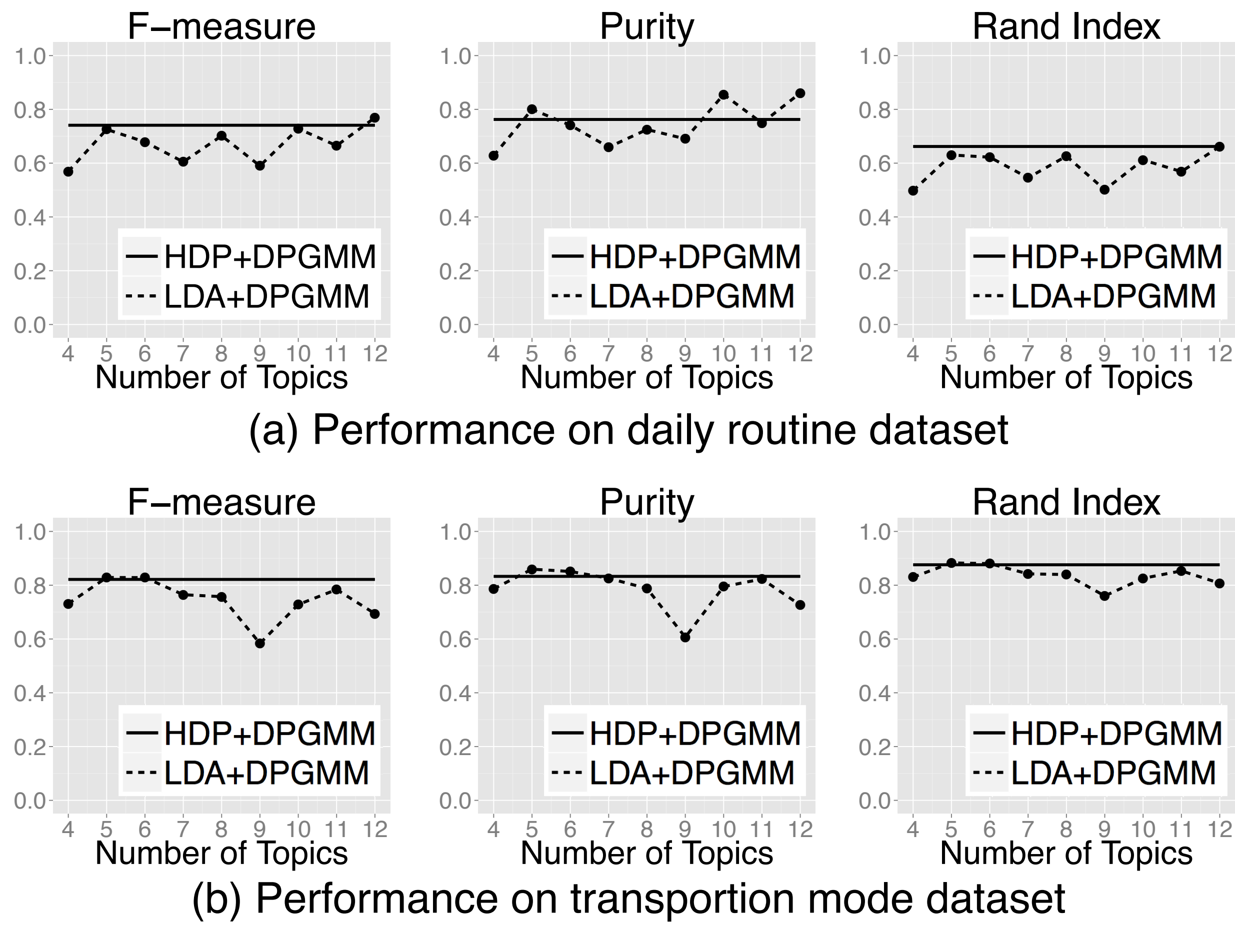


Figure 4.10: Performance comparison of HDP+DPGMM (nonparametric) against LDA+DPGMM (parametric) on the daily routine and the transportation mode dataset in terms of the F-measure (left), the purity (middle), and Rand index (right). We can see that HDP+DPGMM performs as well as the best LDA model without a model selection procedure.

4.5 Discussion

In this chapter, we have presented a novel routine discovery framework which adopts non-parametric topic modeling techniques. Basically, this idea is based on the intuition that human routines can be recognized in a fashion similar to how topics are uncovered from text corpus. Most previous work in topic model based routine discovery used parametric methods such as K-means for low-level feature clustering and LDA for routine discovery. However, model selection (e.g., via a trial-and-error process) is the main challenge for the adoption of parametric models. Our two-phase nonparametric routine discovery framework uncovers latent routines from sensor data in a fully unsupervised fashion. More specifically, the framework automatically finds the size of the low-level activity vocabulary from multi-dimensional feature vectors using DPGMM at the vocabulary extraction phase. At the routine discovery phase, the framework further applies HDP to automatically select the appropriate number of latent topics and discover latent routines. The framework has been validated on two public datasets. Experimental results show that our nonparametric framework can achieve better performance against parametric models without the need of specifying parameters in advance.

Chapter 5

Towards Multimodal Discovery of Human Behavior

In the previous chapter, we presented the core idea, the implementation, and the evaluations of the nonparametric human routine discovery framework. Experimental results show that the proposed nonparametric framework can achieve better performance against parametric models without the need to specify parameters in advance. However, so far we have only incorporated a single sensor modality (e.g., accelerometer or GPS) with the proposed nonparametric framework. We have not incorporated several different modalities of sensor data to increase the performance of the human routine discovery framework.

Most of the related work using topic models only makes use of a single modality of sensor data or applies simple concatenation from different sensor data modalities [22]. A learning model with multimodal data should be the core of many multisensory applications. Previous work has shown that multimodal data is useful for modeling human activities [27, 29]. Hence, it motivates us to extend our nonparametric Bayesian model to take into account multimodal information.

The concept of multimodal topic modeling has shown promising results for tasks such as object categorization and image label prediction in the field of robotics and image recognition [59, 82]. In this chapter, we will present the design, implementation, and evaluation of the multimodal hierarchical Dirichlet process, an extension of hierarchical Dirichlet process (single modality), to model the relationships among multimodal data sources for recognizing human behavior patterns.

5.1 Dataset Description

To evaluate the proposed multimodal nonparametric Bayesian framework for daily activity discovery, we experimented with the realistic public dataset [27]. This dataset is different than the daily life activity dataset containing only a single sensor modality (i.e., accelerometer) used in Chapter 4. The dataset consists of 42 days (11,721 minutes) of multisensory data with four types of sensor features (i.e., accelerometer, GPS, image, and audio) and three types of annotations that are activity, location, and people as shown in Table 5.1. The user carried the smartphone by a neck strap during the daytime. As for the annotation process, a user reviews his/her visual log of data recordings, segments each day's data into a few events, and labels these events with tags [27]. Figure 5.1 illustrates three types of ground truth annotations of one day of sensor data. At any point in time, there are three types of ground truth annotations that are activity, location, and people. The following sections describe the four types of sensory features and how we extracted the feature vocabulary and routines from the sensor data.

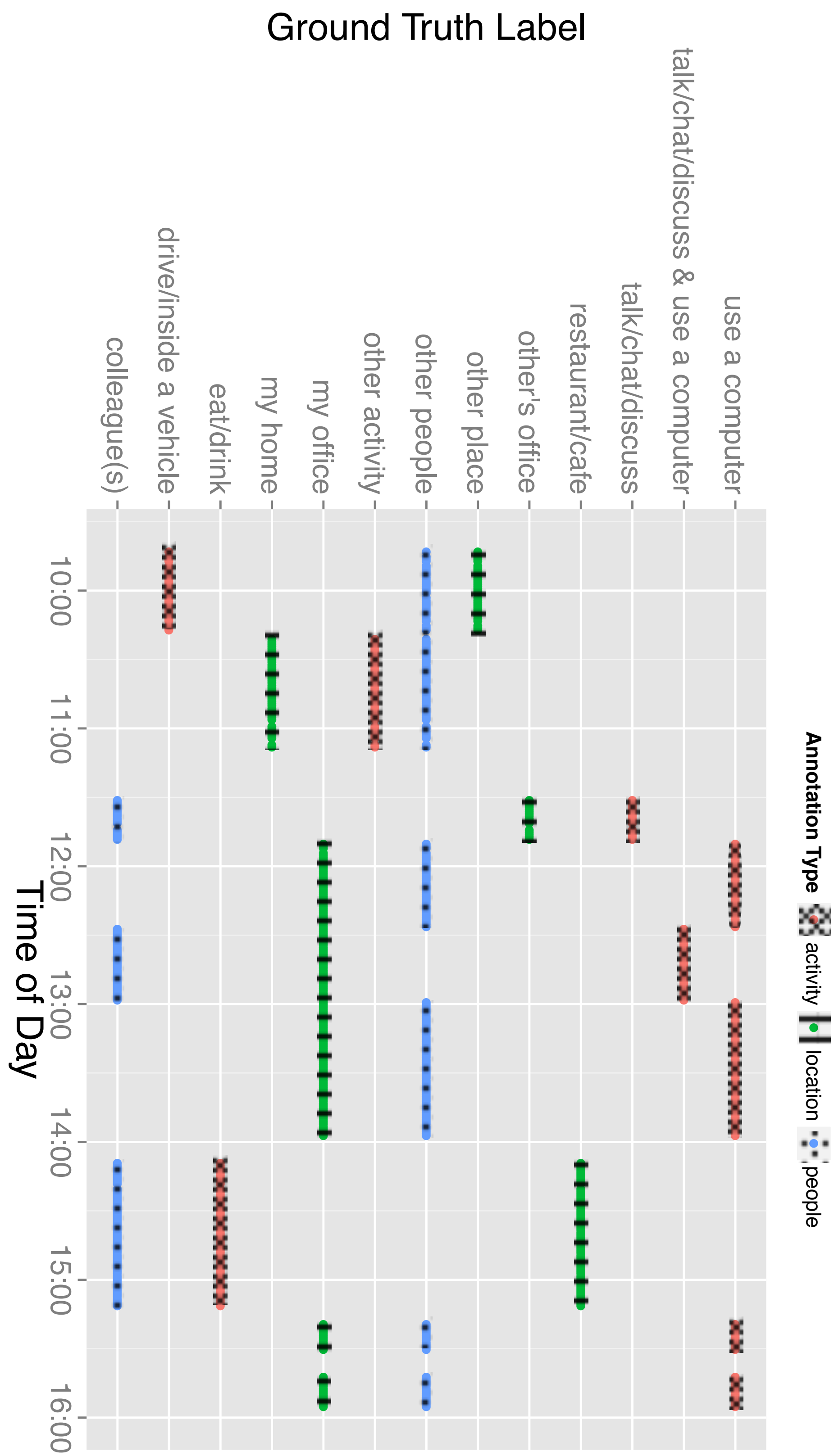


Figure 5.1: Illustration of the ground truth labels of multimodal daily activity dataset.

Annotation type	Annotation classes
Activity	walk, drive/inside a vechicle, eat/drink, talk/chat/discuss, chores, tend to baby, use a computer, read/write on paper/board, other activity
Location	outdoor, restaurant/cafe, other's office, classroom/meeting room, my office, my home, other place
People	family, colleague(s), other people

Table 5.1: List of annotation types and classes used in the daily activity dataset.

5.2 Feature Preprocessing

Accelerometer

The tri-axial accelerometer data was collected at a sampling rate of 16 Hz. Time-domain (i.e., mean, variance, zero-crossing rate, etc.) and frequency-domain features (i.e., Fast fourier transform [FFT] coefficients, spectral entropy, etc.) were extracted to form a 27-dimensional feature vector over a window of 1 second. These features have been proven effective in previous activity recognition work [1].

Audio

Raw audio data was sampled at a rate of $11,025$ Hz in 16-bit pulse-code modulation (PCM) format. The Mel-frequency cepstral coefficients (MFCC) have enough information to reconstruct the partial speech content [49]. In order to further protect privacy, a sparse and short-duration (i.e., 250 ms shorter than the typical duration of a word) audio segment was sampled every 5 seconds. The final 126-dimensional feature vector was composed of MFCC coefficients over non-overlapping windows for each audio segment.

Image

A 24-bit color JPEG image of size 480×640 was captured every minute. To obtain features from raw image data, a subset of MPEG-7 feature descriptors were used [6]. The extracted 64-bin HSV-space color histogram and 80-bin edge histogram were concatenated as a 144-dimensional feature vector.

GPS

A GPS coordinate reading (i.e., latitude, longitude, and altitude) was acquired every minute. However, the coordinate readings were often unavailable due to the lack of signal inside the buildings, and the missing data was handled by computing the mean value within a one-minute window.

5.3 Vocabulary Extraction

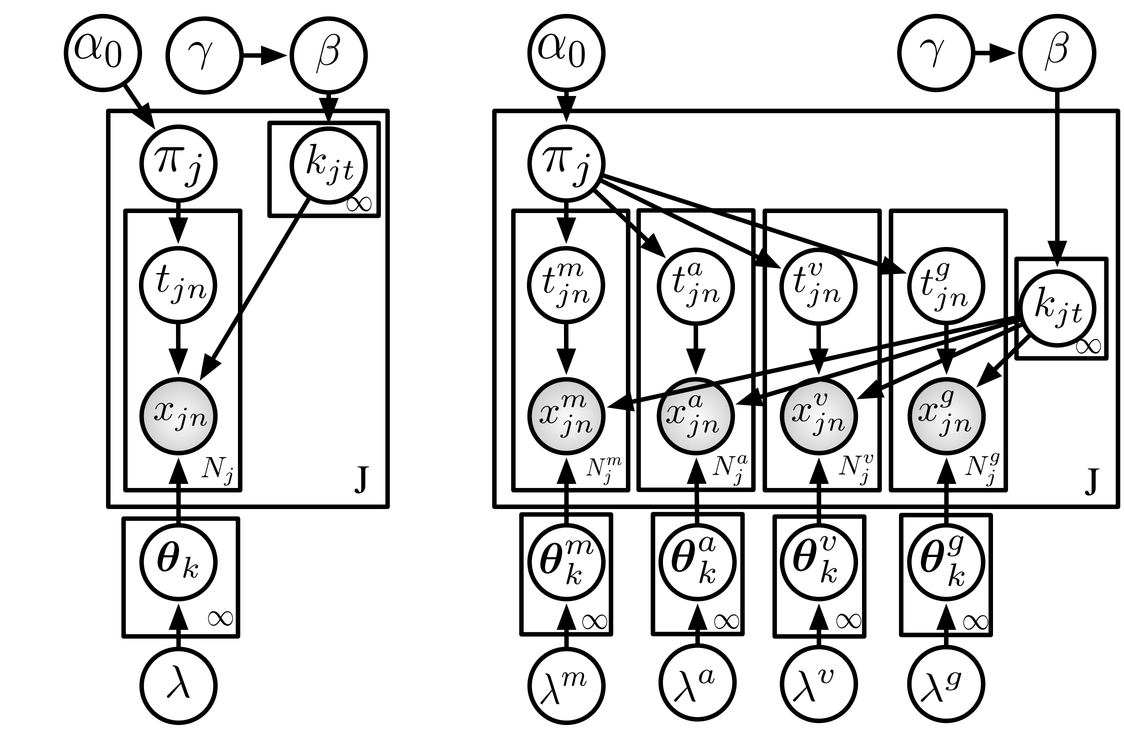
After the feature preprocessing stage, we now have feature vectors for each sensor modality. In order to train topic models for high-level activity discovery, we need to create a set of discrete labels (vocabulary) from the obtained multimodal feature vectors and compose the bag-of-features representation for each data segment. The feature vectors from each modality are mapped into a set of discrete labels using K-means clustering [38]. As mentioned in Chapter 4, feature vectors $\{\boldsymbol{f}_1, ..., \boldsymbol{f}_n\}$, K-means clustering aims to separate the n feature vectors into K clusters, where K is less than n. The process of K-means is to minimize the within-cluster sum of squares. Mathematically, the process of minimizing the objective function (i.e., squared error function) can be written as follows:

$$\underset{\boldsymbol{C}}{\operatorname{argmin}} \sum_{i=1}^{K} \sum_{\boldsymbol{f}_j \in C_i} \|\boldsymbol{f}_j - \boldsymbol{\mu}_i\|^2 \tag{5.1}$$

where $\boldsymbol{\mu_i}$ is the mean of feature vectors belong to cluster C_i and $\|\boldsymbol{f}_j - \boldsymbol{\mu}_i\|^2$ is a chosen distance measure between a feature point $\boldsymbol{f}_j$ and the cluster center $\boldsymbol{\mu_i}$.

In this dataset, K is assigned to be 50 during the pre-processing stage for feature vectors from each sensor modality. Then, we represented each data segment as a histogram of discrete label occurrences (bag-of-features) in a sliding window with overlapping. For the daily routine dataset, the sliding windows were set from 1 to 30 minute durations with 25% overlap.

5.4 Routine Extraction



(a) Hierarchical Dirichlet Process (b) Multimodal HDP

Figure 5.2: Graphical models of (a) the hierachical Dirichlet process (HDP) and (b) the multimodal hierachical Dirichlet process (MHDP).

In this section, we first briefly review the fundamental knowledge of nonparametric

Symbol	Description
α_0	concentration parameter
γ	concentration parameter
β	Dirichlet prior parameter
k_{jt}	index assignment of θ_k associated with random variable x_{jn}
λ	Dirichlet prior parameter for all θ_k
$\boldsymbol{\theta}_k$	parameters of multinomial distributions that generate the value of random variable x_{jn}
π_j	parameters of multinomial distributions that generate index assignment of t_{jn}
t_{jn}	index assignment associated with x_{jn} j
x_{jn}	n-th random variable in the grouped data j

Table 5.2: List of notations used in the graphical models.

Bayesian modeling such as Dirichlet process (DP) and the standard hierarchical Dirichlet process (HDP). The details of nonparametric Bayesian modeling have been reviewed in Chapter 3. Second, we focus on the details of multimodal hierarchical Dirichlet process (MHDP) including the graphical model representation and the parameter estimation process. Finally, we describe how MHDP is applied to obtain the output information in the context of high-level activity discovery.

5.4.1 Dirichlet Process

The basic building block of HDP is the Dirichlet process, which is an infinite-dimensional generalization of the Dirichlet distribution as described in Chapter 3 [75].

The DP is denoted as $DP(\alpha_0, G_0)$, where α_0 is the concentration parameter and G_0 is the base distribution. Like the Dirichlet distribution, the DP is also a distribution over distributions. A distribution $G(\theta)$ is DP distributed if $G \sim DP(\alpha_0, G_0)$.

The DP can be constructed by the Chinese Restaurant Process (CRP) [75]. In a CRP metaphor, the DP process can be described in the following way: (1) A restaurant (document) has an unlimited number of tables θ_k, $k = 1, ...\infty$. (2) The first customer (word) comes in and sits at the first empty table (topic). (3) Given K tables are occupied before the i-th customer comes in, the customer can either sit at table k ($k \leq K$) with probability $\propto \frac{n_k}{\alpha_0+i-1}$ or at a new table $K+1$ with probability $\propto \frac{\alpha_0}{\alpha_0+i-1}$, where n_k is the number of customers already sitting at table k.

For a set of random variables $\phi_1, \cdots, \phi_n$ (the table assignment for each customer), the distribution of the random variable ϕ_i conditioned on the previous random variables can can be written as

$$P(\phi_i|\phi_1, ..., \phi_{i-1}, \alpha_0, G_0) \sim \sum_{k=1}^{K} \frac{n_k}{\alpha_0 + i - 1}\delta_{\theta_k} + \frac{\alpha_0}{\alpha_0 + i - 1}\delta_{\theta_{\overline{k}}} \tag{5.2}$$

where $\overline{k}$ represents a new table $K+1$.

5.4.2 Standard Hierarchical Dirichlet Process (HDP)

The HDP model consists of two levels of DP. The base distribution of the lower level DP is drawn from the upper level DP. Figure 5.2(a) shows the graphical model of HDP for the document classification, and the notations are listed in Table 5.2. Similar to DP's CRP metaphor, the generating process of HDP can be described as a Chinese Restaurant Franchise (CRF). In the CRF metaphor, all restaurants (documents) share the same dish menu. The j-th restaurant has T_j tables, and the n-th customer x_{jn} sits at table t with dish k_{jt} on it. The types of dishes represent the classes of data. CRF is essentially a two-level CRP: (1) Within a restaurant, customers choose tables. (2) Within all restaurants, tables choose dishes from the dish menu.

5.4.3 Multimodal Hierarchical Dirichlet Process (MHDP)

We now develop the extension of HDP to model multimodal data. In this subsection, we focus on the daily activity dataset with four modalities of sensor data. However, MHDP can

Human activity discovery	Chinese restaurant franchise
low-level sensor feature data points	customers
data segments (temporal windows)	restaurants
types of sensor modalities (e.g., motion or audio)	types of customers (e.g., genders or ages)
clusters of sensor features within a data segment	tables in a restaurant
types of latent low-level activities in the dataset	dishes on the menu

Table 5.3: Comparison of terminology in the human activity discovery domain and in the Chinese restaurant franchise metaphor.

be applied to different and more modalities of sensor data without loss of generality.

Graphical Model Representation

Figure 5.2(b) shows the graphical model of MHDP. In the graphical model, $x^m_{jn}, x^a_{jn}, x^v_{jn}$, and x^g_{jn} denote the n-th *motion*, *audio*, *visual*, and *GPS* sensor feature data point in the j-th temporal window. The sensor feature data point from each sensor modality $s \in \{m, a, v, g\}$ is drawn from multinomial distributions parameterized by $\boldsymbol{\theta}^s$. Note that each multinomial distribution θ^s_k is drawn from a Dirichlet distribution parameterized by α^s_0 ($s \in \{m, a, v, g\}$).

Simliar to HDP, the data generation process of the MHDP can be also realized by the Chinese Restaurant Franchise process (two-level CRP). In the following context, we describe the parameter estimation process of MHDP with terms used in the human activity discovery domain rather than in the CRF metaphor.

Table 5.3 lists the detailed terminology mapping in the human activity discovery domain and in the CRF metaphor. All restaurants are data segments (temporal windows) and each sensor feature data point is a customer. Different modalities of sensor data feature can be considered as different types of customers (e.g, different genders or ages) and types of dishes represent different low-level activity classes of data. Tables in a restaurant can be seen as clusters of sensor feature data points within a data segment.

Parameter Estimation of MHDP via Gibbs Sampling

In order to extract the low-level activity proportion of each data segment (temporal window), we need to first estimate the cluster and activity class assignments using Gibbs sampling. Algorithm 1 demonstrates the detailed Gibbs sampling steps for two level DP based on the CRF process: data segment-level cluster selection and overall dataset-level activity selection processes.

Algorithm 1 Parameter estimation process of MHDP.

---

repeat

  for all $j\in\{1,\cdots,J\}$, $s\in\{m,a,v,g\}$, $n\in\{1,\cdots,N_j^s\}$ **do**

    (a) sensor feature data point x_{jn}^s is first removed from cluster $t=t_{jn}^s$:

      N_k^s--, N_{jt}--, $N_{kx_{jn}^s}^s$--

    (b) cluster assignment t is sampled:

$$t_{jn}^s \sim \begin{cases} P(x_{jn}^s|\boldsymbol{X}_{k=k_{jt}}^s)\cdot\frac{n_{jt}}{\alpha_0+N_j-1} & (t\leq T_j) \\ P(x_{jn}^s|\boldsymbol{X}_{k=k_{jt}}^s)\cdot\frac{\alpha_0}{\alpha_0+N_j-1} & (t=T_j+1) \end{cases}$$

    (c) If a new cluster is generated ($t=T_j+1$) in step (b), a corresponding activity class assignment k is sampled:

$$k_{jt} \sim \begin{cases} P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\cdot\frac{M_k}{\gamma+M-1} & (k\leq K) \\ P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\cdot\frac{\gamma}{\gamma+M-1} & (k=K+1) \end{cases}$$

    (d) sensor feature data point x_{jn}^s with a new cluster assignment $t=t_{jn}^s$ is assigned to activity class $k=k_{jt}$:

      N_k^s++, N_{jt}++, $N_{kx_{jn}^s}^s$++

    (e) delete empty clusters that no sensor feature data point belongs to.

  end for

  for all j, t **do**

    (a) activity class $k=k_{jt}$ is first removed from cluster t in the data segment j:

      $N_{kx^s}^s$--, N_k^s-- for all $x^s\in\boldsymbol{X}_t$, M_k--

    (b) a new activity class is sampled:

$$k_{jt} \sim \begin{cases} P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\cdot\frac{M_k}{\gamma+M-1} & (k\leq K) \\ P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\cdot\frac{\gamma}{\gamma+M-1} & (k=K+1) \end{cases}$$

    (c) activity class $k=k_{jt}$ is assigned to cluster t in the data segment j for sensor feature data points $\boldsymbol{X}_t$:

      $N_{kx^s}^s$++, N_k^s++ for all $x^s\in\boldsymbol{X}_t$, M_k++

    (d) delete activity class that no sensor feature data point belongs to.

  end for

  update concentration parameters α_0 and γ

until converge

Data Segment-Level Cluster Assignment

The posterior probability that the sensor feature data point x^s_{jn} ($s \in \{m, a, v, g\}$) is within cluster t is proportional to the product of the prior probability that a sensor feature data point in the j-th data segment is within cluster t and the probability that the sensor feature data point is likely assigned to activity class k (i.e., $P(t^s_{jn} = t|\boldsymbol{X}^s, \lambda) \propto P(t^s_{jn} = t|\lambda)P(x^s|\boldsymbol{X}^s_k)$).

The prior probability that a sensor feature data point in the j-th data segment within a cluster t can be formulated as:

$$P(t^s_{jn} = t|\lambda) = \begin{cases} \frac{N_{jt}}{\alpha_0 + N_j - 1} & (t = 1, ..., T_j) \\ \frac{\alpha_0}{\alpha_0 + N_j - 1} & (t = T_j + 1) \end{cases} \tag{5.3}$$

where T_j is the number of clusters in the j-th data segment, N_j is the number of sensor feature data points in data segment j, and N_{jt} is the number of sensor feature data points within cluster t in data segment j.

Moreover, the probability that the sensor feature data point is likely to be assigned to activity class k can be written as follows:

$$P(x^s|\boldsymbol{X}^s_k) = \frac{N^s_{kx^s} + \alpha^s_0}{N^s_k + d^s\alpha^s_0} \tag{5.4}$$

where N^s_k and $N^s_{kx^s}$ denote the number of sensor feature data points that have been assigned to activity class k and the number of sensor feature data point x^s that belong to activity class k respectively. The vector $\boldsymbol{X}^s_k$ is the set sensor feature data points of type s that have activity class k and d^s represents the dimension of sensor modality s.

Hence, the posterior probability that the sensor feature data point x^s_{jn} ($s \in \{m, a, v, g\}$) is within cluster t can be written as:

$$P(t^s_{jn} = t|\boldsymbol{X}^s, \lambda) \propto P(t^s_{jn} = t|\lambda)P(x^s|\boldsymbol{X}^s_k) = \begin{cases} P(x^s_{jn}|\boldsymbol{X}^s_{k=k_{jt}}) \cdot \frac{N_{jt}}{\alpha_0 + N_j - 1} & (t \leq T_j) \\ P(x^s_{jn}|\boldsymbol{X}^s_{k=k_{jt}}) \cdot \frac{\alpha_0}{\alpha_0 + N_j - 1} & (t = T_j + 1) \end{cases} \tag{5.5}$$

where N_{jt} represents the number of sensor feature data points assigned to activity class k in data segment j. $\boldsymbol{X}^s_{k=k_{jt}}$ denotes a set of sensor feature data points of type s assigned to activity class k. T_j is the total number of clusters in data segment j.

Dataset-Level Activity Class Assignment

Similarly, the posterior probability that the activity class k is assigned to cluster t is proportional to the product of the prior probability that activity class k is assigned to cluster t and the probability that the sensor feature data points $\boldsymbol{X}_{jt}$ within cluster t are likely to be assigned to activity class k (i.e., $P(k_{jt} = k|\boldsymbol{X}, \gamma) \propto P(k_{jt} = k|\gamma)P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)$).

The prior probability that an activity class k is assigned to the cluster t can be formulated as:

$$P(k_{jt} = k|\gamma) = \{ \begin{array}{ll} \frac{M_k}{\gamma+M-1} & (k = 1, ..., K) \\ \frac{\gamma}{\gamma+M-1} & (k = K + 1) \end{array} \tag{5.6}$$

where K is the number of assigned activity classes, M is the total number of clusters, and M_k is the number of clusters with activity class k served. Next, the probability that the sensor feature data points $\boldsymbol{X}_{jt}$ within cluster t are likely to be assigned to activity class k can be expressed as $P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)$.

Finally, the posterior probability that the activity class k is assigned to cluster t can be written as:

$$P(k_{jt} = k|\boldsymbol{X}, \gamma) \propto P(k_{jt} = k|\gamma)P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k) = \{ \begin{array}{ll} P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k) \cdot \frac{M_k}{\gamma+M-1} & (k \leq K) \\ P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k) \cdot \frac{\gamma}{\gamma+M-1} & (k = K + 1) \end{array} \tag{5.7}$$

The activity classes are picked with respect to the posterior probability and the choice of $k = K + 1$ means the generation of a new activity class.

Low-Level Activity Proportion Estimation

Once the parameter estimation process shown in Algorithm 1 converges, we can derive the low-level activity proportion of each data segment by calculating the proportion of activity class assignments of clusters in a data segment. Mathematically, the low-level activity proportion of data segment j can be written as:

$$\Phi_j = \{\phi_{j1}, ..., \phi_{jK}\} = \{\frac{\sum_t^{T_j} \delta_1(\widehat{k}_{jt})\widehat{N}_{jt}}{N_j}, ..., \frac{\sum_t^{T_j} \delta_K(\widehat{k}_{jt})\widehat{N}_{jt}}{N_j}\} \tag{5.8}$$

where $\widehat{k}_{jt}$ and $\widehat{N}_{jt}$ represent converged values of k_{jt} and N_{jt}. The low-level activity proportion in a temporal window can be seen as a feature vector to represent a high-level routine.

5.5 Evaluation

5.5.1 Evaluation Methodology

As for the evaluation of the proposed MHDP model, we consider a supervised approach for three classification tasks: activity, location, and people recognition. For each classification task, we compare the performance using individual modal data and multimodal data. We also investigate the effect of temporal window duration.

In order to perform classification tasks, we first run parameter estimation on MHDP with multimodal data and HDP with each single modal data. Then, we train a support vector machine (SVM) classifier using the obtained low-level activity proportion vectors and ground truth annotations [8]. The results are validated using standard 10-fold cross-validation. To measure the classification performance, accuracy that is the most standard metric to summarize the overall performance for all classes is used [44]. Accuracy can be defined by:

$$Accuracy = \frac{TP + TN}{TP + TN + NP + FN} \tag{5.9}$$

, where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives respectively.

5.5.2 Effect of Modalities

Figure 5.3 shows classification accuracy over different window durations on the activity, location, and people recognition. Each graph contains five curves, which are acceleration, audio, image, GPS, and all (multimodal) sensory data. The results show that the MHDP model achieves a best accuracy of 80%, 87%, and 83% on activity, location, and people recognition, respectively, which outperforms the results using HDP with a single modality of data. Moreover, the results show that audio information gives suboptimal accuracy over three classification tasks. For people recognition in Figure 5.3(c), the audio data with the HDP model even outperforms the MHDP model with a ten-minute window duration setting (80% against 75%). We also see that in Figure 5.3(b), GPS cannot give high location recognition accuracy. Possible reasons for this are that the GPS has missing data in indoor environments and the vocabulary of the GPS data is unable to distinguish between two locations that are located near each other. The similar effect was also reported in previous work [9].

Let us further look into the detailed accuracy rate of each activity, location, and people class shown in Figure 5.4. In Figure 5.4(a), we can see that accelerometer data results in better performance for the activity class such as "*walk*" that is highly correlated to body movement. Moreover, image data outperforms other sensor modalities for activity classes such as "*drive.inside.a.vehicle*" and "*use.a.computer*". The explanation might be that the images of computer displays or windshields are usually brighter and have salient features. However, GPS data have relatively low accuracy across different activities.

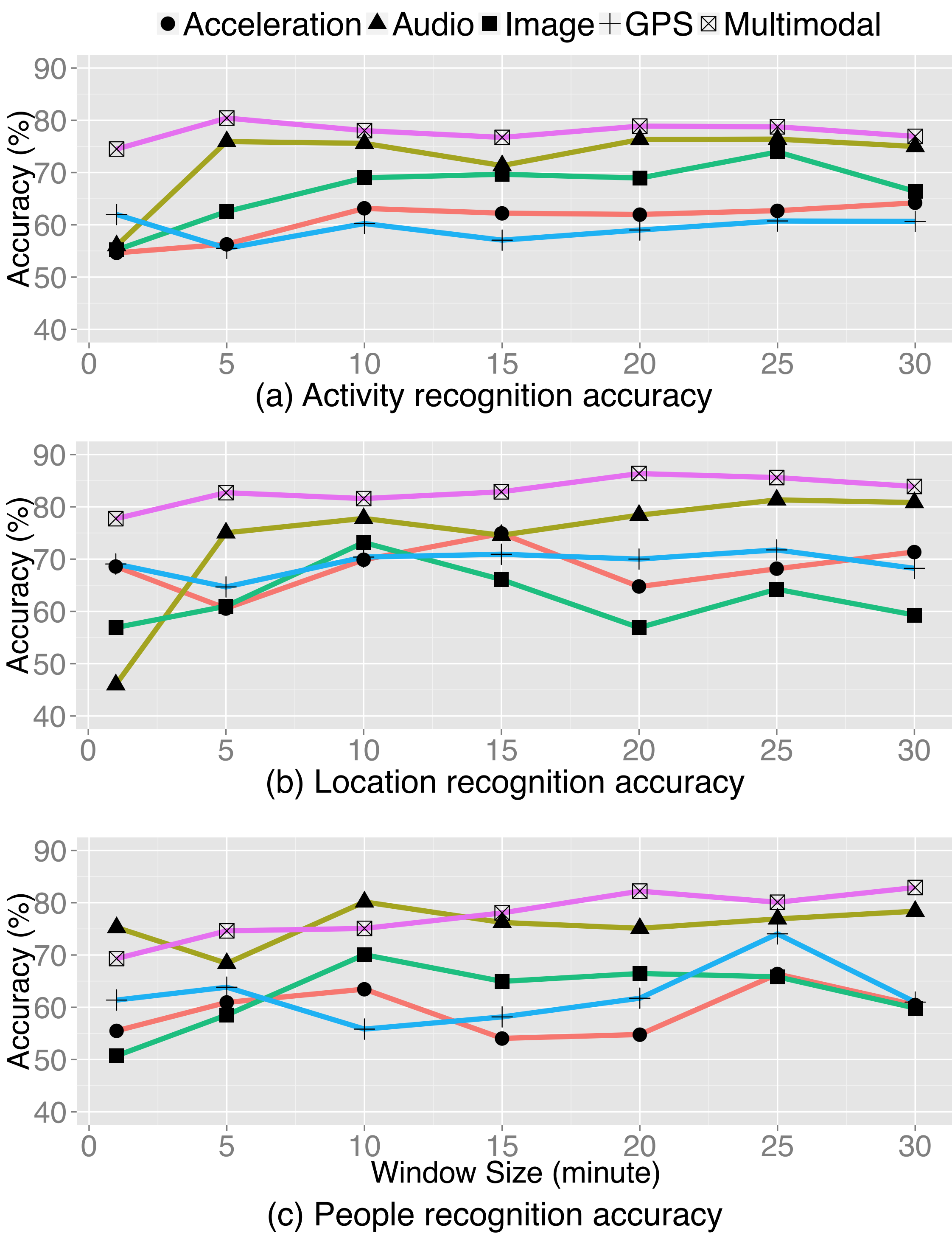


Figure 5.3: Accuracy measure on activity (top), location (middle), and people (bottom) recognition over different window durations (from 1 to 30 minutes). Each single modality of data is used to train a separate HDP model, while MHDP is trained with multimodal data.

In Figure 5.4(b), GPS data have better performance for location classification compared to activity classification. However, as we mentioned before, GPS has missing data in indoor environments so that it cannot distinguish locations between "*my office*" and "*other's office*". Furthermore, audio information provides good accuracies on locations such as "*restaurant/cafe*", "*my home*", and "*my office*" due to the ambient sound signatures are quite distinguishable at these three locations.

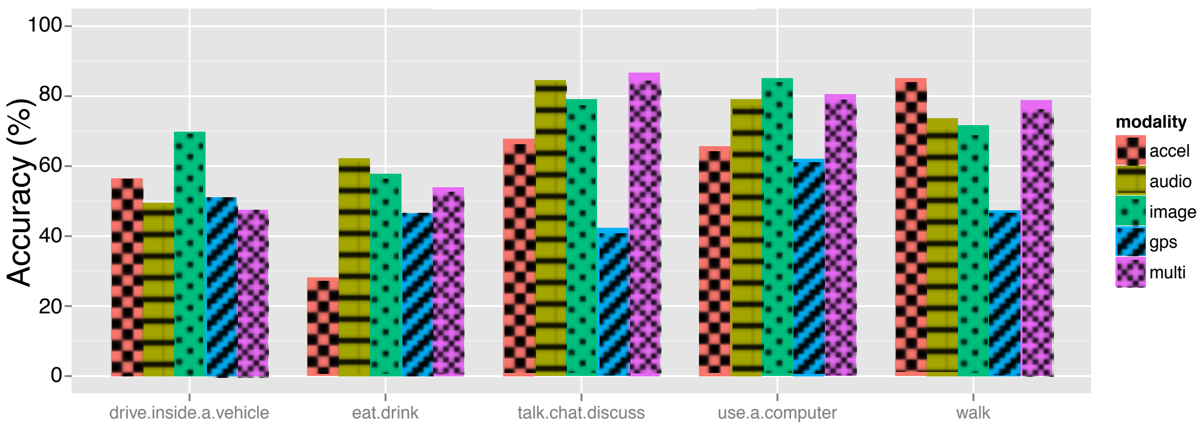
For the detailed accuracy of each people class shown in Figure 5.4(c), we can see that audio data capture the conversation between the user and people surrounding to him/her and results in high accuracy rate.

5.5.3 Effect of Window Duration

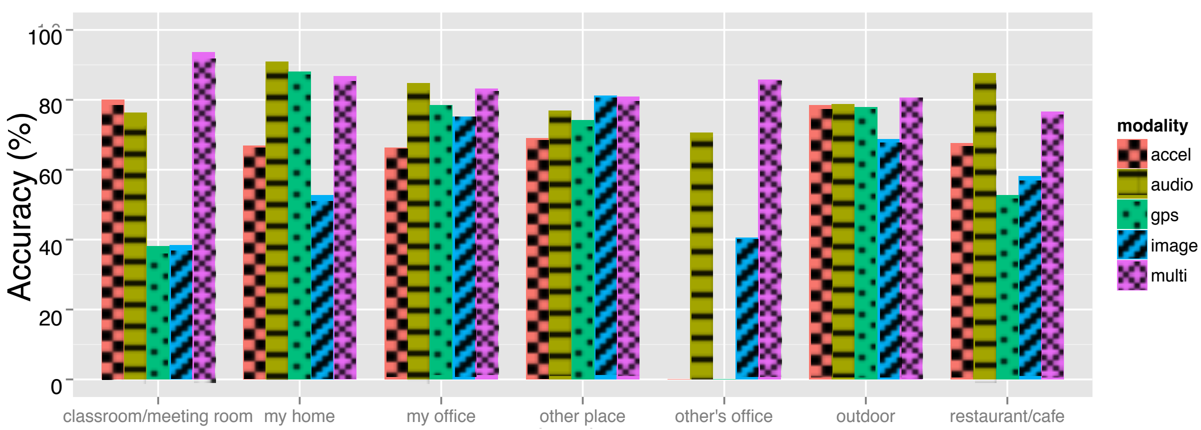
Figure 5.3 shows the MHDP model has optimal accuracy with window durations of 5 minutes (activity), 20 minutes (location), and 30 minutes (people). Moreover, the MHDP model achieves stable ranges of recognition accuracy of $75-80\%$ (activity), $78-87\%$ (location), and $70-83\%$ (people) over different window durations. We also see that the performance of HDP with any modality of data and MHDP with all data tends to increase with the window duration. This effect has also been reported in [68]. In particular, we observe that HDP with audio data significantly improves and results in a stable performance with the increase of window duration from 1 minute to 5 minutes on activity and location recognition.

5.6 Discussion

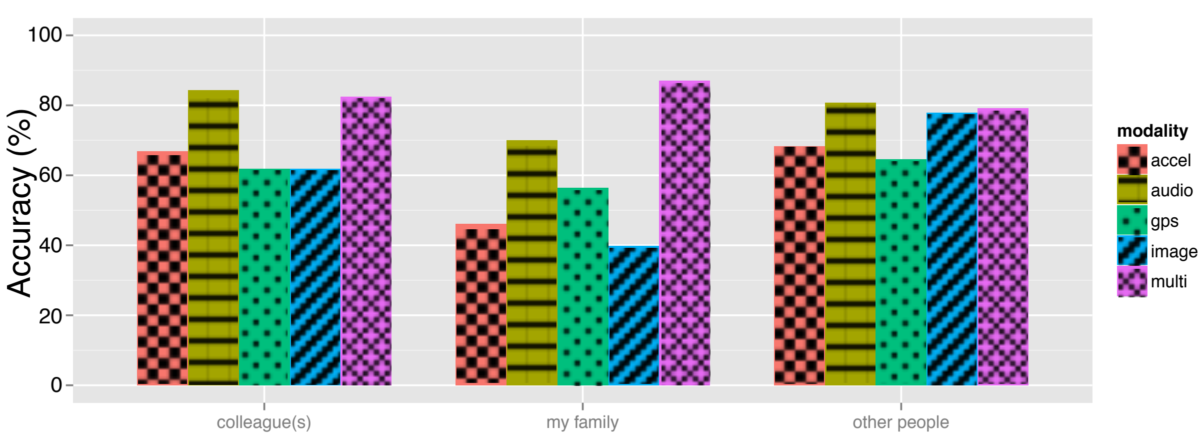
In this chapter, we present the design, implementation, and evaluation of MHDP to model the relationships among multimodal data sources for human behavior discovery. Most of the previous topic model-based approaches incorporate only a single sensor modality (e.g., accelerometer or GPS) or applies simple concatenation from different modalities. They do



(a) Activity Classification



(b) Location Classification



(c) People Classification

Figure 5.4: Accuracy rate of each activity, location, and people class using the multimodal daily activity dataset.

not take into account the informative relationships among different modalities of sensor data. Our proposed MHDP model automatically selects the appropriate number of latent topics and models the relationships among multisensory data streams. Experimental results show that our proposed model can achieve accuracy of 80%, 87%, and 83% on activity, location, and people recognition, respectively, and outperforms the results when using HDP with only a single modality of data.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This thesis presents a study of human behavior pattern discovery and introduces a novel nonparametric framework based on probabilistic topic models that can infer high-level human routines from multimodal sensor data. While there have been a few research projects on high-level human behavior modeling, most existing parametric approaches of mining human behavior patterns have major limitations. Specifically, they assume a fixed, pre-specified parameter regardless of the data. Choosing an appropriate parameter usually requires an inefficient and trial-and-error model selection process. In the context of developing personalized applications, it is even more difficult to find optimal parameter values of models in advance.

Essentially, the proposed framework makes an analogy between sensor data and text in the topic modeling paradigm. Table 6.1 lists the analogy of terms between sensor data and text used in this thesis. In the context of mining a sequence of sensor data, low-level sensor data features are first mapped into a set of discrete labels (vocabulary). Each mapped data feature is equivalent to a text word. Then, the bag-of-features in each temporal window is

Textual data	Sensor data
topic proportions	low-level activity proportions
topics	low-level activities
documents	sensor data segments (temporal windows)
bag-of-words	bag-of-features
vocabulary	feature vocabulary
text words	discretized data features

Table 6.1: List of the analogy of terms between text and sensor data.

referred to as a sensor data segment (document) and used to train the topic model. Sensor data segments belong to the same routine if they have similar topic proportions. The developed approaches discover the latent high-level routines (e.g., "office work" or "commuting") in a fully unsupervised fashion from raw sensor data obtained on mobile devices. Moreover, in order to further model the relationships among concurrent multimodal data, a generative probablistic model is proposed and evaluated by performing activity classification tasks.

To overcome the existing challenges in parametric settings, the proposed approach applies nonparametric Bayesian methods for two major phases: (1) low-level feature vocabulary creation, and (2) high-level routine discovery. At the low-level feature vocabulary creation phase, we applied Dirichlet process Gaussian mixture model (DPGMM) to automatically find the size of the vocabulary from sensor feature vectors and compose the bag-of-features representation for each temporal window. Next, at the routine discovery phase, we further used hierarchical Dirichlet process (HDP) model and its extension to automatically select the appropriate number of latent activities and estimate the latent activity proportion from a collection of sensor data segments represented by the bag-of-features format. Finally, by using affinity propagation (AP) clustering algorithm, sensor data segments with similar low-level activity proportions are discovered as the same latent routine.

The approaches and results presented in this thesis are evaluated on public datasets in two routine domains: two daily-activity datasets and a transportation mode dataset. Experimental results show that our nonparametric framework can automatically learn the appropriate model parameters without any form of manual model selection procedure and the proposed nonparametric models outperforms traditional parametric approaches (i.e., latent Dirichlet allocation [LDA] and K-means) for human routine discovery tasks. Moreover, the proposed generative model with multimodal data (i.e., multimodal hierarchical Dirichlet process [MHDP]) can achieve accuracy of 80%, 87%, and 83% on activity, location, and people recognition, respectively, and have better results when using HDP with a single modality of data.

6.1.1 Contribution

Based on the previous work in the fields of activity recognition and text modeling, this thesis made contributions on developing a nonparametric framework for high-level human routine discovery from multimodal data. More specifically, the design and implementation of a nonparametric framework for high-level human behavior discovery without the need of a trial-and-error model selection process are presented. A probablistic graphical model that models the relationships among a sequence of multimodal sensor data for human behavior discovery is developed. Finally, the proposed framework is evaluated on three realistic public datasets in human activity and transportation mode domains.

6.2 Future Work

Although we have made several valuable contributions to the field of human routine discovery, our framework can be further improved. In the next subsections, we outline possible future research directions to extend the work done in this thesis. First, the temporal information of low-level data features within a temporal window can be taken into account while discovering high-level routines. Second, if we can further learn the hierarchical structure of human routines from sensor data, users' lifestyles can be better profiled. Third, a fixed temporal window duration to compose the bag-of-features representation can be adjusted to a variable duration based on the similarity of sensor data features. Finally, in terms of developing a practical mobile application, an online inference algorithm is important because it only requires to train with new data rather than the entire dataset.

6.2.1 Temporal Order of Low-Level Data Features

The human routine discovery framework presented in this work is based on the "bag-of-features" representation. The bag-of-features representation assumes that the order of features (i.e., mapped sensor data features) within each temporal window is ignored. However, when a person performs an activity, the captured sensor data stream is a time-series signal with strong temporal dependancy. By incorporating n-gram statistics with a unigram topic model, it allows us capture and model the pattern of activity sequences within a data segment [79]. This concept of combining n-gram statistics with the probabilistic topic model has shown promise in the field of computer vision [74, 78]. By extending the proposed non-parametric framework to encode n-gram statistics, we can take into account temporal order of low-level activities within each temporal window and achieve a better routine discovery performance.

6.2.2 Hierarchies of Routines

The approaches presented in this thesis aim to discover high-level routines using topic models. However, using topic models such as latent Dirichlet (LDA) or hierarchical Dirichlet process (HDP), the topics are discovered in a flat structure rather than organized into a hierarchical structure such as that naturally exhibited in text data [40]. There have been previously proposed topic models that learn the correlations and the hierarchical structure of topics for the text corpus [4, 40, 55]. Similarly, human activities in daily life can be categorized into a hierarchy [2, 36]. For example, a user's daily life might consist of major routines (e.g., recreation) and sub-routines (e.g., *eat meal outside* or *visit friends/relatives*) [36]. Considering the hierarchy of human routines as a tree, one branch of the tree might have the routine *recreation* as a parent topic and the routines *eat meal outside* and *visit friends/relatives* as child topics. This hierarchical structure learning can be done with a tree generation process based on the Chinese restaurant process [40]. Hence, as an improvement, we can better understand the structure of human routines by learning hierarchies of human routines from sensor data.

6.2.3 Automatic Selection of Temporal Window Duration

The framework presented in this thesis uses a fixed duration of temporal window (e.g., 30 minutes) to compose bag-of-features for high-level routine discovery tasks. We assume that each temporal window's duration is shorter than each routine to be discovered. We exhaustively evaluated the influence of temporal window duration in terms of routine discovery performance. Since the stability of topic model-based routine discovery might vary according to the durations of routines and temporal windows, it is reasonable to have a more systematic method to determine the duration of sensor data segments [68]. Some previous work in the field of activity recognition focused on recursive segmentation of the sensor data based on the similarity between two adjacent windows [12]. However, we should avoid the possibility that segments are too small that reflect to the changes of low-level features and lack the bag-of-features property for latent routine discovery tasks.

6.2.4 Online Inference of Routines

In the proposed routine discovery framework, we assume that inference algorithms (i.e., topic proportion estimation) run over an entire dataset of sensor streams after they have been collected. But, in many ubiquitous computing applications, large amounts of sensor data are continuously generated from multiple mobile devices and the model must also be adjusted accordingly. It makes inference algorithms infeasible to run offline repeatedly due to the computational complexity. This problem has been addressed in the text mining domain by using online inference algorithms that update topic estimations when a new document is observed for topic modeling [11, 31]. Therefore, future work in this direction can make the proposed routine discovery framework more applicable to real world scenarios that are more dynamic.

6.3 Final Thoughts

With the groundwork of technologies such as sensors, networking, and cloud computing in place, continuous and large amounts of sensory data from heterogeneous sources are more accessible nowadays. It is an exciting time for us to study the field of human behavior pattern modeling from sensor data. As we investigated in this research, text mining and human activity modeling share many common properties. Hence, as a future research direction, it will be interesting to adopt and integrate real-time textual stream mining techniques (e.g., mining the semantics of social media textual streams) in human activity recognition and discovery domains. It will enable us to handle large-scale sensor data streams and develop more proactive context-aware applications.

Bibliography

[1] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. pages 1–17. Springer, 2004.

[2] U. Blanke and B. Schiele. Remember and transfer what you have learned - recognizing composite activities based on activity spotting. In *International Symposium on Wearable Computers*, 2010.

[3] D. M. Blei. Probabilistic topic models. *Commun. ACM*, 55(4):77–84, Apr. 2012.

[4] D. M. Blei, T. L. Griffiths, M. I. Jordan, and J. B. Tenenbaum. Hierarchical topic models and the nested Chinese restaurant process. In *Advances in Neural Information Processing Systems*, page 2003. MIT Press, 2004.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, Mar. 2003.

[6] M. Blighe, H. L. Borgne, N. E. Oconnor, A. F. Smeaton, and G. J. F. Jones. Exploiting context information to aid landmark detection in sensecam images. In *Proceedings of International Workshop on Exploiting Context Histories in Smart Environments*, 2006.

[7] M. Buettner, R. Prasad, M. Philipose, and D. Wetherall. Recognizing daily activities with RFID-based sensors. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, pages 51–60, New York, NY, USA, 2009. ACM.

[8] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data*

Mining Knowledge Discovery, 1998.

[9] S. Buthpitiya. *Modeling Mobile User Behavior For Anomaly Detection*. PhD thesis, Carnegie Mellon University, Pittsburgh, Mar. 2014.

[10] S. Buthpitiya, Y. Zhang, A. Dey, and M. Griss. n-gram geo-trace modeling. In *Pervasive Computing*, pages 97–114. Springer Berlin Heidelberg, 2011.

[11] K. R. Canini, L. Shi, H. W. Neuroscience, and T. L. Griffiths. Online inference of topics with latent Dirichlet allocation. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2009.

[12] P.-W. Chen, S. K. Chennuru, S. Buthpitiya, and Y. Zhang. A language-based approach to indexing heterogeneous multimedia lifelog. In *ICMI-MLMI*, page 26, 2010.

[13] H.-T. Cheng. *Learning and Recognizing The Hierarchical and Sequential Structure of Human Activities*. PhD thesis, Carnegie Mellon University, Pittsburgh, Dec. 2013.

[14] H.-T. Cheng, M. Griss, P. Davis, J. Li, and D. You. Towards zero-shot learning for human activity recognition using semantic attribute sequence model. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 355–358, New York, NY, USA, 2013. ACM.

[15] H.-T. Cheng, F.-T. Sun, M. Griss, P. Davis, J. Li, and D. You. NuActiv: recognizing unseen new activities using semantic attribute-based learning. In *Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '13, pages 361–374, New York, NY, USA, 2013. ACM.

[16] P.-W. Cheng, S. Chennuru, S. Buthpitiya, and Y. Zhang. A language-based approach to indexing heterogeneous multimedia lifelogs. In *Proceedings of the International Conference on Multimodal Interfaces, ICMI*, pages 26:1–26:8, New York, NY, USA, 2010. ACM.

[17] T. Choudhury and A. Pentland. Sensing and modeling human networks using the sociometer. In *Wearable Computers, 2003. Proceedings. Seventh IEEE International Symposium on*, pages 216–222.

[18] D. Cook. Learning setting-generalized activity models for smart spaces. *Intelligent Systems, IEEE*, 27(1):32–38, Jan 2012.

[19] D. Dueck. *Affinity Propagation: Clustering Data By Passing Messages*. PhD thesis, University of Toronto, Canada, June 2009.

[20] D. Dueck and B. Frey. Non-metric affinity propagation for unsupervised image categorization. In *IEEE 11th International Conference on Computer Vision*, pages 1–8, Oct 2007.

[21] N. Eagle and A. Pentland. Eigenbehaviors: Identifying structure in routine. In *Behavioral Ecology and Sociobiology*, pages 1057–1066, 2009.

[22] K. Farrahi and D. Gatica-Perez. Discovering routines from large-scale human locations using probabilistic topic models. *ACM Trans. Intell. Syst. Technol.*, 2(1):3:1–3:27, Jan. 2011.

[23] T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 03 1973.

[24] B. J. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315:2007, 2007.

[25] S. J. Gershman and D. M. Blei. A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1):1 – 12, 2012.

[26] N. Györbíró, A. Fábián, and G. Hományi. An activity recognition system for mobile phones. *Mob. Netw. Appl.*, 14(1):82–91, Feb. 2009.

[27] J. Hamm, B. Stone, M. Belkin, and S. Dennis. Automatic annotation of daily activity

from smartphone-based multisensory streams. In *Proceedings International Conference Mobile Computing, Applications and Services, MobiCASE*, 2012.

[28] B. Hariharan, S. V. Vishwanathan, and M. Varma. Efficient max-margin multi-label classification with applications to zero-shot learning. *Journal of Machine Learning Research*, 88(1-2):127–155, July 2012.

[29] B. L. Harrison, S. Consolvo, and T. Choudhury. Using multi-modal sensing for human activity modeling in the real world. In *Handbook of Ambient Intelligence and Smart Environments*, 2010.

[30] N. Hjort, C. Holmes, P. Mueller, and S. Walker. *Bayesian Nonparametrics: Principles and Practice*. Cambridge University Press, Cambridge, UK, 2010.

[31] M. D. Hoffman, D. M. Blei, and F. R. Bach. Online learning for latent Dirichlet allocation. In *NIPS*, pages 856–864, 2010.

[32] Y.-J. Hong, I.-J. Kim, S. C. Ahn, and H.-G. Kim. Activity recognition using wearable sensors for elder care. In *Future Generation Communication and Networking, 2008. FGCN '08. Second International Conference on*, pages 302–305, 2008.

[33] D. H. Hu, X.-X. Zhang, J. Yin, V. W. Zheng, and Q. Yang. Abnormal activity recognition based on HDP-HMM models. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 1715–1720, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.

[34] T. Huynh. *Human Activity Recognition with Wearable Sensors*. PhD thesis, Technische Universtat Darmstadt, Darmstadt, May 2008.

[35] T. Huynh, M. Fritz, and B. Schiele. Discovery of activity patterns using topic models. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, pages 10–19, New York, NY, USA, 2008.

[36] S. Jiang, J. F. Jr, and M. González. Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery*, 25:478–510, 2012.

[37] P. Kankuekul, A. Kawewong, S. Tangruamsub, and O. Hasegawa. Online incremental attribute-based zero-shot learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3657–3664, 2012.

[38] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Wu. An efficient k-means clustering algorithm: analysis and implementation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):881–892, Jul 2002.

[39] E. Kim, S. Helal, and D. Cook. Human activity recognition and pattern discovery. *IEEE Pervasive Computing*, 9(1):48–53, Jan. 2010.

[40] J. H. Kim, D. Kim, S. Kim, and A. Oh. Modeling topic hierarchies with the recursive Chinese restaurant process. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, pages 783–792, New York, NY, USA, 2012. ACM.

[41] A. Krishnamurthy. High-dimensional clustering with sparse Gaussian mixture models. In *Technical Report*. Carnegie Mellon University, Pittsburgh, 2011.

[42] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity recognition using cell phone accelerometers. *SIGKDD Explor. Newsl.*, 12(2):74–82, Mar. 2011.

[43] O. Lara and M. Labrador. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys Tutorials*, 15(3):1192–1209, 2013.

[44] N. Lavesson, N. Lavesson, N. Lavesson, and N. Lavesson. *Evaluation and Analysis of Supervised Learning Algorithms and Classifiers*. PhD thesis, Blekinge Institue of Technology, 2006.

[45] M. Lavine. More aspects of Polya tree distributions for statistical modelling. *The Annals*

of Statistics, 22(3):1161–1176, 09 1994.

[46] Q. Li, J. Stankovic, M. Hanson, A. Barth, J. Lach, and G. Zhou. Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information. In *Wearable and Implantable Body Sensor Networks, 2009. BSN 2009. Sixth International Workshop on*, pages 138–143, June 2009.

[47] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya, and V. Vasudevan. uwave: Accelerometer-based personalized gesture recognition and its applications. In *Proceedings of the International Conference on Pervasive Computing and Communications*, PerCom '09, pages 1–9, March 2009.

[48] B. Longstaff, S. Reddy, and D. Estrin. Improving activity classification for health applications on mobile devices using active and semi-supervised learning. In *International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth'10, pages 1–7, 2010.

[49] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. 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.

[50] H. Lu, J. Yang, Z. Liu, N. D. Lane, T. Choudhury, and A. T. Campbell. The jigsaw continuous sensing engine for mobile phone applications. In *Proceedings of the ACM Conference on Embedded Networked Sensor Systems*, SenSys '10, pages 71–84, 2010.

[51] M. Mahdaviani and T. Choudhury. Fast and scalable training of semi-supervised CRFs with application to activity recognition. In *Advances in Neural Information Processing Systems*, pages 977–984, Cambridge, MA, 2007. MIT Press.

[52] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.

[53] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks*, BSN '06, pages 113–116, Washington, DC, USA, 2006. IEEE Computer Society.

[54] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell. Sensing meets mobile social networks: the design, implementation and evaluation of the CenceMe application. In *Proceedings of the ACM Conference on Embedded Network Sensor Systems*, SenSys '08, pages 337–350, 2008.

[55] D. Mimno, W. Li, and A. McCallum. Mixtures of hierarchical topics with pachinko allocation. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 633–640, New York, NY, USA, 2007. ACM.

[56] D. Minnen, T. Starner, I. Essa, and C. Isbell. Discovering characteristic actions from on-body sensor data. *Proceedings of International Symposium on Wearable Computers*, pages 11–18, 2006.

[57] M. F. Mokbel and J. J. Levandoski. Toward context and preference-aware location-based services. In *Proceedings of the Eighth ACM International Workshop on Data Engineering for Wireless and Mobile Access*, MobiDE '09, pages 25–32, 2009.

[58] M. Mun, P. Boda, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, and R. West. PEIR, the personal environmental impact report, as a platform for participatory sensing systems research. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services*, pages 55–68, New York, New York, USA, 2009.

[59] T. Nakamura, T. Nagai, and N. Iwahashi. Multimodal categorization by hierarchical Dirichlet process. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 1520–1525, Sept 2011.

[60] R. M. Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265, 2000.

[61] T. Nguyen, D. Phung, S. Gupta, and S. Venkatesh. Extraction of latent patterns and contexts from social honest signals using hierarchical Dirichlet processes. PerCom '13, pages 47–55, Los Alamitos, CA, USA, 2013. IEEE Computer Society.

[62] O. Oreifej and Z. Liu. Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 0:716–723, 2013.

[63] A. S. Pentland, D. Lazer, D. Brewer, and T. Heibeck. Using reality mining to improve public health and medicine. In *Studies in Health Technology and Informatics*, 2009.

[64] C. Perera, A. B. Zaslavsky, P. Christen, and D. Georgakopoulos. Context aware computing for the internet of things: A survey. *CoRR*, abs/1305.0982, 2013.

[65] C. Reed. Latent Dirichlet allocation: Toward a deeper understanding. In *Technical Report*. University of California, Berkeley, 2012.

[66] M. S. Ryoo and J. K. Aggarwal. Recognition of high-level group activities based on activities of individual members. In *Proceedings of the 2008 IEEE Workshop on Motion and Video Computing*, WMVC '08, pages 1–8, Washington, DC, USA, 2008. IEEE Computer Society.

[67] T. S. Saponas, D. S. Tan, D. Morris, J. Turner, and J. A. Landay. Making muscle-computer interfaces more practical. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 851–854, New York, NY, USA, 2010. ACM.

[68] J. Seiter, O. Amft, and G. Tröster. Assessing topic models: How to obtain robustness? In *AwareCast 2012: Workshop on Recent Advances in Behavior Prediction and Proactive Pervasive Computing*, 2012.

[69] J. Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

[70] M. Stikic, D. Larlus, S. Ebert, and B. Schiele. Weakly supervised recognition of daily life activities with wearable sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.

[71] M. Stikic, K. Van Laerhoven, and B. Schiele. Exploring semi-supervised and active learning for activity recognition. In *International Symposium on Wearable Computers*, pages 81–88, 2008.

[72] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss. Activity-aware mental stress detection using physiological sensors. In *Proceedings International Conference Mobile Computing, Applications and Services, MobiCASE*, 2012.

[73] F.-T. Sun, Y.-T. Yeh, H.-T. Cheng, C. Kuo, and M. Griss. Nonparametric of human routines from sensor data. In *Proceedings International Conference on Pervasive Computing and Communications*, PerCom '14. IEEE, 2014.

[74] R. Tavenard, R. Emonet, and J.-M. Odobez. Time-sensitive topic models for action recognition in videos. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 2988–2992, Sept 2013.

[75] Y. W. Teh and M. I. Jordan. *Hierarchical Bayesian Nonparametric Models with Applications.* Cambridge University Press, 2009.

[76] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet processes. In *Journal of the American Statistical Association*, volume 101, 2004.

[77] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea. Machine recognition of human activities: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(11):1473–1488, 2008.

[78] J. Varadarajan, R. Emonet, and J.-M. Odobez. Probabilistic latent sequential motifs: Discovering temporal activity patterns in video scenes. In *BMVC 2010*. Aberystwyth University, BMVA Press, 9 2010.

[79] H. M. Wallach. Topic modeling: Beyond bag-of-words. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pages 977–984, New York, NY, USA, 2006. ACM.

[80] Y. Wang, J. Lin, M. Annavaram, Q. A. Jacobson, J. Hong, B. Krishnamachari, and N. Sadeh. A framework of energy efficient mobile sensing for automatic user state recognition. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services*, MobiSys '09, pages 179–192, 2009.

[81] L. Wasserman. Bayesian model selection and model averaging. *J. Math. Psychol.*, 44(1):92–107, Mar. 2000.

[82] O. Yakhnenko and V. Honavar. Multi-modal hierarchical Dirichlet process model for predicting image annotation and image-object label correspondence. In *SIAM SDM*, 2009.

[83] Q. Yang. Activity recognition: linking low-level sensors to high-level intelligence. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 20–25, 2009.

[84] J. Yin, Q. Yang, and J. Pan. Sensor-based abnormal human-activity detection. *IEEE Transactions on Knowledge and Data Engineering*, 20(8):1082–1090, Aug 2008.

[85] X. Yu. Gibbs sampling methods for Dirichlet process mixture model: Technical details. 2009.

[86] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on GPS data. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, pages 312–321, New York, NY, USA, 2008. ACM.

[87] Y. Zhu, Y. Arase, X. Xie, and Q. Yang. Bayesian nonparametric modeling of user activities. In *Proceedings of the 2011 International Workshop on Trajectory Data Mining and Analysis*, pages 1–4, New York, NY, USA, 2011. ACM.