Title
Building Personal Chronicle of Life Events

Author
Oh, Hyungik

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Building Personal Chronicle of Life Events

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Computer Science

by

Hyungik Oh

Dissertation Committee:
Professor Ramesh Jain, Chair
Professor Michael J. Carey
Professor Nikil Dutt

2019

DEDICATION

To God. I would like to bring all this honor to him.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

John 6:20. But he said to them, “It is I; don’t be afraid.”

Every time that I’d wanted to give up, I could overcome all the things through Christ who strengthens me. I thank His almighty, His forever love and His many blessings.

I would like to thank Professor Ramesh Jain who is my academic advisor, colleague and one of the best friends in my life. In 2013, I was inspired by his vision for helping people live better lives through the Personicle. I still remember the moment when I met him in his office. I, a passionate novice who only had a software development skill, just knocked on his door and told him that I can make this happen. Since the meeting in the fall of 2013, our passion for the objective-self has made the Personicle become a reality. I enjoyed tons of discussion with him, his guide and his feedback on this research direction. I appreciate his full support for my research and my dream.

I would like to thank my thesis committee member Professor Michael Carey. I appreciate his advice and wise counsel on the thesis, and his interest in the Personicle. His active involvement in building an efficient data model for the Personicle has been extremely helpful to make it more robust. I also would like to thank another thesis committee member Professor Nikil Dutt. I enjoyed all the discussion about the Personicle and its application and future. His highly perceptive comments and visions on the Personicle have made it possible to move forward.

I would like to thank my family. I couldn’t make this happen without their love, pray and encouragement. I would like to say I really and truly love you to my mom, Misuk, and dad, Jae Yeon, and thank them for making my dreams come true. I appreciate their sacrifice and unreserved support. I am very proud of my parents. I thank my sister Jiyoun wholeheartedly for her dedication to our family. I also would like to thank my parents in law, Paul Dukshin and Yunho, for their endless love, pray, encouragement, support, and every faith in me. Lastly, my dearest wife, Eunice, thank you so much for believing me and loving me and letting me go on whenever I have challenges. You are the greatest gift of my life. I love you so much.

I also would like to thank my labmate, Laleh Jalali, for initiating the Personicle project and laiding the foundation of its concepts. I would like to thank Kevin Chian for proofreading this thesis. I’m also grateful to Taewoo who helped me from the very first of this journey. I would like to thank especially BCD family, Dohyun, Jihyun, and Ellie. I could get through this tricky time with them. I love you all! Lastly, I would like to thank Santamo family, Jiwon, Christine, Tom, Rose, Sein, Jungeun, Hosung, Joyce, who is always praying for me and encouraging me all the time.

VITA

Hyungik Oh

EDUCATION

Doctor of Philosophy in Computer Science **2020**
University of California, Irvine *Irvine, California*

Master of Science in Computer Science **2015**
University of California, Irvine *Irvine, California*

Bachelor of Science in Information and Telecommunication **2012**
Soongsil University *Seoul, Korea*

RESEARCH EXPERIENCE

Graduate Research Assistant **2015–2020**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant / Reader **2015–2017**
University of California, Irvine *Irvine, California*

PATENT APPLICATION

Multimodal food journaling **2019**
UC Case 2019-217-0, Co-Inventors: Ramesh Jain

Automatic personal daily activity tracking **2018**
UC Case 2018-228-0, Co-Inventors: Ramesh Jain

Polaris: Lifestyle Guide for Diabetes **2018**
UC Case 2018-832-0, Co-Inventors: Ping Wang, Ramesh Jain

PUBLICATIONS

Detecting Events of Daily Living Using Multimodal Data **2019**
arXiv preprint arXiv:1905.09402

Multimodal Food Journaling **2018**
Workshop on Multimedia for Personal Health and Health Care (ACM MM)

Converting Data to Actionable Health Information **2018**
Short Workshop on Next Steps Towards Long Term Self Tracking (ACM CHI)

From Multimedia Logs to Personal Chronicles **2017**
ACM on Multimedia Conference (ACM MM)

Cybernetic Health **2017**
arXiv preprint arXiv:1705.08514

Human Behavior Analysis from Smartphone Data Streams **2016**
Workshop on Human Behavior Understanding

An intelligent notification system using context from real-time personal activity monitoring **2015**
IEEE International Conference on Multimedia and Expo (ICME)

Personicle: personal chronicle of life events **2014**
Workshop on Personal Data Analytics in the Internet of Things (VLDB)

ABSTRACT OF THE DISSERTATION

Building Personal Chronicle of Life Events

By

Hyungik Oh

Doctor of Philosophy in Computer Science

University of California, Irvine, 2019

Professor Ramesh Jain, Chair

Human beings have always been interested in understanding themselves and their surroundings. Learning about the relationship between the two can reveal facts of the present and help predict the future, a critical part to live a better life. With the proliferation of IoT sensor devices, it is now possible to collect quality data for each individual and utilize this data for building personal models that can help to understand the self and environment. However, since this sensor data have different granularities and semantics, the semantic gap becomes even more formidable. Thus, there are challenges in aggregating, integrating, and synchronizing this heterogeneous data to a form such that it effectively describes the life experiences of each individual. In this dissertation, we design a personal chronicle, which contributes a solution to the aforementioned challenges, called Personicle, in which all kinds of personal data streams can be correlated with one another to form a model of a person.

To implement the Personicle, we first attempt to bridge the semantic gap between the low-level multimedia logs and high-level semantics by developing a common daily event model through the data unobtrusively obtained from smart devices. To do this, we define an atomic interval, which brings together the scattered sources of heterogeneous data to partition the data into manageable pieces. This atomic interval lets us segment a day into sequences of similar patterns and use the segments for daily event recognition.

Secondly, we design an event-triggered Ecological Momentary Assessment (EMA) to maximize the chance of aggregating the semantic data from the users. Unlike the traditional EMA process, which mainly depends on user initiative and intervention, we contribute to overcoming the problems endemic to persistent data collection, such as missing a logging moment or early abandonment, by initiating the EMA process from the system side at the right moment.

Lastly, we propose a fully-automated approach to obtain latent semantic information from all the integrated data aiming to maximize the opportunity of both qualitatively and quantitatively capturing one's life experiences. To show a concrete example of this enrichment, we perform an experiment with "Eating" and "Working", a complex event central to human experiences. These enhanced daily events can then be used to create a personal model that could capture how a person reacts to different stimuli under specific conditions.

Chapter 1

Introduction

Observational data has been used heavily in experiments for building scientific models of individuals. Internet-based companies, such as Google, Amazon, and Facebook, put a lot of effort into monitoring users' activity logs on their respective websites. This data can then be used to construct the personal profiles which drive the companies' revenue, fueling well-placed advertisements. In the field of multimedia, life logs have been lauded as meaningful observational data that can create a log of a person's life activities. Continuously monitoring a person's everyday life may provide valuable data for building personal models, leading to deep insights on his or her current state and predicting future situations. Recently, sensors have become present in millions of personal devices and thus can provide data about one's everyday life. All the measurements obtained from these sensors can be used to detect concrete events of daily living as well as the relevant attributes of the events. Therefore, it is possible to effectively organize a chronicle of the person's daily events, called Personicle, in which all kinds of personal data can be correlated with one another to form a model of a person. The medical community has also shown the importance of observational data on healthcare research. Over the years, various attempts have been made to predict epidemics, cure disease, and improve the quality of life by observing a person's everyday lives and

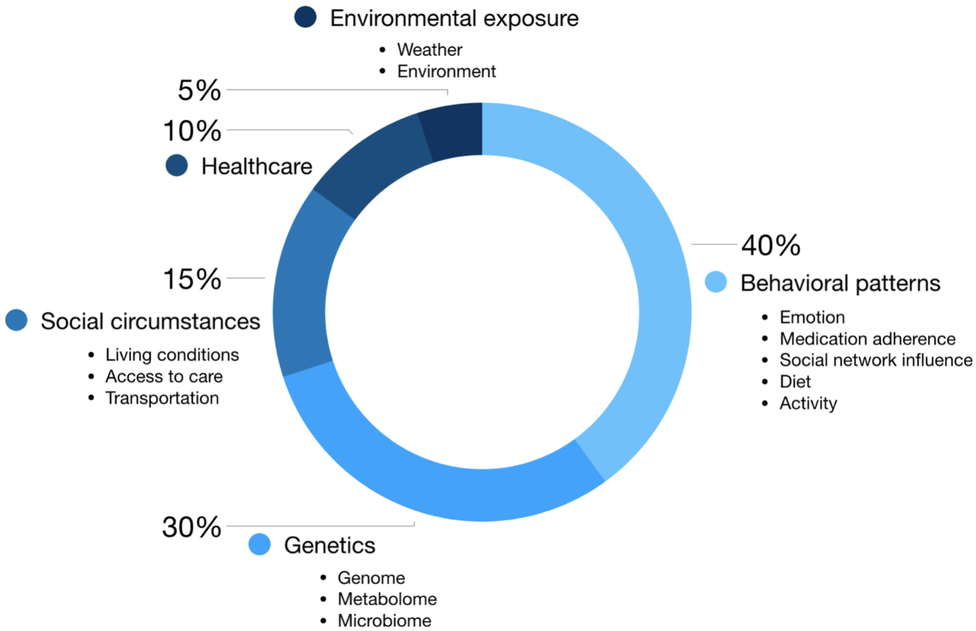


Figure 1.1: The factors influencing health released by the World Health Organization (WHO).

relevant data as shown in Figure 1.1 [16, 33, 120]. Such trends pave the way toward "P4 health", which is an approach to make healthcare more predictive, preventive, personalized, and participatory. There is now a growing interest in using this observational data for studies relating to humans. Although these approaches have shown great potential, a serious challenge, and hence a great opportunity for multimedia research, is that all the multi-modal data should be effectively represented in the context of each individual.

To understand the preferences and particularities of an individual, researchers must build personal models. The personal model can be represented as a personalized knowledge of how a person reacts to different stimuli under specific conditions or how their physiology changes from an intervention [81]. Therefore, the primary consideration in building the personal model, which establishes the premise that each person is a unique entity, is to describe how various factors uniquely drive each individual's personal state. To extract this personalized knowledge, it requires aggregating long- and short-term information about a person and analyzing how these data interact with each other. We can take a hint for what the long- and short-term information is from an example shown in Figure 1.1. Long-term information of a person could come from the genetic factors as well as the history of behavioral patterns. Biological and physical sensors could capture the short-term information through the factors like glucose level, resting heart rate, physical activity, and environmental exposure. In addition, more complex state factors such as emotion or stress levels can be determined by processing the short-term information. Both short-term and long-term information constantly change and interact with one another, and therefore make model-building a process of dynamic and causal understanding [81].

More importantly, to build the personal model, it is necessary to efficiently aggregate and integrate heterogeneous data available to us in a format that can represent all the correlated data together. To solve this challenge, and thus contribute to building the personal model, in 2013, we began this Personicle project, the first of its kind aimed at collecting different sorts of personal data streams in the form of event chronicle. More specifically, we developed an idea that a chronological and continuous analysis of each individual's life events can help decipher the correlation and causality of their personal state [52]. We believe that life events can include the unique lifestyle of each individual that may impact their personal state, such as work they do, food they eat, their interaction with people as well as health-related habits like smoking. It has become more obvious in recent decades that these life events can also significantly contribute to health and survival and even disease exposure and mortality [90]. For this reason, Personicle was designed in a way that can parallelize a sequence of life events with other personal data streams on the same timeline, and thus relate the different streams to one another in the context of life events.

Following that, we identified several challenges, which are necessary to be explored in depth to implement Personicle.

(i) Much of multimodal sensor data has different meanings and granularity and is scattered in isolated silos. This data is in the form of data streams and thus is necessary to be synchronized with one another. Most of all, the primary challenge for obtaining a person's life experiences from these data streams is to bridge the semantic gap between the low-level multimedia logs and high-level semantics. It is required to study how to effectively aggregate, integrate and synchronize multimodal data streams from different silos, and then derive knowledge of each individual's evolving situation for understanding their life experiences.

(ii) Although there is a possibility of bridging the semantic gap by using state of the art technology, to recognize a person's enhanced life experiences, there still needs to aggregate and integrate as many high-quality data as possible. For that reason, Ecological Momentary Assessment (EMA) has been used to collect a person's current behavior and experiences in the natural environment via a smartphone application [18]. On the one hand, it has shown many potential benefits in regards to semantic data collection but on the other hand, it tends to be unreliable and requires many actions on the user's part which can then lead to the problems endemic, such as missing a logging moment and early abandonment. There needs to be an advanced EMA mechanism that can maximize the chance of aggregating the semantic data from the users.

(iii) To qualitatively capture one's life experiences, and thus contribute to building a robust model, personal data streams need to be semantically enriched. Moreover, to quantify the enriched data streams as much as possible, it is required to adopt and develop a fully automated method. Most of the contemporary studies either focused on recognizing simple semantics with specific experimental settings or worked on context enrichment for particular research, such as the care of dementia, rather than understanding the overall life experiences of human beings. This challenge motivated our attention to dive into recognizing events of daily living and developing its fully automated method.

Therefore, this thesis handles the above challenges in three major parts. The summary of each part is as follows:

From Multimedia Logs to Daily Activities: Chapter 3 describes a way to collect, integrate, and segment multi-modal sensor data streams, and how to capture the daily experiences of people with the collection of data. Multi-modal data streams are essential for analyzing personal life, environmental conditions, and social situations. Since these data streams have different granularities and semantics, the semantic gap becomes even more formidable. To make sense of all the multi-modal correlated streams we must first synchronize them in the context of the application, and then analyze them to extract meaningful information. In this chapter, we consider the problem of capturing the daily experiences of each individual by using daily activity. The first step is to correlate diverse data streams with atomic-interval, and segment a person's day into her daily activities. We collect the diverse data streams from the person's smartphone to classify every atomic-interval into a daily activity. Next, we use an interval growing technique for determining daily-activity-intervals and their attributes. Then, these daily-activity-intervals are labeled as the daily activities by using Bagging Formal Concept Analysis (BFCA). Finally, we build a personal chronicle, which is a person's time-ordered list of daily activities. This personal chronicle can then be used to model the person using learning techniques applied to daily activities in the chronicle and relating them to biomedical or behavioral signals. We present the results for this daily activity segmentation and recognition by using lifelogs of 23 participants [83].

Event-Triggered Ecological Momentary Assessment: Chapter 4 represents an event-triggered method to complement the lack of important semantic information that has significant effect on health. A food journal is essential for improving health and well-being. However, journaling every meal is extremely difficult because it depends on user initiative and intervention. Current approaches to food journaling are both potentially inaccurate and tedious, causing people to abandon their journals very soon after they start. In this chapter, we propose a proactive and reactive mechanism that can significantly reduce user initiative while still remaining highly accurate. We first suggest a novel eating moment recognition technique using heart rate and activity patterns to trigger a food journaling process in a proactive manner. We then begin the food journaling process via voice command which utilized natural language processing when logging meals, which increases the ease of reactive self-reporting. Lastly, we enhance the food journal by automatically assessing ecological moments of eating through our Personicle system. We verified the method from a feasibility study conducted with three people for three months in their day-to-day lives. Our approach is designed to be unobtrusive and practical by leveraging multi-modal sensor data through the most common device combination of a smartphone and wearable device [87].

Enhancing Events of Daily Living: Chapter 5 presents how an activity of daily living can be unobtrusively enhanced by heterogeneous signals, and thus become an event of daily living. Events are fundamental for understanding how people experience their lives. It is challenging, however, to automatically record all events in daily life. An understanding of multimedia signals allows recognizing events of daily living and getting their attributes as automatically as possible. In this chapter, we consider the problem of enhancing a daily event by employing the commonly used multimedia data obtained from a smartphone and wearable device. We develop an unobtrusive approach to obtain latent semantic information from the data, and therefore an approach for enhancing a daily event based on semantic context enrichment. We represent the enhancement process through an event knowledge graph that semantically enriches a daily event from a low-level daily activity. To show a concrete example of this enrichment, we perform an experiment with eating, which may be one of the most complex events, by using 14 months of data for three users. In this process, to unobtrusively complement the lack of semantic information, we suggest a new food recognition/classification method that focuses only on a physical response to food consumption. Experimental results indicate that our approach is able to show automatic abstraction of life experience. These daily events can then be used to create a personal model that can capture how a person reacts to different stimuli under specific conditions [85].

Contribution

This thesis is the first of its kind where life experience is dealt with in terms of chronological events of daily living and relevant data streams. The Personicle is a powerful concept for capturing the unique stories of each individual since all kinds of personal data can be easily collected and are parallel with life events in one place. We believe this is an effective form to explore the correlation and causality between life events and all other data streams, and therefore extract the personalized knowledge that can help build the personal model. To the best of our knowledge, before Personicle, there have not been any implementation or experimental validation for this sort of research although there is a few of conceptual thinkings [51, 57]. Our main contributions to build the Personicle are as follows:

- **Defining an atomic interval as a base unit of daily life events.** We defined an atomic interval to break the silos of heterogeneous data scattered in different storage and thus bring the data into a daily life space. We proposed this idea by referring to the fact that daily life in a time-line is similar to an object in two-dimensional pixel space like objects. The daily life can be conceptualized as a correlation between times and pixels. It can facilitate synchronizing the low-level observations of daily life in an interval and then to handle a day as a collection of the sequence of these intervals.

- **Daily life segmentation.** We developed a method to segment a day into similar patterns of atomic intervals. This was based on our observations that indications of the changes in physical activity patterns can be involved in the changes of other attributes, which can be considered as ending one daily event and starting another. This technique can facilitate preprocessing one's daily life for further analysis, and thus can give an idea of what one's daily life look like.

- **Easy quantification of high-quality life experience data.** Unlike contemporary studies using impractical experimental settings that require users to wear customized on-body sensors, we attempted to develop a common daily event model through the data unobtrusively obtained from common IoT devices, such as a smartphone and wearable device. Moreover, our model reflects the constraints of the real-world, such as missing values and the limited number of samples, while maintaining reasonable performance. Thus, it allows us to easily quantify each individual's life experience, which includes data from low-level life logs to high-level life events.

- **Designing an event-triggered Ecological Momentary Assessment (EMA).** Although EMA was carried out in subjects' natural environments through a smartphone application, the prompts for the assessment questions had difficulty finding the right moment. Thus, it still relied on the subject's initiative and intervention, which can make users abandon the tracking even at the early stages. The event-triggered EMA contributes to overcoming the lack of intervention by initiating the EMA process from the system side at the right time. Personicle recognizes the EMA moment and triggers the questionnaires with basic contexts about the moment so that we can encourage the subjects to provide their important information.

- **Unobtrusively enriching events of daily living.** We suggested an unobtrusive approach to enhance the events of daily living based on semantic context enrichment. This can provide better abstractions to correlate the current states of a human being with daily experience, semantic context, and physiological signal. Such enriched daily events could play a very important role in building a model of the person reflecting the dynamics of his reactions under specific conditions.

- **Introducing a Personicle open source software.** We present an open-source software[1] and mobile application[2] co-developed by Clearsense[3] that can facilitate building the chronicle of daily life event by applying theoretical knowledge proved in this thesis. This platform would be used for bringing in data from all the wearable devices and data sources to enrich event characteristics and attributes in Personicle. With this release, we expect to create an interactive system in which we allow researchers and developers and even individual users to add their customized data streams to Personicle as well as to obtain enhanced events of daily living.

Thesis Outline

We first introduce the background of Personicle in Chapter 2. After that, we explain how Personicle can be implemented in Chapters 3, 4, 5 and 6. In Chapter 3, we discuss the theoretical foundation for building Personicle. In Chapter 4, we describe an Event-triggered EMA to complement the lack of missing information in Personicle. In Chapter 5, we present a fully-automated method to enhance events of daily living with an actual example, eating, and therefore show how event enrichment is may possible. In Chapter 6, we provide one more example of complementing the lack of the informational aspect for an event that can help enrich the events of daily living. Finally, Chapter 7 concludes the thesis and discusses future challenges.

[1] https://personicle.com/
[2] https://play.google.com/store/apps/details?id=personicle.os&hl=en_US
[3] https://clearsense.com/

Chapter 2

Preliminaries

In this preliminary chapter, we focus on introducing the Personicle in detail with the definition of key terminology. We also discuss the importance of chronicle of multimodal data, especially for personalized knowledge extraction, by providing an actual example of a Personicle user. In addition, we introduce a Minimum Viable Product (MVP) version of the Personicle system with its software architecture and data model. With these things in mind, we will go on to discuss how to implement the Personicle starting from Chapter 3.

2.1 Personicle

2.1.1 Overview

In many schools of thought, the most important object or entity is ourselves. People are interested in understanding the self to maximize their own satisfaction [51]. For this reason, continuously collecting one's personal data can play an important role in helping the person gain a better understanding of themselves based on his or her own contexts. However, the data alone represents very little. To extract meaningful information or knowledge from the data and thus evaluate the current state of an individual, all the data streams need to be correlated with one another in the context of its application. For example, let's suppose that there is an active food logger who has type 2 diabetes and therefore has been trying to keep a healthy, balanced diet. He attempts to achieve his daily goals in food consumption by carefully calculating and journaling the portions of vegetable, fruit, grain, protein, and dairy of every meal. Although his food journal can indicate whether or not he maintains his own intuition on a well-balanced diet, that alone is not good enough to understand the effects of certain foods or nutrients on his health, critical knowledge the diabetes patients and their doctors need to know. The objective when setting a diet is inevitably how it affects one's health, not the diet itself, which cannot be evaluated with a single data stream. Therefore, we need to collect and analyze personal data in a way such that we can find correlation and causality of different data streams to understand the personal state more effectively. To do this, we suggest a new type of personal data in which all kinds of lower-level personal data is aggregated, integrated and are parallel with one another in order to find their unique relationships.

2.1.2 Definition: Personicle, Event, Activity, and Life log

We believe that the sequence of daily life events an individual undergoes can contain one's life experiences, behavioral patterns, and even his/her emotions, and therefore contribute to a better abstraction of the person's current states. Thus, we define a Personicle as: "A personal chronicle of life events". To understand the concept of Personicle, it is necessary to define events, activities, and a life log in the context of daily life. Figure 2.1 shows a semantic hierarchy of these three types of data in our daily lives. First, a life log can be defined as a personal record of one's daily life in a varying amount of detail [42]. This log contains a comprehensive dataset of a human's activities, such as physical activity (e.g., walking,

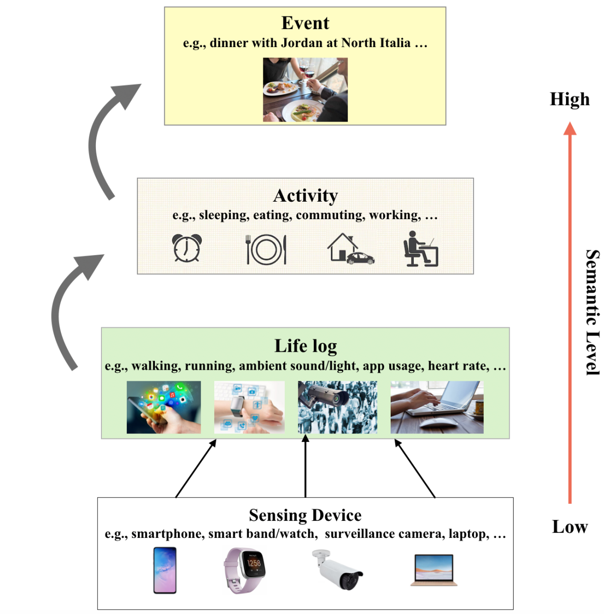


Figure 2.1: From life logs to an event

running), ambient sound/light, and app usage, and can be obtained through diverse sensing devices. By analyzing this life log, we can resolve the semantics of higher-level activities of daily living. We define these activities to be routine daily activities, such as eating, bathing, dressing, using the toilet, and transferring, that a normal person performs without external help [35, 80]. With this activity and life log, we can finally objectively understand one's events of daily living. We define an event as a common concept in human daily life that represents the aggregation of activities and other attributes (e.g., life log), such as dinner with Jordan at North Italia for celebrating his 33rd birthday.

2.1.3 Chronological Observation of Multimodal Data Streams

Sensors are now present in millions of personal devices, such as smartphones, wearable devices, and home appliances. These sensors are often continuously recording information about the world around it, particularly about a person. Thus, it is logical that the analysis of this data can reveal what an activity an individual is engaged in at nearly any given moment.

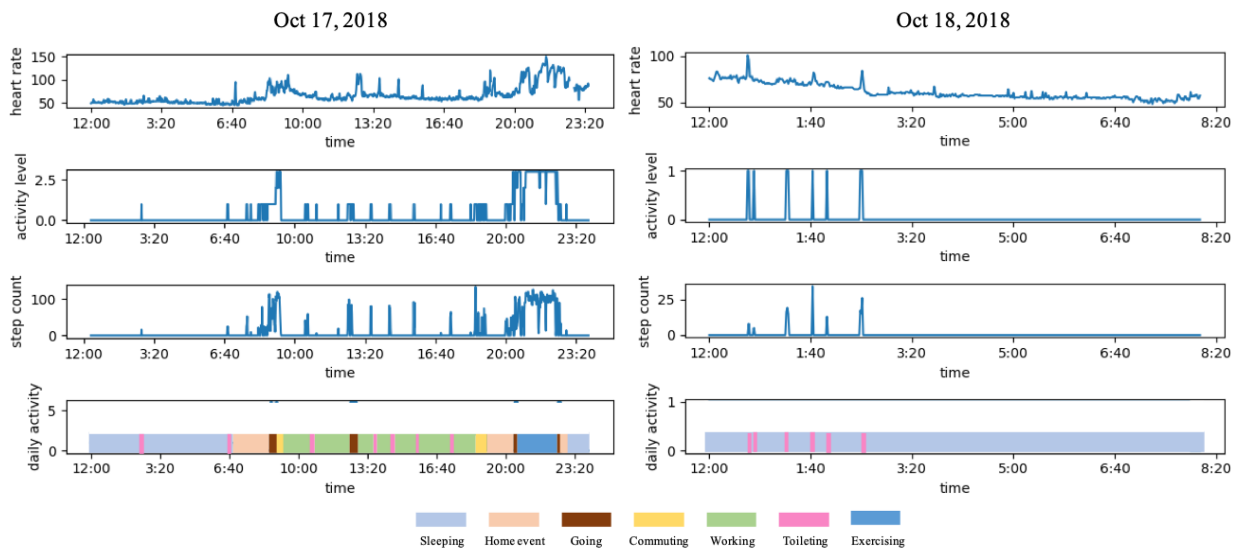


Figure 2.2: A simplified Personicle obtained from an actual user of the system.

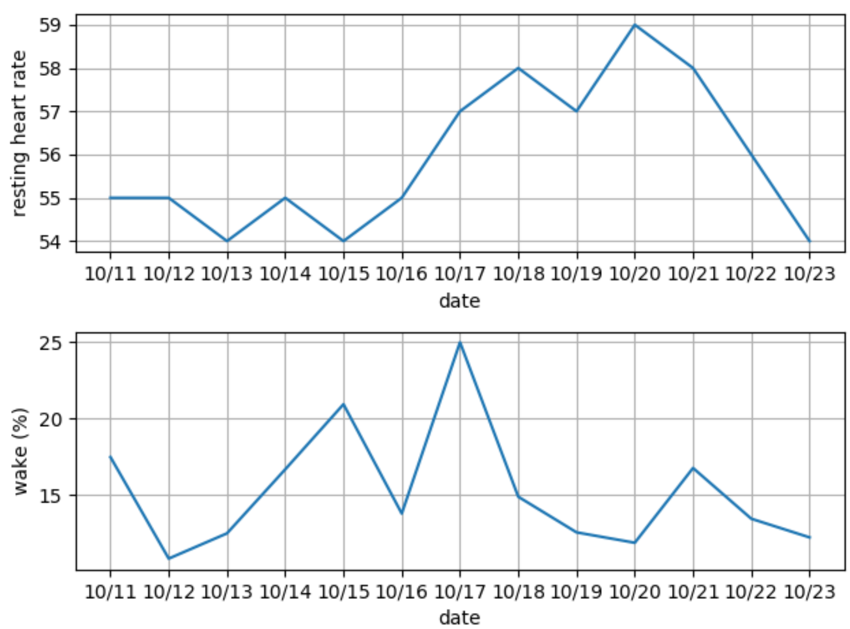


Figure 2.3: The actual user's resting heart rate and awake time during sleep on a daily basis.

We can also understand a person's responses to an activity. Physiological reactions, for example, can be evaluated by using relevant sensors such as a continuous glucose monitor (CGM), galvanic skin response sensor (GSR), or photoplethysmogram (PPG) sensor. The huge scale of measurements obtained from these sensors, GPS, social media and other related sources can be utilized to detect one or more of the predetermined classes of life events we defined earlier. We can then effectively compile this higher level data into a chronicle containing a sequence of events. The Personicle has definite advantages for personalized knowledge extraction. Compared with a single data value/stream, the Personicle utilizes a better, multimodal data structure to help decipher the correlation and causality of one's personal state. For instance, Figure 2.2 shows a simplified Personicle obtained from an actual user of our system. This Personicle can provide meaningful knowledge of how heavy exercise affects this user's resting heart rate and sleep efficiency. We can see from data on Oct 17, 2018, that this user went to the bathroom two times in the middle of the night, which is his average frequency for the last 6 months. We can also see that his heart rate fluctuated around 50 while he was sleeping during the night. If we compare these observations to those with data on Oct 18, 2018, we can notice that something came up that can lower sleep efficiency as well as increasing resting heart rate. On Oct 18, his heart rate during sleep fluctuated around 70, which is much higher than his normal heart rate range and even similar to when he was in working event. In addition, he went to the bathroom 6 times during the night, which is three times higher than usual. From these observations, we can notice that playing soccer event on the night of Oct 17 could be the reason causing his unusual sleep pattern. When we analyzed his last 6 months' worth of Personicle, we could find that he neither played intense sports nor exercised other than walking for a few minutes at a time. It means that this user does not get enough exercise for a long time. Thus, we can extract knowledge from this sequenced chronological observation that a sudden heavy exercise late at night would affect his sleep efficiency and resting heart rate. Furthermore, Figure 2.3 shows that this heavy exercise event affected the user's resting heart rate even for the next one week.

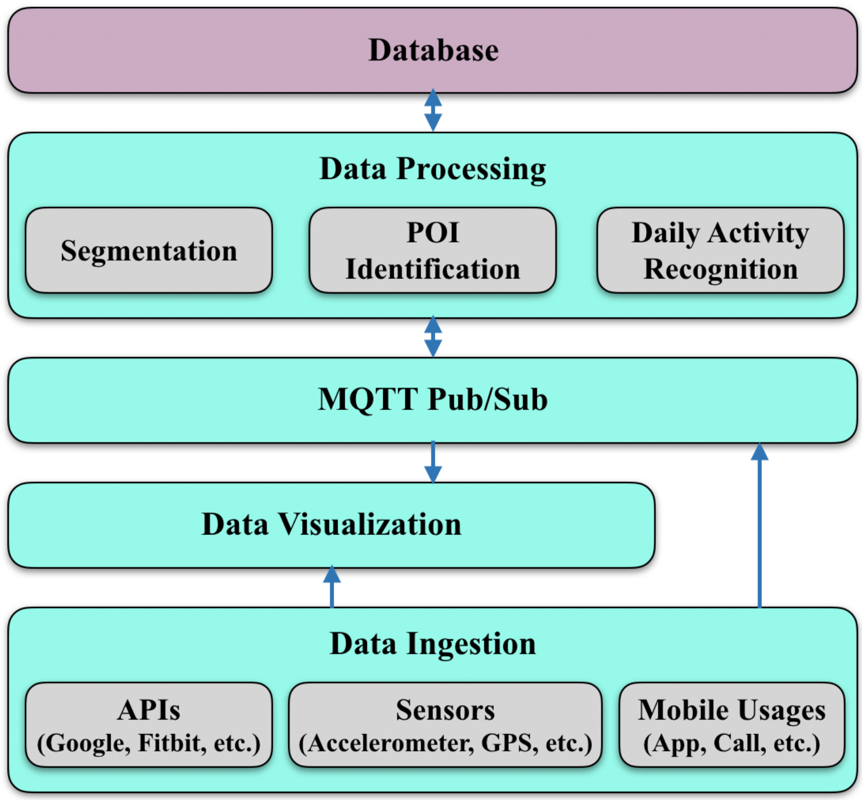


Figure 2.4: Personicle MVP System Architecture

His normal resting heart rate fluctuated around 55 before the night, but it was increased by 59 after playing soccer. Given that his average resting heart rate was 55 for the last 6 months, this rapid increase must be an important indicator that can tell about his current health state.

2.2 Personicle System

When the Personicle project was first initiated in 2013, we designed and implemented a research version of a data pipeline using an Android smartphone and Google Cloud Platform (GCP) to automate the process involved in extracting, transforming, combining, and loading life logs for further analysis. Recently, this research-oriented system has been entirely redesigned and re-implemented with a company called Clearsense for the purpose of creating an open-source software. With this effort, we released a MVP of Personicle platform in which we implemented the fundamental features of Personicle, such as data ingestion, data visualization, data processing, and data store. The following architecture shows the high-level design, which can be further broken down into more discrete modules in later versions.

As shown in Figure 2.4, the Personicle system consists of several layers. The bottom-most layer, which is the Data Ingestion layer, pulls multimodal data streams from different sources. In this layer, we not only collect low-level life logs from the built-in sensors in a smartphone, but also obtain higher-level data through APIs such as Google places[1] and Google play services[2]. We also provide an interface to connect with wearable devices, such as Fitbit, so that we can cover all other collectible data in one system. In a future release, we plan to make it available for other users to add their own data streams into the Data Ingestion layer, and thus make it an interactive system. In addition, we run several loggers in this layer to collect the user's mobile device usages in real-time. The data that we are now collecting is as follows:

- APIs: Google places (e.g., place type - cafe, restaurant, school, etc.), Google play services (e.g., latitude, longitude), Fitbit (e.g., activity level, heart rate, resting heart rate, step count, sleep - start time, end time, efficiency, length of deep sleep, rem sleep, light sleep and awake), Google Activity Recognition (e.g., in vehicle, on bicycle, running, still, walking)
- Sensors: ambient light (0 lm - 1000 lm), accelerometer, gyroscope, GPS, step
- Mobile Usages: application usages (e.g., app category, name, duration), media play (e.g., duration), wifi connection (e.g., ssid, duration), phone on/off count, calendar (e.g., event, start time, end time, location), call usage (e.g., duration, type - missed,

[1] https://developers.google.com/places/web-service/intro
[2] https://developers.google.com/android/guides/overview

Figure 2.5: Personicle MVP Data Visualization

incoming, outgoing)

Once data is collected in the Data Ingestion layer, it is delivered to the Data Processing layer through MQTT. The Data Processing layer is built in Nifi to effectively automate and manage the flow of multimodal data streams. In this layer, we identify Points of Interest (PoI) by analyzing the user's frequently visited locations, segment different data streams through the analysis of physical activity patterns, and build a machine learning model to recognize daily activities. The Data Processing layer also stores the ingested data, segmented results, and recognized daily activities to the database as well as sending the aggregated results back to the Data Visualization layer. The Data Visualization layer depicts the user's daily activities and relevant attributes as shown in Figure 2.5. In this layer, we can display a pie chart that summarizes the daily activities in any given day, where each daily activity is labeled by different colors. We also provide another graph that can facilitate the comparison of multiple signals on the same timeline, allowing users to find their correlations if they wish.

Figure 2.6 shows the data model designed for the Personicle MVP. We designed and released

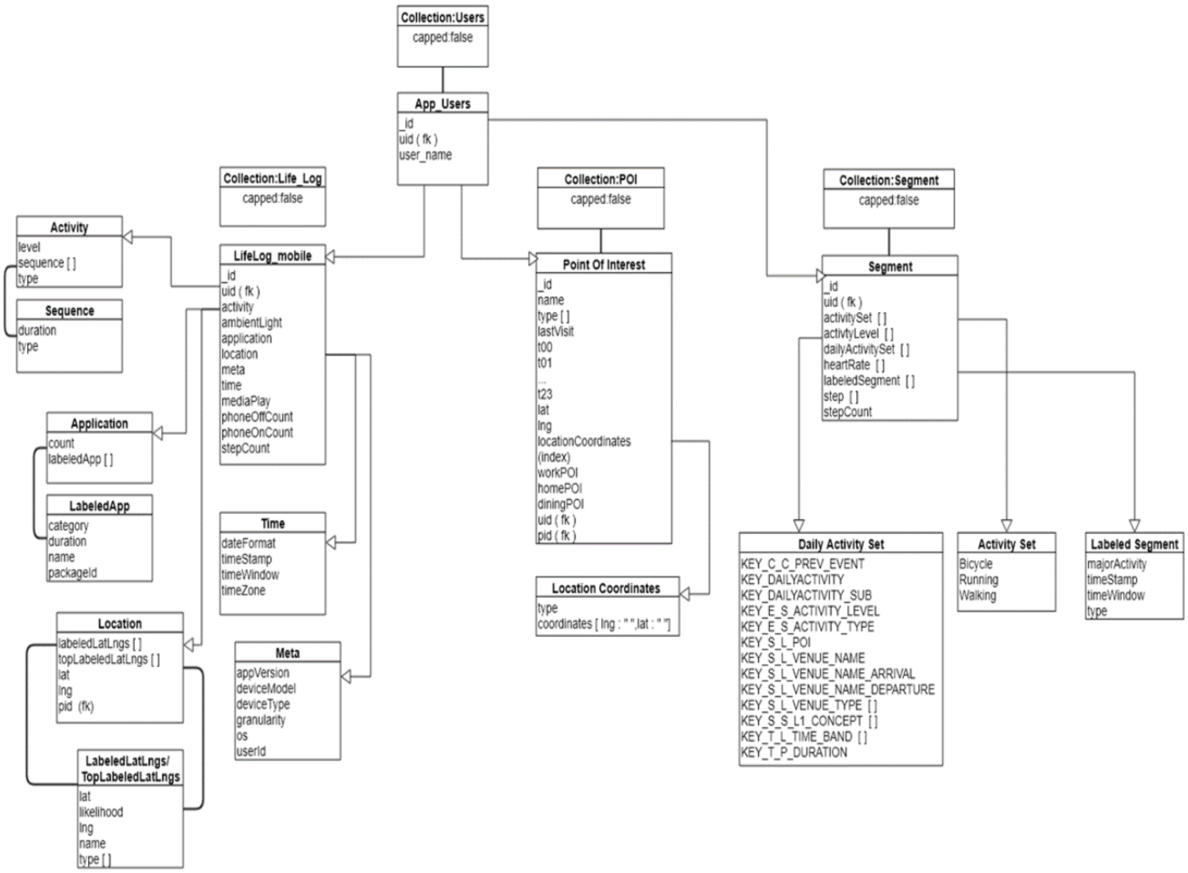


Figure 2.6: Personicle MVP Data Model

this model with the intent of organizing the various elements of currently collectible data and then show how they relate to one another, and thus encourage other researchers to participate in further discussion of a large, comprehensive multimedia database. Indeed, Peng Cheng Lab (PCL)[3] has started redesigning the data model by considering the correlations among events, measurements (e.g., data streams coming from different data sources), and attributes (e.g., semantics derived from measurement streams) for their Personicle research. We are also collaborating with AsterixDB group in University of California, Irvine to redesign our early data model with AsterixDB [8] to manage all types of big multimedia data more fast and effectively. As such, the efforts on implementing a real database are underway, but it is beyond the scope of this dissertation.

[3]http://www.szpclab.com/

Chapter 3

From Multimedia Logs to Activities of Daily Living

3.1 Introduction

Understanding the daily lives of human beings, what people have experienced, how people have spent their time, when and where they have been and whom they have been with, has long been the subject of scientific inquiry. This interest has led people in the field of multimedia to develop scientific approaches to monitoring and analyzing personal lifestyles and behavioral patterns. Multimedia researchers have tried to extract semantic level information from visual content so that they can analyze people's lives, and even environmental conditions and social situations. They also have analyzed real-time behavior data, which is collected via wearable devices, such as smartphones or smartbands, and social media, to understand more about personal lifestyles and behavioral patterns. However, recognizing people's daily lives at higher cognitive and more abstract levels (e.g., working, exercising, shopping, or relaxing) than low-level multimedia lifelogs (e.g., step count, GPS, venue, or physical activity), which makes inferring and predicting people's lifestyles more intuitive, remains relatively undeveloped.

Advances in sensor technology have increased the number of quantitative and qualitative multimedia lifelogs that are captured via wearable devices. Thus, we can now automatically aggregate and analyze heterogeneous multimedia data streams. Since these data streams have different granularity and semantics, the data streams need to be correlated by synchronizing them in the context of the application. The synchronized data streams can then be raised up to higher-level forms, so-called daily activity, by analyzing relationships between the daily activity and their temporal, causal, spatial, experiential, informational, and structural aspects [113]. Finally, the personal chronicle of the daily activity can be generated by chronologically ordering the recognized results [52]. In this thesis, we automatically recognize these daily activities using multimodal data streams from each individual's smartphone. Figure 3.1 shows the steps of our recognition approach: collecting multimedia lifelogs, synchronizing and segmenting the data streams, recognizing daily activities, and generating the personal chronicle.

We consider the problem of modeling an individual to ultimately help them with personalized health management. We believe that objectively understanding the daily activities of human beings has a strong potential to improve health research, given that these daily activities and the sequences of these high-level data abstractions contain their life experiences, behavioral patterns, and even their feelings. According to Kahneman et al., quantifying information about time usage and its frequency, as well as stress level, pleasure, and other affective states of each individual user, is potentially useful for health research [57]. More specifically, they tried to find this information by identifying each person's daily activity. Thus, the authors first conducted a survey categorizing people's common daily activities, and then described how to quantify them. Jain and Jalali's research on objective self models has also shown that analyzing the personal chronicle of daily activities can be used to build sophisticated

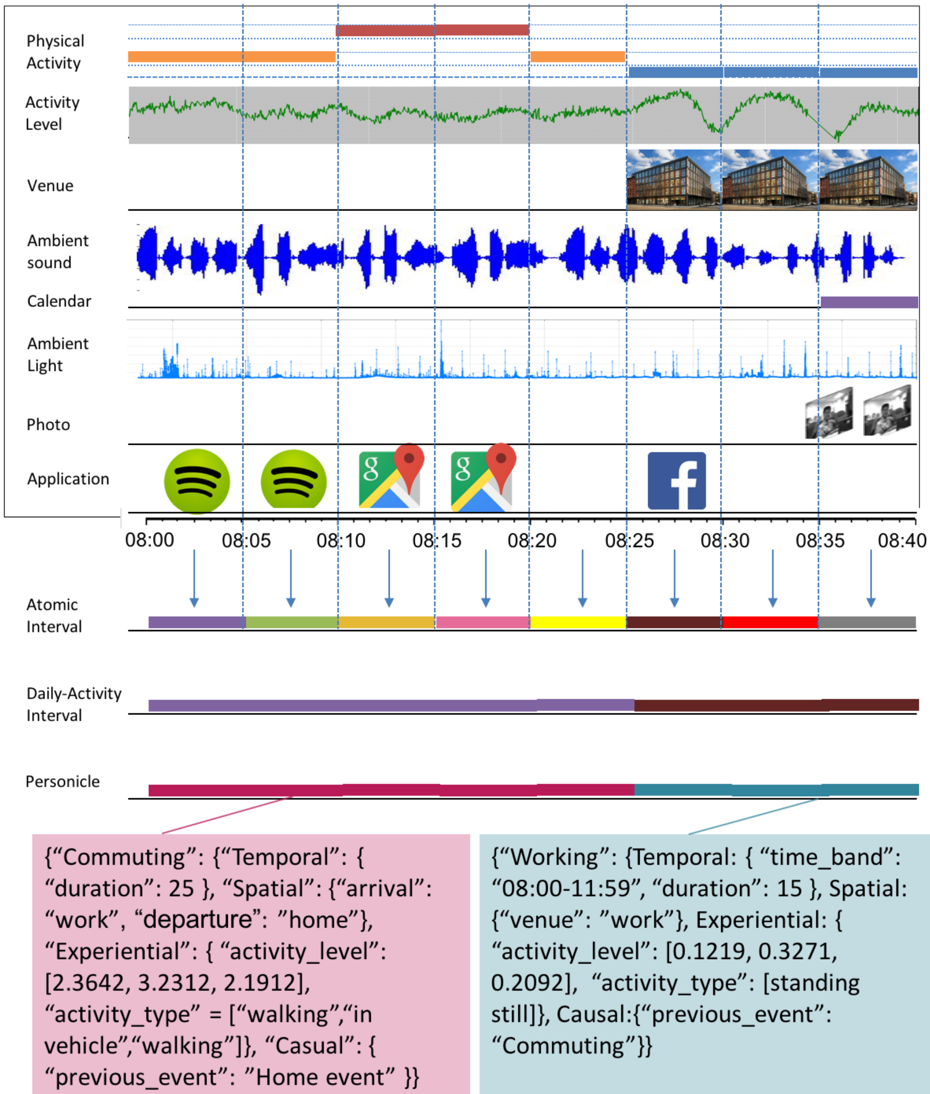


Figure 3.1: From multi-modal sensor data streams to atomic-interval, daily-activity-interval and chronicle of daily activities.

models that will help the monitoring of individual health and building disease models [51]. They built a complete infrastructure for the objective self, but it has not yet had actual implementations and experimental validations. Thus, with the same goals in mind, we recognize Kahneman's common daily activities and generate personal chronicles of the daily activities in order to build objective self models.

To automatically quantify the daily activity of each individual, the recognition method should be unobtrusive and effortless, and user tracking should only use common devices. More importantly, we should not intervene in users' life patterns by pushing them to do something or putting them in specific situations in order to recognize their daily activity. However, one major technical challenge is that this sort of fully-automated tracking is not always a guarantee of high recognition accuracy [24]. Some daily activities might require more diverse features than current smartphone sensors, and some others might be user-dependent or subjective daily activities, which need user feedback for personalization. This paper describes how to overcome these challenges for fully-automated tracking and explores to what extent Kahneman's daily activities can be recognized.

Our approach begins with understanding daily-activity-intervals by classifying every atomic-interval into a daily activity. We propose the idea that daily activities in a time-line are similar to objects in two-dimensional pixel space in that both the daily activities and objects are determined by a correlation between the times/pixels. We first collect multimedia logs via each individual's smartphone. Then, the collected logs are used to segment a person's day into their daily activities. We use diverse data streams from the person's smartphone to classify every atomic-interval into a daily activity. Next, we use interval growing techniques for determining daily-activity-intervals and their attributes. Then, these intervals are labeled as the daily activities by using Bagging Formal Concept Analysis (BFCA). Finally, we build a personal chronicle represented as events.

We believe that recognizing atomic-level daily activities, which can be automatically recognized, is one important step towards higher-level activity recognitions. Our main contribution in the area of activity recognition is 1) revealing and quantifying these atomic-level daily activities with our automated and unobtrusive approach, and 2) increasing the possibility of automatically recognizing the higher cognitive daily activities, and thus 3) quantifying the personal chronicle of these daily activities, as in Figure 3.1.

3.2 Related Work

Research on human behavior analysis is not a new area. It has been around for decades in many different forms. In 1945, Vannevar Bush's "Memex" vision had already presented a systematic approach, which organized a person's life-time knowledge, such as books, records and communication, by providing a user-authored data store, its linkages, and labels of the data to understand personal experiences [42, 19]. However, the capability to greatly develop this vision has recently become possible with advancements in technology [42]. The significant advances in computer storage, processing power, sensing technology, and network systems have encouraged researchers to participate in the field of human behavior recognition [51]. Classification techniques have also contributed to the recognition of higher-level semantics, such as physical activity [10, 116, 62, 46], more so sensor measurements.

There have been several data-driven studies analyzing the contexts or lifelogs of each individual user. A. K. Dey devised an architecture named Context Toolkit, which would allow the combination of data resulting in an abstraction that can be used to better understand how people experience the real-world [1, 28]. They provided higher-level contexts by aggregating and interpreting lower-level contexts in the conceptual framework. Since the Context Toolkit was introduced in 2001, the agent for human sensing has moved from the computer-based toolkit to mobile/wearable sensor-based loggers [14, 20, 36, 49]. With the trend of using the Internet-of-Things for data-driven studies, the so-called lifelogging, which is focused on a process of pervasively collecting, processing, and reflecting on an each individual's life experience data, has become more popular [42]. For example, Gordon Bell recorded many aspects of his everyday life by capturing a series of real-world images using the wearable camera, called SenseCam [40, 39], for the purpose of aiding recollection of past experiences. He expected that "total capture" of daily life would lead to "total recall" of our lives [12, 99]. P. Wang and A. F. Smeaton have also highlighted the importance of visual lifelogs because they identify various semantic concepts across individual subjects. They automatically identified high-level human activities such as eating, drinking, or cooking using SenseCam images, and data models [110, 111, 112].

Many human activity recognition systems have been based on situation specific capture. MIT's "PlaceLa" installed hundreds of sensors in all parts of a home seeking to automatically record activities [99]. Kasteren et al. collected location data and voice labeled annotations for each activity, such as breakfast, sleeping, or toileting, from the house. They constructed a probabilistic model using a hidden markov model (HMM) to predict future sensor readings [107]. Research on situation specific capture has drawn much attention in Activity of Daily Living (ADL) recognition. To automatically recognize ADL (e.g., toileting, grooming, bathing, showering or sleeping, etc.) for the purpose of preventative medical monitoring or building a smart home, researchers have set up low-cost sensors at critical locations in a home [104, 37, 107, 22, 79, 50] and then have predicted activities using naive bayes classifier [104, 69], HMM [107], ontologies and semantic reasoning [22], and Formal Concept Analysis [79], etc. Luštrek et al. used smartphone data, such as location (GPS), physical activity (accelerometer), and sound, and combined machine learning algorithms and symbolic reasoning to recognize high-level activities of a diabetic patient [69].

Another research group seeks to segment events on a lifelog of images. Doherty and Smeaton extract MPEG-7 features from images, such as accelerometer sensor values, light-level, ambient temperature, and passive infrared, and compare the similarity to those of adjacent images for the purpose of event segmentation [30]. There is another group who plans to recognize Kahneman's daily activities by analyzing taken photos from smartphones [7]. However, to the best of our knowledge, there is no approach for daily activity recognition that begins with understanding physical activity patterns by using non-visual smartphone lifelogs, and then gradually finding daily-activity-intervals in order to recognize daily activity. Moreover, we have not seen any approach to identify atomic-level daily activities to recognize higher cognitive and more abstract levels.

3.3 Methodology Overview

In this section, we describe our overall methodology for recognizing daily activity. We first explain what daily activity is, and then finalize the target corpus of the daily activity. Next, we categorize the daily activity corpus into three levels that describe their characteristics in terms of recognition possibility. Lastly, we provide the definitions of each daily activity. Since we ask our participants to label their daily activities with the exact name of that moment, we must synchronize the exact meaning of each daily activity. We refer to dictionaries, such as the Oxford and Cambridge English Dictionaries, and modify the meanings to match our contexts. We explained these definitions to each participant, and encouraged them to correctly label their daily activities according to the definitions.

We consider that daily activity is a brief name for each episode, such as "commuting to work" or "eating lunch", that can generally happen in the daily lives of human beings. Thus, we think that the continuous series of the daily activities can imply the person's lifestyle, behavioral patterns, and even their feelings. Kahneman et al. have also insisted that quantifying these daily activities would potentially be useful for research on human well-being. Furthermore, they have tried to categorize common daily activities by conducting a survey, and suggested 16 common daily activities. We refine our daily activities into Kahneman's

Table 3.1: Kahneman's daily activity on concept levels

Level 1	**Level 2**	**Level 3**
Still	Working	Watching TV
Walking	Commuting	Preparing food
Running	Exercising	Socializing
Cycling	Religious event	Housework
Driving	Shopping	Intimate relations
Direct communication	Eating	Relaxing
Remote communication	Using toilet	Taking a break
On the smartphone	Home event	Sleeping

daily activity corpus, which has already been verified for human well-being research [57].

We classify Kahneman's common daily activity in three levels. The level definitions are as follows:

- **Level 1 (L1):** a daily activity which can be automatically recognized. It can be seen as the atomic-level.
- **Level 2 (L2):** a daily activity which has the possibility of automatic recognition in the near future using sensing technology, but can not yet be recognized.
- **Level 3 (L3):** a daily activity which is not possible to be automatically recognized, but is soon to be recognized once richer data is gathered. We also deem subjective or user-dependent daily activities as level 3.

Since there are limits and restrictions on smartphone-based recognition, such as the lack of sensor data, or difficulties in understanding user-dependent or subjective activities, we think that it is not possible to recognize all the daily activities at the current stage. Our approach is to focus on recognizing daily activities, which can be automatically recognized via smartphone first (atomic-level), and then gradually try to recognize the daily activities which have a high possibility of automatic recognition (L2). Once the daily activity is recognized, we start considering that activity is the atomic-level activity, and using it as an attribute for other daily activity recognitions. Table 3.1 is the classification of Kahneman's daily activities in these three levels. In this thesis, we refine the target activities to L1 and L2 activities, and see to what extent L2 activities can be automatically recognized.

We define Kahneman's L2 activities based on their dictionary definition. People might have different definitions for each daily activity. Thus, we provide them with the following general definitions for correctly labeling their daily activities:

- **Working**: the activity of doing a job at the workplace (indoors)[1].
- **Commuting**: the activity of traveling regularly between work and home[1].
- **Exercising**: the activity of performing physical actions to make or keep your body healthy[1].
- **Religious event**: the activity occurring at religious places.
- **Shopping**: the activity of looking for things to buy in a shopping mall[1].
- **Eating**: the activity of taking food in a restaurant[2].
- **Using toilet**: the activity of going to the bathroom.
- **Home event**: the activity occurring in a structure in which a person lives, esp. a house or apartment[1].

3.4 Life logging

Lifelogging signifies the process of gathering, processing, and storing data regarding personal life experiences [42]. We collect, process, and record a user's contextual information while

[1]http://dictionary.cambridge.org/us/dictionary/english
[2]http://www.oed.com/

the user is carrying their smartphone. As shown in Figure 3.1, each exclusive data receiver, which is responsible for the generation of each data stream, pulls or processes the collectable data independently using built-in smartphone sensors and different APIs. The agent is always running in the background of each smartphone, logging the data without any user interventions, and storing the derived results locally on the device for user-studies. We collect the following lifelogs:

- **time:** time_window (e.g., 20161028_59), time_band (e.g., 0: 00:00 - 03:59, 1: 04:00 - 07:59, 2: 08:00 - 11:59, 3: 12:00 - 15:59, 4: 16:00 - 19:59, 5: 20:00 - 23:59), week (e.g., 0: week, 1: weekend), long_time (e.g., 1477655468)

- **location:** latitude[3], longitude[3], venue_name[3] (e.g., [Cheesecake Factory, Starbucks, Yogurt Land]), venue _type[3] (e.g., [restaurant, cafe, food]), venue_likelihood[3] (e.g., [30%, 10%, 5%]), point_of_interest

- **activity:** activity_type[3] (e.g., [still, walking]), duration[3] (e.g., [250, 50]), activity_level (e.g., 0.4012)

- **phone oriented lifelog:**

  1. **application:** count, name (e.g., [off, Facebook]), category[4] (e.g., [none, communication]), duration (e.g., [200, 100])
  2. **photo:** count, concept[5] (e.g., [person, pasta, dish, man, woman])
  3. **media:** play time
  4. **sound setting:** silence, bell, vibration
  5. **calendar:** event (e.g., birthday party), where (e.g., Cheesecake Factory), start_time, end_time

[3] https://developers.google.com/android/guides/overview
[4] https://play.google.com/store
[5] https://clarifai.com/

Table 3.2: Atomic-interval sample dataset. a1: still, a2: walking, a3: running, a4: bycle, a5: vehicle

atomic interval	**activity level**	**activity type**	**venue type**	**...**	**app type**
59	0	[a1]	building	...	-
60	1.15	[a1,a2,a1,a2]	route	...	fitness
61	1.99	[a3,a2,a1]	park	...	music
...	...	...	...	...	...
288	0	[a1]	building	...	music

We collect not only low-level lifelogs, such as latitude and longitude, but also high-level semantics. For example, we provide venue name set (e.g., [Cheesecake Factory, Starbucks, Yogurt Land]), which is the exact names of a given GPS point, and the categories of that venue (e.g., [restaurant, cafe, food]). Considering one GPS point may contain multiple venues, we also provide the probabilities of being at each venue (e.g., [30%, 10%, 5%]). In addition, we analyze the places the user frequently visited, and provide the user's point of interests. Furthermore, we accumulate a sequence of the user's physical activity, calculate activity level, which is an average score of the physical activity set [86], and then provide these as high-level lifelogs.

Since these lifelogs are collected as data streams, and they have different granularities and semantics, we must synchronize the data streams by correlating them with a periodic time-interval. We define this periodic time-interval as atomic-interval. Atomic-interval is a 1 x N matrix having N kind of lifelogs collected for a given time-interval. Each row in Table 3.2 shows the atomic-interval. The numbers in the first column indicate the order of the atomic-interval of the day. The sequentially collected lifelogs, such as activity type, are chronologically collected in an array. Average value, such as activity level, calculated based on pre-defined weights and their amount. Semantic data, such as activity type, venue, photo concept, or application category, are gathered by trustworthy APIs. The length of the atomic-interval can be decided by the designer depending on the precision requirement of the application, and thus there can exist the following separated atomic-intervals per day if we assume the unit of interval as minute.

$$number\_of\_atomic\_intervals = \frac{24hours \times 60minutes}{time\_interval} \tag{3.1}$$

We organize these atomic-intervals as json format in Figure 3.1, and then store them in the mobile phone database. We also define daily-activity-interval as a length of the daily activity. This daily-activity-interval can be determined by using our interval growing technique. This technique analyzes the characteristics of sequential atomic-intervals, and groups similar atomic-intervals together to make the daily-activity-interval. This is also shown in Figure 3.1.

3.5 Daily Activity Recognition

3.5.1 Daily Activity Segmentation

Daily activity segmentation is the process of partitioning a day into multiple sets of daily-activity-intervals. We pull diverse data streams from a user's smartphone, synchronize each data stream by using atomic-intervals, and then segment a day with our interval growing technique when chronological atomic-intervals have similar patterns of physical activity. For these reasons, determining a length of the atomic-interval must be the first step. We have proven that a five-minute time interval can be a reasonable amount for the atomic-interval. We have tried to find situation transition moments by comparing the similarities of sequential five-minute atomic-intervals, and then proved that this amount of time can be a base unit of daily activity segmentation [86, 53]. Thus, we use five-minutes as the length of the atomic-interval, and then divide a day into 288 atomic-intervals. Most importantly, we assume that indications of the changes of physical activity pattern can be involved in the changes of other attributes, which can be considered as ending one daily activity and starting another. For example, let's say a user has been working at the office, and he has been sitting on the chair. After 10 minutes, if he moves towards the cafeteria for lunch, we recognize this change of physical activity, segment this moment, and make a daily-activity-interval by segmenting from the first atomic-interval to now. In other words, our daily activity segmentation focuses on the interval-growing technique appropriate for daily activity segmentation in which the relevant atomic-intervals are identified by the patterns of physical activities.

Binary Interval Growing (BIG): More specifically, we apply our binary interval growing technique to determine whether consecutive atomic-intervals have similar patterns of physical activities. In order to compare the similarities, we classify each atomic-interval into the moving or the non-moving type of interval, and then deal with the atomic-intervals as one or the other. Algorithm 1 shows the procedures of how to segment atomic-intervals into daily-activity-intervals. We first set up a seed atomic-interval S_j, and then keep calculating $\delta(i)$ every five minutes to determine the similarity between sequential atomic-intervals. $\delta(i)$ can be represented by the following formula:

$$\delta(i) = \|f(S'_j) - f(I'_i)\|_2^2 \tag{3.2}$$

where S'_j is $\{l_j, t_j\}$, I'_i is $\{l_i, t_i\}$, $f(x)$ is a classification algorithm to classify the non-moving (0) or the moving (1) type of atomic-interval, and $\delta(i)$ is a distance between S_j and I_i. Thus, we segment atomic-intervals when $\delta(i)$ is equal to 1, and then make a daily-activity-interval by segmenting from I_j to I_i. For example, if the type of the seed atomic interval is non-moving, then $f(S'_j)$ is equal to 0. After 5 minutes, if the type of the current atomic-interval is also non-moving, $f(I'_i)$ will be 0, and thus $\delta(i)$ is also equal to 0. However, after another 5 minutes, if the type of the current atomic-interval is moving, $f(I'_i)$ will be 1, and we will finally get $\delta(i) = 1$. Then, we segment this moment, make a daily-activity-interval by

---

Algorithm 1 Solution for BIG

Input: current atomic-interval $\mathbf{I}_i$, seed atomic-interval $\mathbf{S}_j$
Output: daily-activity-interval set $\mathbf{R}$;
1: Set $\mathbf{S}_j = \mathbf{I}_i$ if $i = 0$ and $j = 0$, or $\mathbf{S}_j = \varnothing$, and then set $k = 0$;
2: **repeat**
3: Wait for next atomic-interval, $\mathbf{I}_i = \mathbf{I}_{i+1}$;
4: Extract activity level $\mathbf{l}_i$, and total amount of moving time $\mathbf{t}_i$ from $\mathbf{I}_i$;
5: Extract activity level $\mathbf{l}_j$, and total amount of moving time $\mathbf{t}_j$ from $\mathbf{S}_j$;
6: Calculate $\delta(i)$;
7: Make a daily-activity-interval $\mathbf{r}_k$ by segmenting from $\mathbf{I}_j$ to $\mathbf{I}_i$, increment k and j, set new seed atomic-interval $\mathbf{S}_j = \mathbf{I}_i$ if $\delta(i) = 1$;
8: **until** the system is terminated.
9: **return R**

---

segmenting from I_j to I_i, and repeat this process again.

3.5.2 Daily Activity Recognition

To recognize the daily activities, we now build a common daily activity model. Westermann et al. have built a common multimedia event model by identifying the global unique properties of each individual event. This model addresses several fundamental aspects of events, such as temporal, spatial, experiential, causal, structural, and informational aspects [113]. Specifically, Westermann et al. approach the common event modeling by understanding physical (e.g, event occurrence time stamp and interval), logical (e.g, temporal domain), and relative (e.g, temporal relationships to other events) relationships between each aspect and an event. We bring in these general aspects as the categories of our modeling attributes, and modify the physical, logical, and relative components to match the daily activities.

We build the common daily activity model by using Formal Concept Analysis (FCA) based on these general aspects of events. FCA is one powerful technique when data sources are limited, and even when they are uncertain, due to its specialty for discovering implicit information

Table 3.3: Simplified cross table defining relationships between daily activity and their attributes.

		Attribute		
		Walking (Experiential)	Medium time-duration (Temporal)	Work (Spatial)
Object	Working		X	X
	Using Toilet	X		X
	Commuting	X	X	

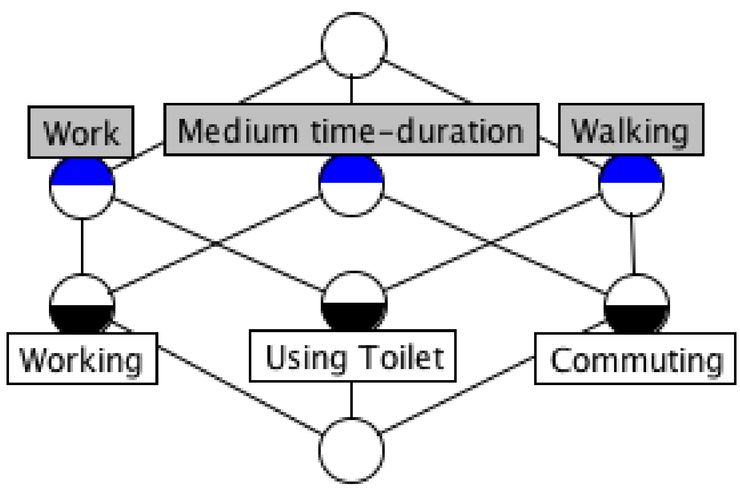


Figure 3.2: Sample concept lattice derived from Table 3.

based on pre-defined binary relationships between object and attributes. FCA can be applied for daily activity recognition as follows. One daily activity D can be represented by a triplet $T = (D, A, R)$, where A is a set of attributes, and R is the binary relationships between D and A, $R \subseteq D \times A$. Once each daily activity is defined by the triplet, the triplet can then be converted into a cross table (e.g., Table 3.3). Then, all possible formal concepts (X_i, Y_i), where $X_i \subseteq D_i$, and $Y_i \subseteq A_i$, are extracted from the cross table, and then are set up as nodes in the concept lattice, which is a graphical representation of the partially ordered knowledge. The hierarchy of formal concepts can be constructed by the following relations:

$$(X_1, Y_1) \leq (X_2, Y_2),\ if\ X_1 \subseteq\ X_2 \leftrightarrow Y_1 \supseteq Y_2 \tag{3.3}$$

X_i and Y_i satisfy the following relations:

$$X_i^{'} = \{a_i \in A_i \mid \forall d_i \in X_i, (d_i, a_i) \in R_i\} \tag{3.4}$$

$$Y_i^{'} = \{d_i \in D_i \mid \forall a_i \in Y_i, (d_i, a_i) \in R_i\} \tag{3.5}$$

Table 3.3 shows the simplified relationships between daily activity and their attributes. In order to build the FCA model, formal concepts should be derived from the cross table, such as (*Working*, {*Medium time-duration, Work*}), (*Using Toilet*, {*Walking, Work*}), (*Commuting*, {*Walking, Medium time-duration*}), ({*Working*, *Using Toilet*}, *Work*), ({*Working, Commuting*}, *Medium time-duration*), and ({*Using Toilet, Commuting*}, *Walking*). These formal concept pairs become each node in the concept lattice, and their hierarchy is determined by formula (3). Figure 3.2 shows the concept lattice reflects the partially ordered knowledge between each node. The top node and the bottom node indicate ({*Working, Us-*

ing Toilet, Commuting}, $\emptyset$), and ($\emptyset$, {*Walking, Medium time-duration, Work*}), respectively. To navigate the concept lattice to obtain the expected results, depth first search is carried out with input attributes. For example, if input attributes are *Medium time-duration* and *Work* in Figure 3.2, these two nodes will indicate one daily activity, *Working*.

Basically, FCA finds an expected result depending on the structural similarity between an input attribute set and pre-defined attribute sets. Thus, different kinds of input attributes can significantly affect the structural similarities. Because of this, it is necessary to estimate what attributes are important keys to separating each different daily activity, and find all unique daily activity structures composed of those attributes. Moreover, we also need an effective method for estimating missing data while maintaining accuracy, considering that we recognize the daily activities in real-time on smartphone, and the smartphone status will not always be in the best condition. Lastly, given that the amount of actual user data is not always enough to train a powerful model, we also need to come up with how we can make a strong learner by using a group of weak learners. We believe that an ensemble classifier that consists of many concept lattice bags, and its voting process to obtain a majority result from all the recognitions, helps to overcome these challenges. We suggest Bagging Formal Concept Analysis (BFCA), which applies the ensemble approach to FCA, in order to solve those challenges. Bagging Formal Concept Analysis (BFCA) consists of the following steps:

1. Categorize all the labeled daily-activity-intervals, which obtained from 23 participants for two weeks, by each daily activity.

2. Make n number of classifiers where n is the number of the recognizable daily activity, make m number of bags per classifier, and bootstrap training data for each bag.

3. In each bag, use one third random attributes $\frac{p}{3}$, where p is the number of total attributes, and extract all unique relationships between the labeled daily activity and their randomly picked attributes.

4. Build the cross table in each bag by using those unique relationships, and generate the concept lattice. This concept lattice only determines whether the given input attribute set can be the labeled daily activity.

5. When an input attribute set is given, which is an unlabeled daily-activity-interval, we navigate all the concept lattices for each daily activity classifier, calculate the possibility of being each daily activity, and then choose the highest possibility among the results.

Given that FCA requires discrete attributes, we convert our time-series values C, such as activity level, or time duration of daily-activity-intervals, into discrete space, such as w-dimensional space $\{high, medium, low\}$, by a vector $\bar{C} = \bar{c_1}, \bar{c_2}, ..., \bar{c_i}$. We use a discretization technique, SAX (Symbolic Aggregate ApproXimation), which reduces the time series of arbitrary length n into the w-dimensional space by the following equation [68]:

$$\bar{c_i} = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} c_j \tag{3.6}$$

3.5.3 From Activities to Events

We proceed to create events with all facets by using all collectible data sources from multiple devices. We insist that an event is just a single unit in itself, but it can form the chronicle once it is stored in the database. Thus, we store all the recognized daily activities in the database as events with as many data sources as possible, such as *Personicle* in Figure 3.1, and quantify the chronicle. This personal chronicle can then be used to model the person by using learning techniques and relating them to biomedical or behavioral signals. In the current version, we use only the signals from smartphones, but pulling heterogeneous signals from multiple devices, and then analyzing a person with all the facets of the events will be

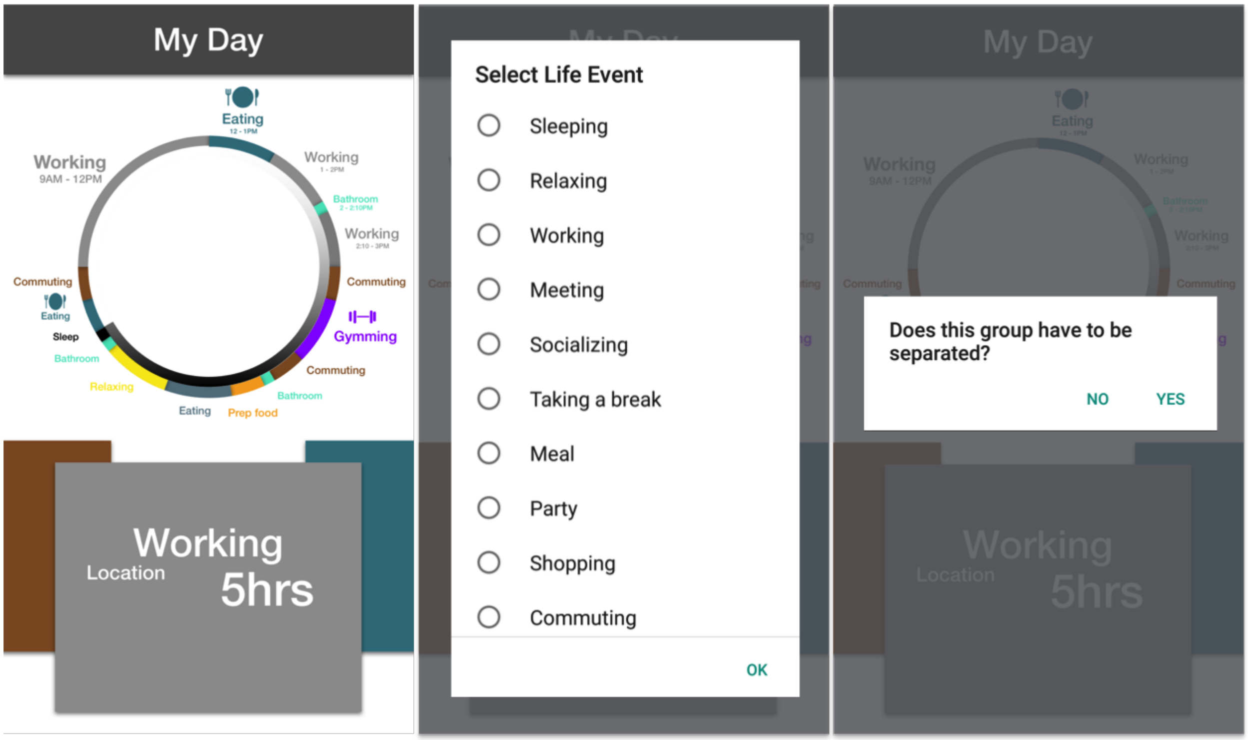


Figure 3.3: The system running for daily activity segmentation and recognition.

an important topic for further research.

3.6 Experimental Validation

We implemented an android application to test our segmentation and recognition methods. As shown in Figure 3.3, we asked 23 participants to give feedback on the results of their segmentations as well as label their daily activities for each segmented result during an average of two weeks. We stored all the collected lifelogs in their smartphone database, and then gathered these databases after the experiment had been completed. The total number of collected daily-activity-intervals was 35,967.

Table 3.4: Overall segmentation results of 23 participants.

Algorithm	Segmentation accuracy			
	Best J_c	Worst J_c	Average J_c	Stdev
BIG	**0.9583**	**0.7896**	**0.9050**	**0.0432**
Clustering	0.7841	0.5803	0.6564	0.0601
Thresholding	0.8556	0.4370	0.5863	0.1467

3.6.1 Segmenting User's Day

We assume that segmentation moments can mostly be affected by their adjacent atomic-intervals since atomic-intervals are on a one-dimensional time-line. Thus, interval growing based approaches, which compare contiguous atomic-intervals, must show better performance for the daily activity segmentation than those of statistical methods using all collected lifelogs. To verify the performance of BIG, we compare the BIG results to 1) ground truth, which was obtained by participants' feedback, and 2) the results to those of statistical techniques, such as clustering (k-means), and thresholding (otsu). We use the jaccard coefficient, which has the obvious advantage of similarity evaluation between two sets of binary data, for verifying the performance. The jaccard coefficient is calculated as follows:

$$J_c(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{3.7}$$

where A is the ground truth and B is the algorithmic result. We see if BIG can be uniformly applied in all the users by achieving relatively higher results than those of other algorithms. For this reason, we handle each user's experimental data separately, calculate each user's jaccard coefficient, and then compare the best, worst, and average results as in Table 3.4.

We prove the BIG's performance by comparing the results to the statistical approaches. From the results in Table 3.4, we can see that the average accuracy of BIG is higher than those of the others. Even the worst result of BIG is almost similar to or a little less than other techniques' best results. Furthermore, the standard deviation of BIG shows that each user's accuracy is nearly the same; however, we can also see a 0.1687 difference between the best and worst accuracy. There are two reasons. Given that we depend on the API results for physical activity prediction (e.g., still, walking), some incorrect API results may lead to incorrect segmentation results. More specifically, the API returns "walking" or "in vehicle" activity when a user slightly shakes his legs or has minute-long movements. We found that the user who obtained the worst result in BIG had many of these cases, and thus these unexpected cases resulted in the incorrect segmentations results. The different awareness of segmentation moments between users and us also caused the incorrect results. We handle 5 minutes time-intervals, and thus we don't consider the short changes as segmentation moments. For example, if a user walks only for 5 or 10 seconds, and then immediately starts a non-moving activity, we consider this as one continuous non-moving segment given 5 minutes length of granularity. However, some of the users who participated in our experiment gave feedback many of these moments were segmentable moments.

The lower result of clustering and the thresholding technique show that reflecting past physical activity patterns for current segmentation moments can cause a bed effect on segmenting results. For example, if a user is a very active person, those techniques will not segment small movements even though these are a sufficient amount for the daily activity segmentation.

3.6.2 Recognizing Daily Activity

With the identified daily-activity-intervals, we now try to recognize L2 daily activities. 23 participants had labeled L2 daily activities on these daily-activity-intervals.

In the lifelogs, we observed that at times our system was unexpectedly killed by OS, which made the data discontinuous. At other instances, participants did not label their segmented results, or the participants' phones ran out of battery. We tried to avoid these exceptional cases by immediately restarting the system when it was terminated by OS, or asking the user to label the segments with the pop-up messages in the system. However, there were still many non-labeled and non-consecutive segments. We first clean these unclear data in order to precisely verify the performance of recognitions, and thus the total number of considered daily-activity-intervals are 15,087 samples of 35,967. And then we split these samples into 30% training dataset, and 70% test dataset to show that the model of BFCA can be robust despite the relatively small training dataset.

In order to maximize the recognition performance, we assume that each daily activity has a specific combination of the common event attribute sets that most represent the daily activity. This means that all the aspects of the common event model (e.g., temporal, spatial, experiential, structural, informational, and causal aspects) are not vital elements for every daily activity recognition. For example, according to the definitions in Section 3, the "Commuting" activity, which refers to the activity of traveling regularly between work and home, can be recognized by only using spatial (e.g., work or home), structural (e.g., L1 daily activity, such as going, or still), and causal (e.g., the relations between current and previous daily activity) aspects. To verify this idea, we experimented with the different combinations of the common event model aspects, and figured out the best combinations by calculating their accuracy. We roughly use 10 bags of concept lattice for this experiment, and thereby calculate their f-measures to see the weighted harmonic accuracy between precision and recall. From the results in Table 3.5, we can see that some combinations of the attributes have better results than those of others, such as S_5 for D_1, D_6, D_7 and D_8, and S_6 for D_2. It shows that unnecessary information results in the confusion of modeling, and thus we use the specialized combination sets for each daily activity modeling.

We now try to find the best number of concept lattice bags, which also can maximize the recognition performance. First, we train separate BFCA models on different numbers of bags, which are from 1 to 1000, by using the selected attribute sets. Then, we experiment

Table 3.5: F-measure (%) for combination of attribute sets. D_1: Commuting, D_2: Eating, D_3: Exercising, D_4: HomeEvent, D_5: ReligiousEvent, D_6: Shopping, D_7: UsingToilet, and D_8:Working. S_1: Temporal + Experiential, S_2: Temporal + Spatial, S_3: Spatial + Experiential, S_4: S_1 + Spatial, S_5: S_4 + Causal, S_6: S_5 + Structural aspect.

			Attribute set combination					
		# sample	S_1	S_2	S_3	S_4	S_5	S_6
Daily Activity	D_1	393	66.7	66.7	55.5	75.6	**90.4**	76.6
	D_2	404	28.2	71.9	43.2	70.7	77.8	**79.6**
	D_3	15	0	**100**	**100**	**100**	**100**	**100**
	D_4	10698	60.6	94.7	65.6	91.8	**96.6**	**96.6**
	D_5	588	0	**98.5**	**98.5**	97	76.4	**98.5**
	D_6	53	0	40	22.2	25	**44.4**	40
	D_7	28	56.3	0	38.5	9.5	**81.2**	55.2
	D_8	2908	6.9	69.5	44.9	81.8	**90.3**	89.1

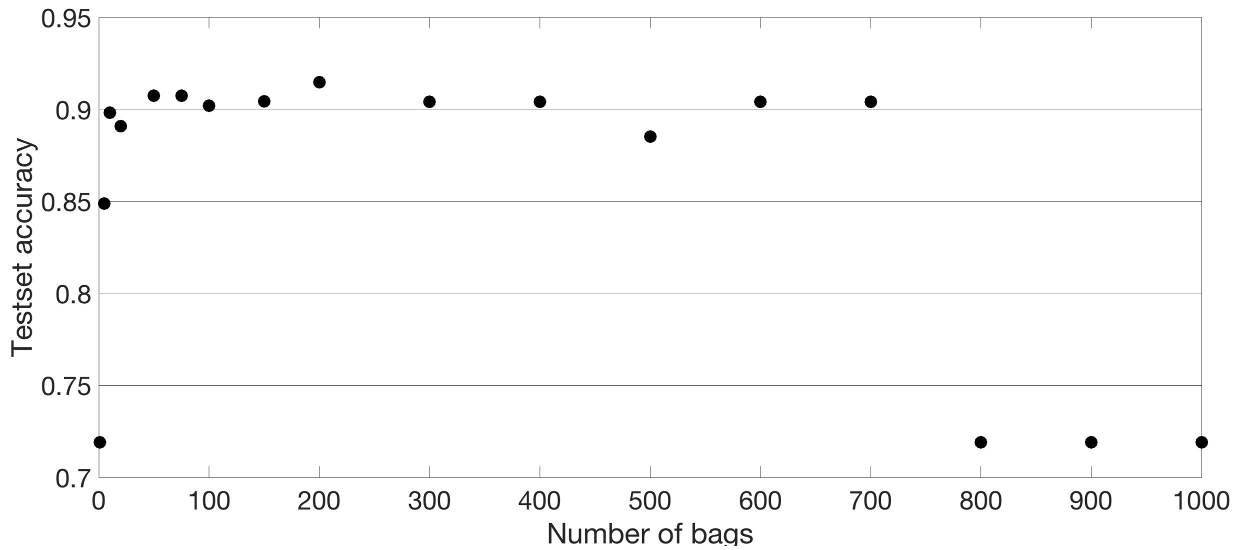


Figure 3.4: The Variations of BFCA accuracy on different number of concept lattice bags.

Table 3.6: Confusion matrix of the BFCA. D_1: Commuting, D_2: Eating, D_3: Exercising, D_4: Home Event, D_5: Religious Event, D_6: Shopping, D_7: Using Toilet, and D_8: Working.

		Predicted (%)							
		D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8
Targeted (%)	D_1	**95.8**	0	0	4.2	0	0	0	0
	D_2	0	**97.8**	0	0	0	2.2	0	0
	D_3	0	0	**100**	0	0	0	0	0
	D_4	0	4.3	0	**95.7**	0	0	0	0
	D_5	0	2.9	0	0	**97.1**	0	0	0
	D_6	0	16.7	0	0	0	**66.7**	16.7	0
	D_7	5.3	0	0	0	0	0	**94.7**	0
	D_8	5.6	9.3	0	0	0	0	0	**85**

Table 3.7: Accuracy for the daily activity recognition on 23 participants.

	Precision	Recall	F-measure
FCA	0.2522	0.5114	0.3378
BFCA	**0.9098**	**0.9204**	**0.9151**
Decision Tree	0.6604	0.6826	0.6643
Random Forest	0.7358	0.7464	0.7411
Support Vector Machine	0.6981	0.7081	0.7031

with the daily activity recognition on those trained models, respectively, by using the same test dataset. Finally, we calculate their f-measures to see what numbers of bags would return the best recognition accuracy. Figure 3.4 shows the variations of accuracy on the different number of bags. In our results, we can see that the accuracy for under 700 bags is nearly the same; however, the accuracy rapidly decreased by 0.7191 once bags are over 800. Basically, the higher the number of bags, the higher the recognition performance in ensemble technique. However, a large number of bags in BFCA can confuse the voting process given that these bags can make all the classifiers robust. Therefore, among the good results under 800 bags, we choose the best accuracy, 0.9147 (bags=200).

Then, we build the confusion matrix to see the specific results of each daily activity recognition. In Table 3.6, we can see that 5 minutes length of granularity results in an ambiguous segmentation boundary between "Commuting" activity and "Home Event" activity (4.2%). We also can see that randomly picked $\frac{p}{3}$ attributes cause confusion in the daily activity modeling. For example, "Home Event" activity can be considered as "Eating" activity (4.3%), and "Shopping" activity can be classified either as "Eating" (16.7%) or "Using Toilet" activity (16.7%), if spatial aspects are missed. However, the overall accuracy of all the daily activity recognition (>90%) shows that using the randomly picked attributes, and a certain number of concept lattice bags can minimize the misclassification of daily activities. This is proven in Table 3.7.

As shown in Table 3.7, BFCA has greatly improved the recognition performance compared to the FCA. FCA only depends on the structural similarity between an input attribute set and pre-defined relations. Thus, it sometimes recognizes multiple daily activities if they have similar structures to the pre-defined relations. This issue is a critical problem, which can cause lower performance, given that FCA does not have any statistical methods to choose the most probable result. The result of BFCA shows that applying a statistical method to FCA, such as the ensemble approach, can be one solution to overcome the problem.

Since BFCA brings the idea from random forest, which uses the ensemble technique bagged by decision trees, we also compare BFCA to random forest. In Table 3.7, we can see that BFCA has better results than the random forest. Basically, our dataset is imbalanced data because some daily activities occupy the better part of the day. For example, the "Sleeping", "Home Event", and "Working" activities used to be the majority of the daily activities. Moreover, these daily activities mostly share similar lifelogs to each other, and thus the decision tree and random forest must have difficulty clearly classifying them. This can also explain why the support vector machine, which is one of the most powerful classification algorithms, has lower recognition accuracy than BFCA.

3.7 Conclusion

Kahneman, who is a Nobel Prize winner, showed the importance of daily activities in human life experiences. This paper builds towards the research to develop techniques for objectively and automatically understanding the daily lives of human beings via common wearable devices. Specifically, this paper focuses on recognizing human daily activity to understand their lifestyle and behavior patterns for the purpose of building objective self model. Thus, it describes the methodology behind automatically recognizing daily activity with the goal to build a personal chronicle. We develop a logging application that runs on Android device, collects data, and converts the data into personal chronicle. Using the chronicle, one may proceed to determine individual models using machine learning techniques. Such models may play very important role in applications for health and behavior modification. We begin with synchronizing multimodal data streams by using atomic-intervals, and then use an interval growing technique for determining daily-activity-intervals and their attributes. Next, we use the common event model and BFCA to classify each daily activity. Lastly, the daily activities are stored in the database and consist of the chronicle of daily activities. Results obtained across different FCA and classification algorithms show the potential of such an approach for recognizing daily activities. Further research would allow for increasing the number of detectable atomic-level daily activities by combining more heterogeneous and higher cognitive multimedia logs, and thus recognizing more various daily lives.

Chapter 4

Multimodal Food Journaling

4.1 Introduction

You are what you eat.

The foods and drinks we put in our bodies have a direct impact on our health and well-being. There have been numerous medical studies showing that unhealthy dietary habits can be a major cause of diseases such as obesity, kidney disorder, CVD[1], cancer, and diabetes [60, 105]. Clearly a well-balanced diet is very important to stay healthy.

Food journaling has been demonstrated to encourage people to develop healthier dietary habits since it provokes self-reflection that can play a significant role in behavior change. Therefore, health care professionals and people who suffer from health-related disorders have tried to maintain a food journal so that they can analyze the health effects of their dietary intake [115]. However, even though food journaling has been the main method of monitoring dietary intake for a long time, unobtrusive ways of keeping a food journal remain relatively undeveloped.

[1] Cardiovascular disease

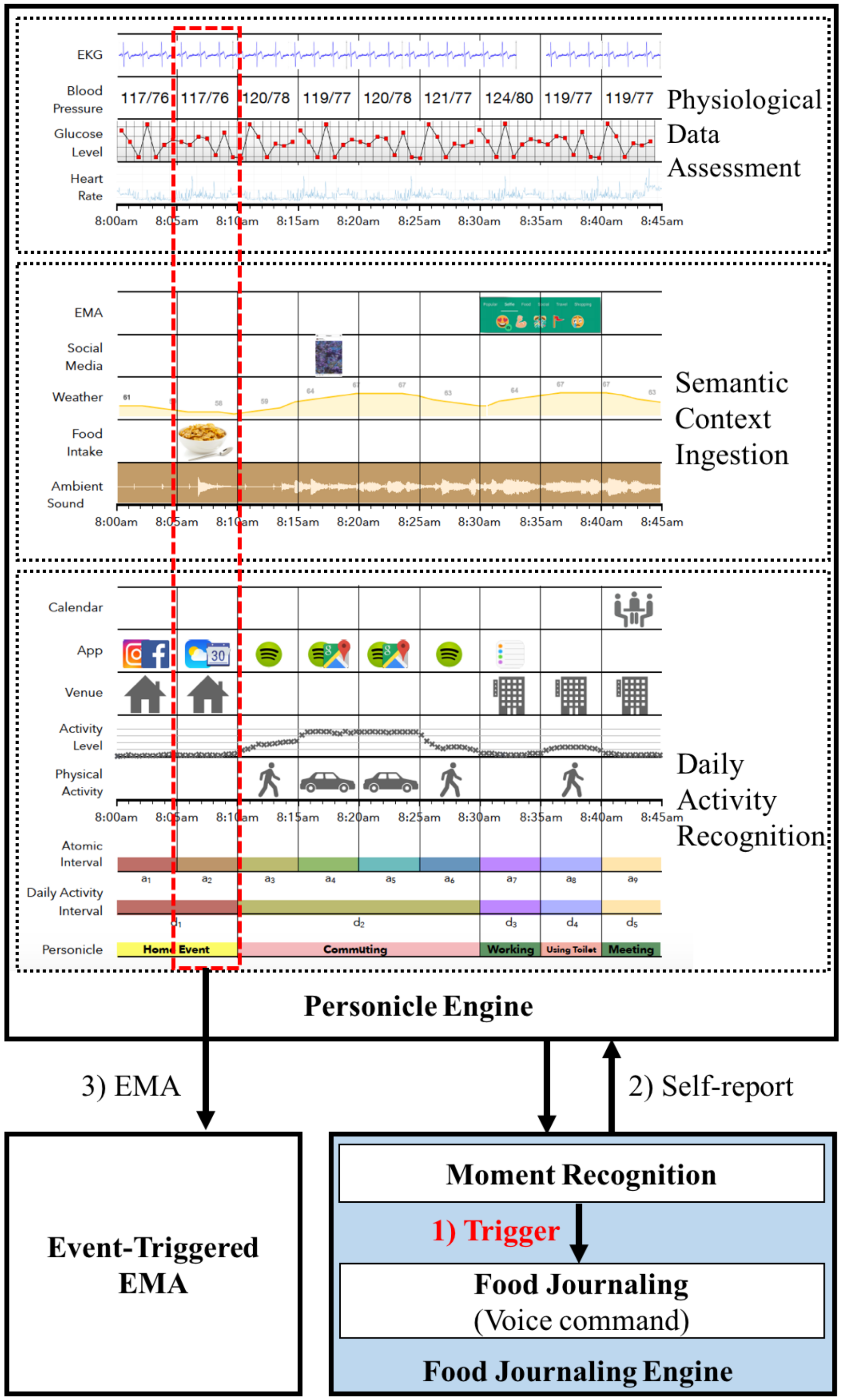


Figure 4.1: Assessing an enhanced ecological moment of eating activity through *Personicle* based food journaling.

The traditional method of keeping a food journal is manually recording meals in as much detail as possible by including the portion size, number of servings and calories, time, location, or even the people around us. This detailed description is effective, but it is very easy to forget or procrastinate logging food entries, which then results in more difficulty recalling meals eaten or even early abandonment of the journaling process. Although remarkable technical progress in automating the food journaling process has been made, it is still highly dependent on the user to take initiative and then requires them to do things such as taking pictures of their food, scanning barcodes, or searching for foods in a food database. These methods tend to be unreliable and require many actions on the user's part which can then lead to the problems endemic, such as inaccurate or missed food entries and early abandonment.

There are currently two main challenges in improving food journaling: (1) triggering a food journaling process in a timely, proactive manner, and (2) improving the reactive self-reporting procedure while preserving high measuring accuracy. In this thesis, we offer an alternative method to current food journaling through a event-triggered Ecological Momentary Assessment (EMA). We try to consider both the proactive and reactive perspectives that can unobtrusively enhance the event-triggered EMA as in Figure 4.1 and thus move forward as fully-automated food journaling.

To solve the first challenge in food journaling, a timely reminder is essential. The best time for food journaling is when people start eating a meal since they know what they are eating at that moment. Thus, our approach begins with finding eating moments so that we can trigger a food journaling process at the correct time. More specifically, we find two kinds of eating moments; one is "eating at a restaurant", and the other one is "eating at home". Our previous research [84, 86, 53, 52] has developed technology to automatically recognize the former eating moment through smartphone based *Personicle*, which provides a person's time-ordered list of daily activities. However, there have been difficulties recognizing "eating at home" due to the lack of available smartphone sensors. In this thesis, we try to solve the latter problem by pulling heart rate signal in *Personicle*.

We then try to keep a food journal through what we call the event-triggered EMA. To do this, we provide an environment that the user can log their meals by describing what they just ate via voice commands. Essentially, taking pictures of foods and barcodes to create food entries have shown to be inaccurate or inconvenient. For this reason, we offer an alternative, which is to use the voice commands to create food entries by using speech-to-text technologies and natural language processing. Meanwhile, the *Personicle* system automatically assesses the user's ecological moment by including the food entries as well as various contexts of the eating moment and thus unobtrusively complete the event-triggered EMA.

Our main contribution in the area of food journaling is 1) providing an event-triggered EMA to automate the food journaling process, thereby 2) encouraging people to keep a well-balanced diet, as well as 3) helping them develop healthier dietary habits. We make these contributions by providing a general eating moment model that can automatically recognize the starting moment of eating, and then prompting the user to begin a voice command food journaling method. We validate our approach with an experiment for 3 months with 3 users who are using the *Personicle* system with Fitbit Charge 2 or Blaze. Our food journaling scenario is as follow:

1. Users install *Personicle* on their Android phone and start using it with a Fitbit device, such as Charge 2, Blaze, Ionic, or Versa, which are the most common devices in the market.

2. After a cold start period lasting a week, the *Personicle* system starts requesting the voice command food journaling whenever it recognizes a starting moment of breakfast, lunch, or dinner. It generates a unique pattern of vibration so that it can let the user know that it's time to make a food journal.

3. Then, the user simply speaks whatever he is eating at that moment, such as "I'm eating a slice of pizza with buffalo wild wings and a cup of Coke for lunch".
4. After that, the *Personicle* system extracts food items (e.g., pizza, buffalo wild wing, coke), quantity of the food (e.g., one slice, a cup), and meal type (e.g., lunch).
5. Finally, the *Personicle* system makes an event-triggered EMA by capturing other contexts around the eating moment, such as glucose level, stress level, emotion, weather, location, other people with the user, or even past events before the eating activity.

4.2 Related Work

Food journals are currently the most commonly used method for analyzing dietary intake. An early method of keeping a food journal was through anecdotal summaries, such as lengthy interviews and questionnaires [75, 26]. This method has shown to be a cumbersome and inefficient way of monitoring dietary intake. Recall-based paper diaries have been another popular alternative to understanding dietary habits of people [17]. However, both of these methods as of recently have been phased out in favor of mobile food journals driven by advancements in technology. Commercial mobile applications such as MyFitnessPal[2], Fitbit[3], or Bixby Vision [4], support food databases so that users can easily and accurately log calories and nutritional information. Utilizing databases takes the guess work out of logging calories and result in more accurate food journals. Additionally, most of these mobile applications also support features such as barcode scanners and shortcuts for commonly eaten foods in order to quickly journal food information [106]. However, even though these technologies increase convenience and usability for maintaining a food journal, it is still highly dependent on the user to take initiative and remain consistent in their food logging [92].

[2]https://www.myfitnesspal.com/
[3]http://www.fitbit.com/
[4]https://www.samsung.com/global/galaxy/apps/bixby/vision/

Researchers in the field of computer vision have started to incorporate food image recognition in order to make food journaling more convenient and consistent for users. They have addressed the challenges in image recognition by developing machine/deep learning algorithms to recognize food items [25, 15, 56, 59, 76, 119]. For example, FoodLog has contributed to a record of users' food intake simply by taking photos of their meals [5, 4]. It also allows users to input textual descriptions based on image retrieval techniques. This kind of approach mainly uses mobile applications or wearable cameras (e.g., DietCam [61], Menu-Match [11], FoodCam [59]) for food recognition, assessment, and journaling. In addition to food image recognition, food quantity estimation has been another important aspect of the research to automate the assessment of food intake. [91, 89, 108]. However, the classification of food image is still a very difficult task and is still fairly inaccurate, since there are various confounding factors, such as visually similar foods, home made foods, quality of photos, and lighting conditions [92].

Another important challenge of automated food intake monitoring involves eating moment recognition. Since Stellar et al. recognized eating event by measuring tongue pressure through oral strain gauge in 1980s [102], researchers have used various sensing modalities for eating moment recognition. One of these modalities used an acoustic sensor to monitor swallowing and chewing sound through the ear, laryngopharynx [98], or neck [117, 23]. Some others have utilized on-body inertial sensor to detect eating or utensil (e.g., fork, spoon) gesture [105, 9, 55, 31]. More recently, researchers have looked to analyzing the heart rate response for eating moment recognition. Shinji et al. analyzed short-term and long-term features of heart rate changes, and revealed that there is another heart rate peak after eating for few hours [48]. Despite all the progress made in this field, most of these proposed methods are impractical for real-life usage, requiring multiple on-body sensors, or suffer from several limitations, such as weak gesture model, or experimental constraints (e.g., time, situation). To the best of our knowledge, there is no approach seeking for unobtrusive food journaling that automates the process of keeping a food journal by utilizing all the contexts around eating activity. The biggest difference our work offers is that we generate a event-triggered EMA by automatically assessing ecological moments of eating activity at the correct time.

4.3 Eating Moment Recognition

We try to trigger a food journaling process for two different kinds of eating moments, "eating at restaurant", or "eating at home". Currently, we have successfully recognized eating activity if people are eating outside of their homes, such as restaurants, or their favorite breakfast/lunch/dinner spot [84]. However, it has been difficult to recognize when the user is eating at home since we were unable to find useful features that can classify "eating" from "home event". In this section, we propose a novel method of recognizing eating moment by pulling heart rate signal in the chronicle of daily activity. This approach mainly focuses on finding the starting moment of the eating.

We first hypothesize that heart rate is increased when people start eating a meal, and then maintain that increased rate while they are eating. This is supported by the studies published in psychophysiology, nutritional science, and electro-cardiology, which have proved that heart rate is generally higher after meals [54, 45, 97]. We also propose another hypothesis that there is a unique activity pattern before "eating at home", such as preparing food, or moving to dinning room. We try to build a general eating moment model, which is designed to learn the aforementioned features. Figure 4.2 shows the sensor data processing pipeline starting with multi-modal sensor data tracking, such as heart rate and step count. We segment the low-level signals whenever the pattern of physical activity is changed, and then extract features from the segmented results, and finally recognize eating moments through machine learning algorithms.

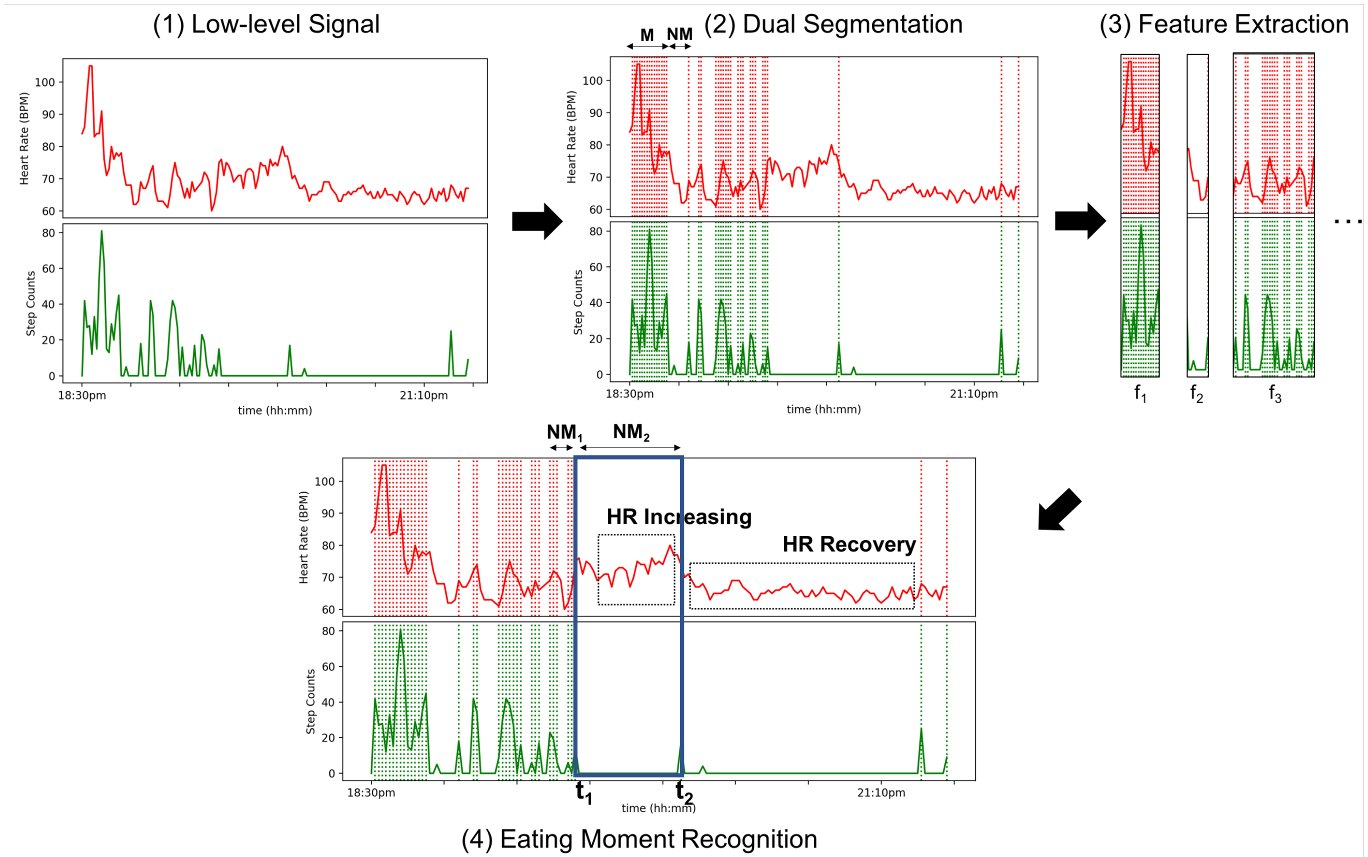


Figure 4.2: The sensor data processing pipeline for the eating moment recognition. M: a moving type of daily activity interval, NM: a non-moving type of daily activity interval, NM_1: a sample of non-moving type of daily activity interval before eating, NM_2: a sample of non-moving type of daily activity interval indicating an eating moment.

4.3.1 Double Segmentation

We define segmentation as a process of partitioning a day into multiple sets of daily activity intervals. To segment a day, we assume that transitions of physical activity pattern (e.g., moving to non-moving, or non-moving to moving) can be involved in the changes of other attributes, such as location, which can then be considered as ending one daily activity (e.g., "home event") and starting another (e.g., commuting) [84, 86]. Thus, as shown in Figure 4.2, we segment sequential atomic intervals of the day into moving (M) or non-moving (NM) type of daily activity intervals.

As shown in Figure 4.1, we have used five-minutes as the length of an atomic interval in order to segment the sequential atomic intervals into coarse-grained daily activity intervals. In our previous research [84], we finally classified this coarse-grained daily activity intervals into daily activity, such as "home event". However, in this research, we decrease the length of atomic interval from five-minutes to one-minute so that we can re-segment coarse-grained daily activity intervals into fine-grained daily activity intervals, and thus classify "eating at home" activity from "home event". To do this, we suggest a recursive binary interval growing technique (RBIG).

Recursive Binary Interval Growing (RBIG): Algorithm 1 shows the process of recursively re-segmenting a coarse-grained daily activity interval into fine-grained daily activity intervals. We first use five-minute for segmenting a coarse-grained daily activity interval. Once we segment the coarse-grained daily activity interval, we come back to the seed atomic interval, and then re-segment the coarse-grained daily activity interval into fine-grained daily activity intervals by using one-minute atomic intervals. As shown in Algorithm 1, we assign the seed atomic interval S_j at the beginning of the process, and then start calculating the similarity $\delta(i)$ between the seed atomic interval S_j and the incoming atomic intervals A_i,

Algorithm 1 Double Daily Activity Segmentation using RBIG

Input: current atomic interval $\mathbf{A}_i$, $\mathbf{a}_i$, seed atomic interval $\mathbf{S}_j$, $\mathbf{s}_j$ where $\mathbf{a}_i$ and $\mathbf{s}_j$ are 1 minute interval

Output: daily activity interval set $\mathbf{R}$;

1: Set $\mathbf{S}_j = \mathbf{A}_i$ if $i = 0$ and $j = 0$, or $\mathbf{S}_j = \varnothing$, and then set $k = 0$, $m = 0$, $p = 0$;

2: **Repeat**

3: Wait for next atomic interval, $\mathbf{A}_i = \mathbf{A}_{i+1}$, $i = i + 1$;

4: Extract activity level $\mathbf{l}_i$, and total amount of moving time $\mathbf{t}_i$ from $\mathbf{A}_i$;

5: Extract activity level $\mathbf{l}_j$, and total amount of moving time $\mathbf{t}_j$ from $\mathbf{S}_j$;

6: Calculate $\delta(i)$;

7: If $\delta(i) = 1$, make a daily activity interval $\mathbf{R}_k$ by segmenting from $\mathbf{A}_j$ to $\mathbf{A}_i$;

8: Set $p = i$, reset $i = j$, $\mathbf{a}_i = \mathbf{S}_j$, $\mathbf{s}_j = \mathbf{S}_j$;

9: **Repeat in the daily activity-interval $\mathbf{R}_k$**

10: Assign next atomic interval $\mathbf{a}_i = \mathbf{a}_{i+1}$;

11: Extract activity level $\mathbf{l}_i$, and total amount of moving time $\mathbf{t}_i$ from $\mathbf{a}_i$;

12: Extract activity level $\mathbf{l}_j$, and total amount of moving time $\mathbf{t}_j$ from $\mathbf{s}_j$;

13: Calculate $\delta(i)$;

14: If $\delta(i) = 1$, make a daily activity-interval $\mathbf{r}_m$ by segmenting from $\mathbf{a}_j$ to $\mathbf{a}_i$;

15: Set $m = m + 1$, $j = i$, set new seed atomic interval $\mathbf{s}_j = \mathbf{a}_i$

16: **until** i == p

17: Set $k = K + 1$, $j = p$ set new seed atomic interval $\mathbf{S}_j = \mathbf{A}_p$

18: **until** the system is terminated.

19: **return R, r**

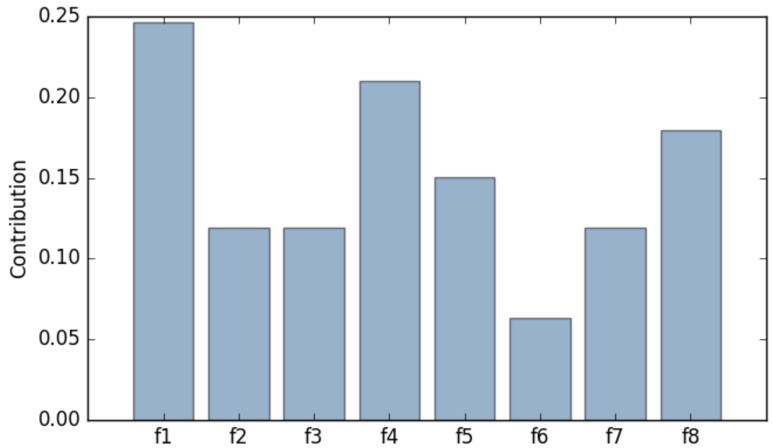


Figure 4.3: Contribution of each feature for eating moment classifier. $f1$: past average heart rate, $f2$: current average heart rate, $f3$: Δ average heart rate, $f4$: the amount of past NM time, $f5$: the amount of past M time, $f6$: the number of past steps, $f7$: the amount of moving time (past 30 min), $f8$: heart rate variation.

every interval minutes (e.g., 5 min, 1 min). $\delta(i)$ can be formulated as follow:

$$\delta(i) = \|f(S'_j) - f(I'_i)\|_2^2 \tag{4.1}$$

where S'_j is $\{l_j, t_j\}$, I'_i is $\{l_i, t_i\}$, $f(x)$ is a classifier to identify if an atomic interval is the moving (1) or non-moving (0) type of interval. When $\delta(i)$ is equal to 1, we segment the atomic intervals from I_j to I_i, and then make it as a coarse-grained daily activity interval R_k. After that we try to segment R_k into fine-grained daily activity intervals r_m by repeating the same process with one-minute atomic interval.

4.3.2 Feature Extraction and Selection

We extract and select eating moment features from the fine-grained daily activity interval. We heuristically explored the 3 months worth of daily activity intervals so that we can identify latent features underlying the visible sensor data streams. In our exploration, as shown in

Figure 4.2, we first could see that heart rate is increased when people start eating, as seen in NM_2 between time t_1 and t_2, and then the increased heart rate remained high during the meal time. We next tried to see how average heart rate is different between the eating moment and the past since it can be another unique feature that determines the starting moment of eating. To do this, we excluded all the moving type of daily activity intervals, which can highly affect the increase of heart rate, and then tried to correctly compare the difference in average heart rate. Such heart rate effects of dietary intake can be seen in between NM_2 and NM_1. Based on this finding, we extracted eating moment features, such as average heart rate, heart rate variation, and heart rate difference between current and past daily activity intervals. In addition, we also extracted more features from activity patterns before eating. As shown in Figure 4.2, there will always be a certain amount of moving time just before beginning an eating activity due to events like preparing the food, or moving to the dining room. Thus, we also extracted features such as step count, and the amount of moving and non-moving time.

Lastly, we used the Correlation-based Feature Selection (CFS) criteria so that we can select the best subset of extracted features [43]. This algorithm evaluates how accurately all the features in the feature subset are indicative of the target class. It also can evaluate which features are not correlated with each other by providing complementary information for each of them [92]. To evaluate our heuristically extracted features, we first trained the eating moment classifier with all the features, and then compared the recognition performance (F-measure) to those of other classifiers, which are trained without a particular feature subset. As shown in Figure 4.3, all the extracted features show some performance degradation, which means these features are all highly indicative of the starting moment of eating activity. Based on this result, we selected all the extracted features.

4.3.3 Eating Moment Recognition

With the eating moment features, we tried to build a general eating moment model that can classify the fine-grained daily activity intervals into eating or non-eating moments.

There are several things to keep in mind regarding general eating moment modeling. First, it requires relative values due to the fact that everyone has varying heart rate ranges. Thus, we converted the heart rate C of each individual data into heart rate levels, which are discretized w-dimensional space, by a vector $\bar{C} = \bar{c_1}, \bar{c_2}, ..., \bar{c_i}$. To do this, we used a discretization technique, Symbolic Aggregate Approximation (SAX) that reduces the time series of arbitrary length n into the w-dimensional space as follows [68]:

$$\bar{c_i} = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} c_j \tag{4.2}$$

Second, we also converted the activity and time features (e.g., moving pattern, step count, and the amount of moving and non-moving time into discretized levels) given that these also differ from person to person. Furthermore, we reflected day-night differences in body temperature and heart rate by training breakfast, lunch, or dinner model separately.

For training the classifier, a Support Vector Machine (SVM) with a Radial Basis Function kernel (RBF) was applied to the training dataset [21]. In the test phase, we ran the classifier whenever the *Personicle* system finds the fine-grained daily activity intervals. If the interval is classified as a starting moment of “eating at home”, we assume the whole range of this interval as the eating moment, as shown in t_1 and t_2 in Figure 4.2.

4.4 Voice Command Food Journaling

As a first step towards building a voice command food journaling engine, we define a basic sentence protocol that has to be spoken by a user to apply to text analysis. Then, the prototyped solution accepts a voice based input describing food intake, transcribes the input to text by using Google voice API[5], breaks down the input sentences according to the pre-defined protocol, and then extracts information for keeping a food journal. In this section, we take the full advantage of using APIs in order to ensure maximum results.

4.4.1 Protocol

The major components that are important on a food journal is food item, meal type and quantity. This information enables the ability to obtain nutrition information and calorie intake by querying a food database, such as USDA Food Composition Databases[6]. Therefore, we suggest users to include the aforementioned information with an actuating verb, such as "eat" or "have", when they describe what they are eating. The actuating verb would help to increase the accuracy of voice command analysis since it points out the most important sentences of all conversation. The quantity, food item, and meal type information should be listed sequentially after this actuating verb. Here are some examples.

- **Protocol 1:** *I'm eating* (actuating verb) / *a* (quantity) / *cheeseburger* (food item) / *for lunch* (meal type)
- **Protocol 2:** *Two* (quantity) / *garlic naans* (first food item) / *and a cup of* (quantity) / *Coke* (second food item)

The first example shows a complete protocol that we want to see from the voice command.

[5]https://cloud.google.com/speech-to-text/
[6]https://ndb.nal.usda.gov/ndb/doc/index

It has an actuating verb “eat”, and then food item “cheeseburger”, quantity “one” and meal type “lunch”. In addition to the complete form, we also can accept a simplified protocol as can be seen in second example. We then try to obtain food items “garlic naan”, “Coke”, and quantities of the foods “two”, “a cup”. Multiple food information also can be acceptable once it is listed sequentially.

4.4.2 Information Extraction

After the speech has been converted to text via Google Voice API, the next task is to extract the key information out of the text. We utilize a natural language processing API, TextRazor[7], so that we can accuarately extract the keywords in a sentence and the classification results of those keywords. Based on the result, we first find the sentences, which include actuating verbs, such as “eat”, or “have”. We then filter out all the keywords from the sentences by checking for foods, meal types and numbers, keeping only the necessary information to create a food entry. After that, we extract the nearest numbers from the food items to map the quantity to the food. Lastly, if there is no meal type in the sentence, we extract this information from the tense of the verb or current time that the food entry was created.

4.5 Event-Triggered EMA

The current phase of our event triggered EMA is shown in Figure 4.1. We can automatically assess personal ecological moments without any questionnaires since our *Personicle* system continuously monitors the chronicle of daily events as well as semantic contexts and physiological signals. Therefore, once the *Personicle* system recognizes an eating moment, we can create an EMA of the eating activity consisting of stress level, glucose level, emotion,

[7] https://www.textrazor.com/

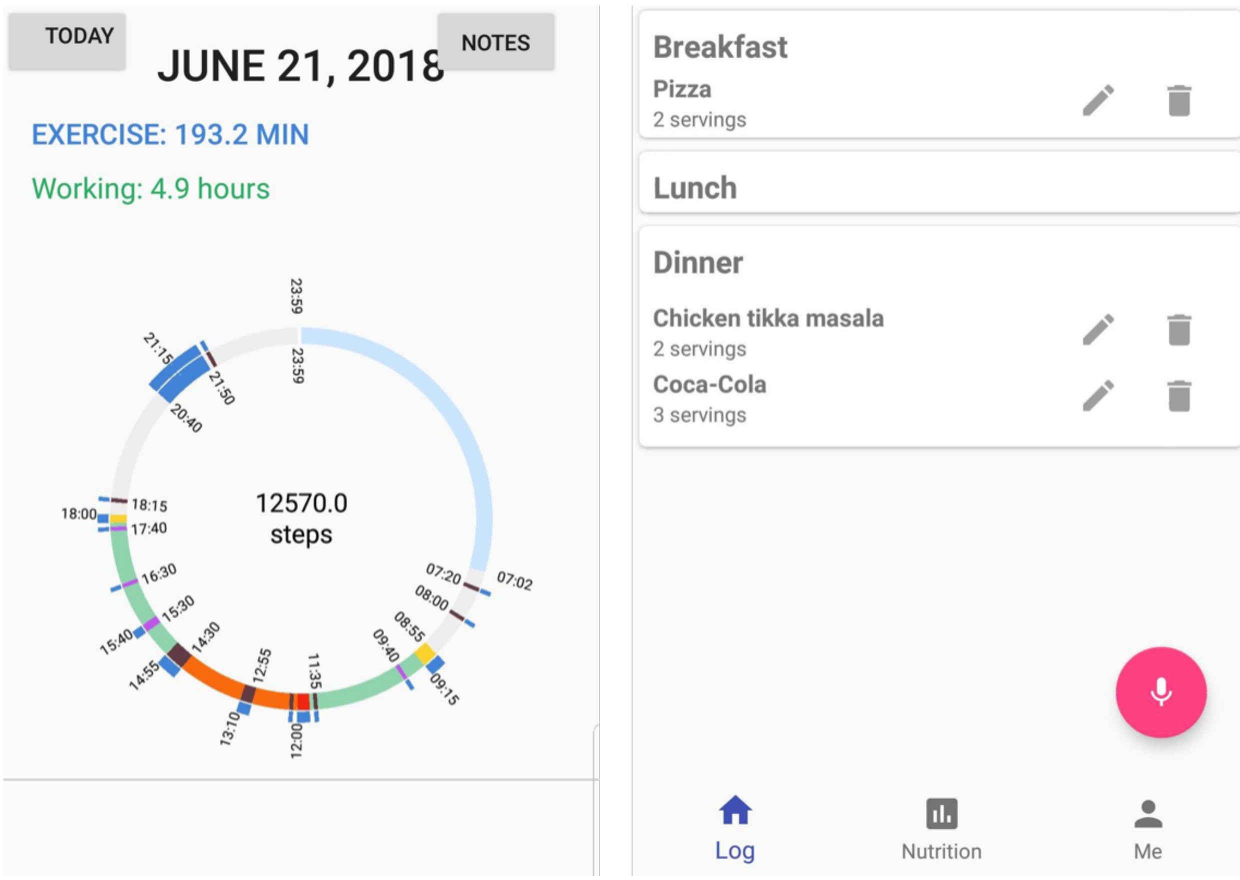


Figure 4.4: Screenshots of *Personicle* system including daily event recognition and food journaling.

weather, location, other people with the user, and even past events before eating and their frequency. Additionally, if the user reacts to the voice command request, we also include the food eaten, the quantity of the foods, the nutrition value, and the calorie intake in the event-triggered EMA. The ultimate goal of our event triggered EMA is to fully automate the food entry process, and thus keep a food journal without any user interventions, such as taking pictures. We see the potential of automating the food entry process in that there are distinct differences in heart rate patterns depending on the food type. Such research could open many opportunities to support innovative studies in pervasive health.

4.6 Experimental Validation

In this section, we first elaborate how participants collected multi-modal sensor data and how they labeled starting moments of their eating activity. We then validate our eating moment recognition method by evaluating the performance of the general eating moment classifier. After that, we verify the voice command food journaling using the two kinds of protocols defined in Section 4.1.

4.6.1 Data Collection

We implemented a *Personicle* system on Android as in Figure 4.4. This *Personicle* system always runs in the background, collects both smartphone and wearable sensor data as described in Figure 4.1, and stores the data in Google Firebase Database[8] in real-time. To collect experimental data, we hired three participants who are using Galaxy S9 plus (OS 8.0.0) with Fitbit Blaze, Galaxy S8 (OS 8.0.0) with Fitbit Blaze, and Google pixel (OS 8.1.0) with Fitbit Charge2 for three months, and asked them to install the *Personicle* application

[8]https://console.firebase.google.com/

on their smartphone.

The three participants manually labeled eating moments by using Samsung Health[9], Khana-Pal[10], and the *Personicle* application, respectively. We then used the dual segmentation technique (RBIG) so that we can extract fine-grained daily activity intervals, which include the labeled moments as described in Section 3.1. Therefore, the participants didn't have to provide all the start and end times of each eating activity, but simply labeled a time stamp in the meal time. To guarantee the quality of the ground truth data, we requested them not to label the eating moment by guessing if they miss the food journaling time. The total numbers of atomic intervals, daily activity intervals, and eating labels were 90720, 1785, and 255 respectively.

4.6.2 Eating Moment Recognition

We evaluated our model by calculating precision, recall, and F-measure. We performed 10-fold cross validation on each participant's daily activity interval data and then averaged the results to obtain an overall result.

Since eating-labeled daily activity intervals had different segment sizes, from five minutes to more than an hour, we needed to make their size uniform so that we can train and test the eating moment classifier. Thus, we tried to find the optimal sub-segment size of the eating-labeled daily activity intervals that most represents a starting moment of eating activity. To do this, we first explored all the sub-segment sizes of eating-labeled daily activity intervals and found that 80.9% of them belong to a range between five minutes and twenty minutes. Then, we chose the intervals of 5, 10, 15, and 20 minutes among those ranges considering that *Personicle* system has a five-minute data processing interval. Afterwards, we trained the classifiers separately with the chosen sizes and tested the classifiers to see which sub-

[9] https://www.samsung.com/us/support/owners/app/samsung-health
[10] https://play.google.com/store/apps/details?id=com.foodie.android

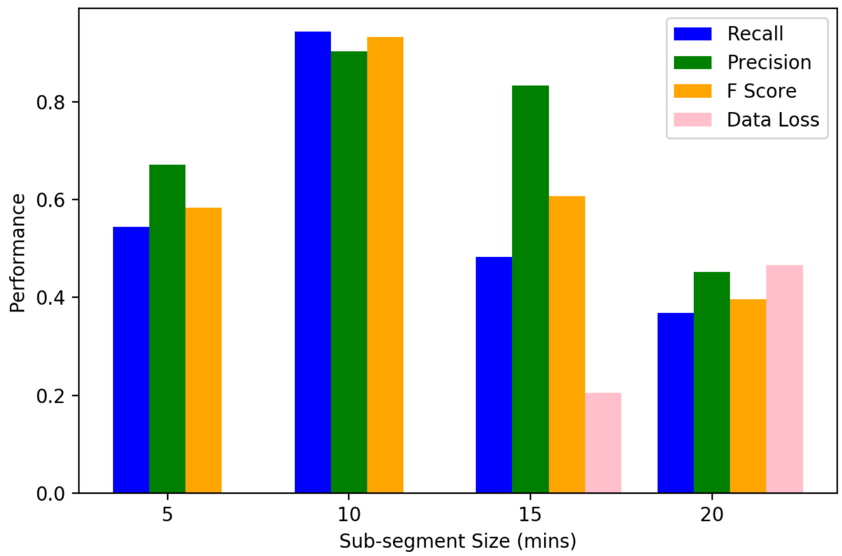


Figure 4.5: F-measure of the eating moment recognition across different sub-segment sizes.

segment size most accurately recognized the starting moment of eating. Figure 4.5 shows the performance of eating moment recognition across the four different sub-segment sizes. From this experiment, we can see that using a 10-minute sub-segment size provides better results compared to other sub-segment sizes. This indicates that if a sub-segment size is too long or too short, it is difficult to properly represent the features of eating moments. We also can see that there is data loss when the sub-segment sizes are too long. It means that there are many number of eating activities that are less than 15 minutes. Furthermore, the results shown in Table 4.1 indicate that the performance of eating moment recognition across different sub-segment sizes is not person-specific but can be generally applied for all the users. Based on these findings, we found that 10-minute is the most optimal sub-segment size for accurately reflecting the selected features as described in Section 3.2.

With the 10-minute sub-segment size, we next verified the performance of our general eating moment classifier. Table 4.2 presents the precision, recall, and F-measure of the eating moment classifiers with and without SAX algorithm. From this comparison, we can clearly see that the performance of all classifiers improved with the SAX algorithm. This indicates

Table 4.1: F-measure of each user's eating moment recognition across different sub-segment sizes.

User	Sub-segment Size			
	5 minutes	10 minutes	15 minutes	20 minutes
User 1	0.7500	**0.8686**	0.8236	0.8571
User 2	0.6000	**0.8889**	0.6667	0.0000
User 3	0.4000	**1.0000**	0.3333	0.3333

Table 4.2: Performance of eating moment classifiers trained with and without SAX algorithm in terms of Recall (R), Precision (P), and F-measure (F). *10-NN*: 10 Nearest Neighbors, *NB*: Naive Bayes, *RF*: Random Forest, *SVM*: Support Vector Machine.

Model	Without SAX			With SAX		
	R	P	F	R	P	F
10-NN	0.4286	1.0000	0.6000	0.8750	1.0000	0.9333
NB	0.8571	0.4615	0.6000	0.7778	0.8750	0.8235
RF	0.5714	0.8000	0.6667	0.7500	1.0000	0.8571
SVM	0.2857	0.6667	0.4000	**0.8750**	**1.0000**	**0.9333**

that the pattern of heart rate and activity around the eating moment are relatively similar from person to person. More interestingly, Table 4.2 shows that all the tested classifiers achieved significantly good performance. Even the 10-NN classifier shows comparable results to SVM on the precision, recall, and F-measure. It can signify that the eating moment sub-segment set can be easily separable even though it is in high dimensional space. However, the relatively low recall in the classifiers with SAX means that more unique features are still needed to classify more diverse eating moments. Considering that we are planning to add more features to the eating moment model, we chose the SVM classifier, which is low cost, and high speed, and fairly robust against over-fitting, especially in high-dimensional space.

Lastly, we tested the performance of eating moment recognition over time. According to Lin et al., the small subset of data can deteriorate the efficiency of SAX since the discretization technique is based on normal distribution [68]. Considering that the *Personicle* system began with zero user data, our initial performance also could be highly affected by incorrectly discretized features. Because of this, we tried to find the cold start of our general eating moment classifier in order to understand how long it takes to obtain reasonable results. As

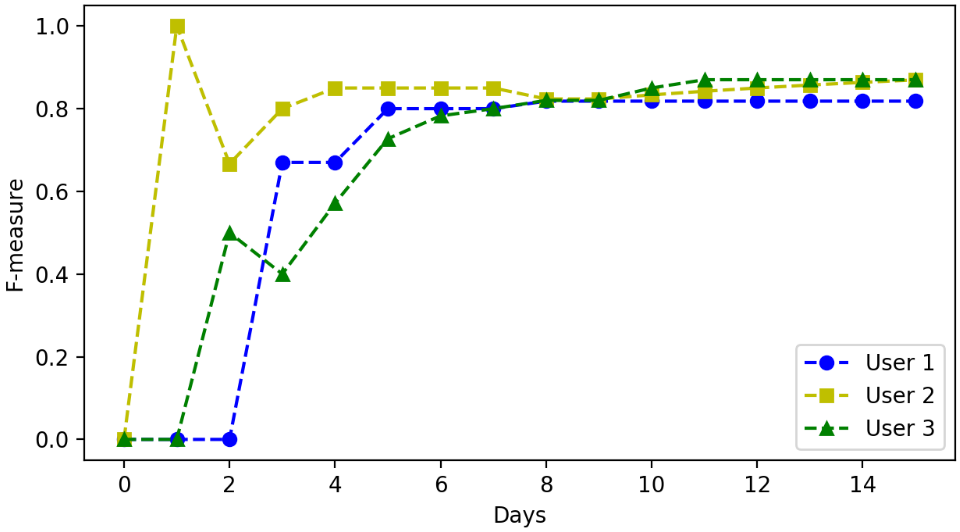


Figure 4.6: Performance of eating moment recognition using SVM classifier over time.

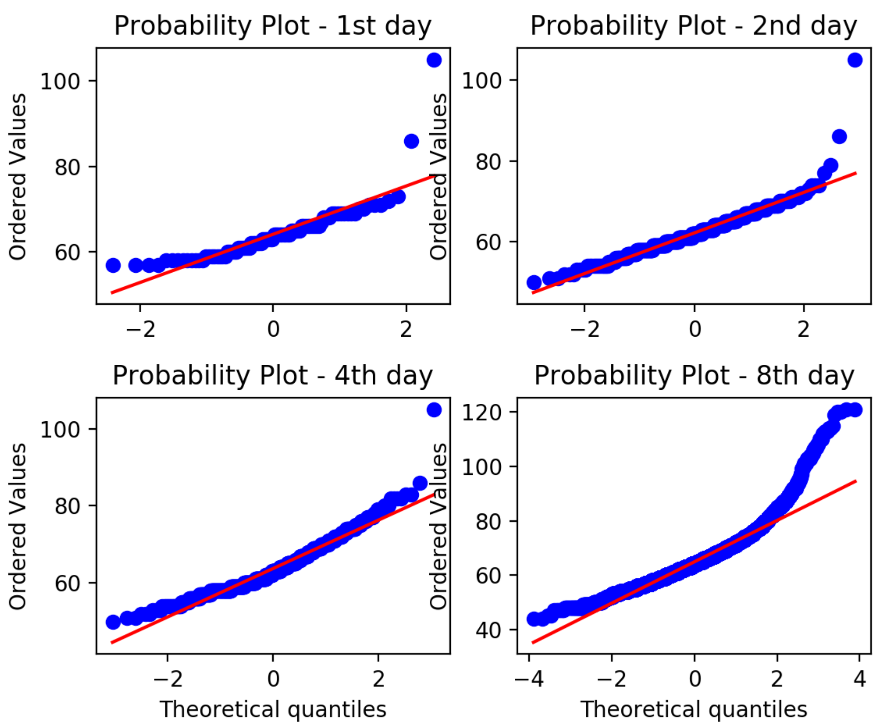


Figure 4.7: A normal probability plot of the distribution of evening heart rate over time.

we can see in Figure 4.6, there was a wide fluctuation in the recognition performance for the first few days. To determine the reason for this, we tried to draw a normal probability plot of heart rate data as shown in Figure 4.7. In the figure, the highly linear nature of the plots indicates that heart rate data comes from a Gaussian distribution. However, the first few days of each plot show that there was lack of heart rate data in high range, and thus resulted in incorrect heart rate discretization, especially between 70 bpm and 120 bpm. As an example, 70 bpm was discretized as level 8, 9, 8, and 7 out of 10 in the 1st, 2nd, 4th, and 8th day's normal probability plot, respectively. This small difference could significantly affect the eating moment recognition performance given that recognizing eating moment is highly dependent on heart rate variations.

Our results show great potential for automating the eating moment recognition in a practical manner using a smartphone and wearable device combination. However, the current research has a few limitations, namely sub-segment size (e.g., 10 minutes) and location (e.g., home) in the recognition process. While these constraints can be effective for the recognition of "eating at home" events, these can preclude the possibility of recognizing other eating events, particularly those that are less than 10 minutes, or those that happen during physical activity, or those occurring in outdoor places. Therefore, in our future research, we aim to analyze more distinct heart rate patterns, which are not only the analysis of starting moments, but also overall fluctuations in the entire eating moments. To do this, we will collect more diverse eating cases with more participants and try to analyze the effect of different factors on heart rate so that we can better understand correlations between heart rate and foods.

4.6.3 Voice Command Food Journaling

In this section, we evaluated the performance of voice command food journaling by testing predefined protocols. We selected the most popular food items from the following seven

Table 4.3: List of incorrect results among all the test cases. A_1: Korean accent, A_2: Indian accent, A_3: Native English accent.

Food	**Voice Command**	**Food Extracted**		
		A_1	A_2	A_3
Indian	Idly	Italy	Elite	O
	Hot spicy sambhar rice	Some hard	O	sambar
Korean	Bingsu	Kingsville	O	O
American	I had large stack of 6 pancakes	O	6 cake	O
Chinese	Szechuan chilli chicken	Sachin	Sichuan	2chan
	Wonton	1 ton	O	O
Italian	Lasagna	O	Sonya	O
Mexican	Quesadilla	O	Jesse Diaz	O

different food types: Indian, Korean, American, Chinese, Italian, Mexican, and Japanese. From this list of different food types, 41 test cases of the two protocols were made (sixteen of Protocol 1, and twenty five of Protocol 2). We evaluated these 41 test cases with three users who have different English accents, which are native English, Indian, and Korean, so that we can also see the effect of accent on the voice command results. Table 4.3 shows all the incorrect results that we obtained from our experiment. The meal types and quantities of the foods were correctly extracted from the test cases, and thus were excluded in the Table.

In the experiment, we found that there are three issues when trying to recognize food items. First, if the user is not familiar with food items, such as Wonton, Idly, Szechuan, Lasagna, Quesadilla, or Sambhar, these are not correctly recognized and extracted as shown in Table 4.3. Second, the voice recognition API has been mainly trained by a native English accent. For example, even though Bingsu and Idly were pronounced by their native accents, which are Korean and Indian respectively, the voice recognition API converted them into wrong words, such as Kingsville and Elite. Third, if there is a short silence between or within words, such as a pause in the word "pancakes" in Table 4.3, recognition accuracy is compromised. Articulating the voice recording method in the programming level will be able to improve this issue. To solve all of the aforementioned problems, we will use more contexts around the voice command moment or apply approximate string matching algorithms so that we can find the most close food item to the converted text.

4.7 Conclusions

Food journaling is a great way to improve health because it lets us monitor our dietary intake, but it can potentially be inaccurate and difficult to maintain. This paper builds towards the research to develop a unobtrusive food journaling method that automates the process of keeping a food journal via common wearable devices. Specifically, this paper focuses on recognizing a starting moment of eating activity to trigger a food journaling process in a timely, proactive manner. Thus, it describes the methodology behind automatically recognizing eating moment with the goal to build an event-triggered EMA. We also propose a voice command food journaling method which makes it simple to keep a food journal while still remaining highly accurate, and thus include the food entries in the event-triggered EMA. Such methods could play a very important role in applications for health and well-being study using personal food journaling data. Results obtained from the three participants show the potential of such an approach for the unobtrusive food journaling. We expect that our further research would allow for automatically recognizing food items at broad level considering that there are distinct differences in heart rate patterns depending on the food type, and thus building a fully-automated food journaling engine.

Chapter 5

Enhancing Events of Daily Living

5.1 Introduction

Activities of Daily Living (ADL) have attracted attention in the multimedia field and other communities since Kahneman [57] demonstrated the effectiveness of ADL in judging quality of life. ADL refers to routine activities, such as eating, bathing, dressing, toileting, and transferring, a normal person performs without external help. The extent to which people continuously perform ADL is one of the significant measures in evaluating their current state and quality of life. Researchers in the multimedia field have attempted to recognize this unobtrusively, using different multimodal streams [35, 80]. By continuously and objectively recognizing ADL, they have showed that ADL can open up new possibilities for research, especially in disease-centric healthcare, such as the assessment of gait in Parkinson's disease [16] or for individuals with dementia [78].

Though ADL is effective, we think this concept can be extended and made significantly more effective. Activities may be used to define events and to characterize, understand, and guide lifestyle. However, as shown in Figure 5.1, a person's single daily activity, such as eating

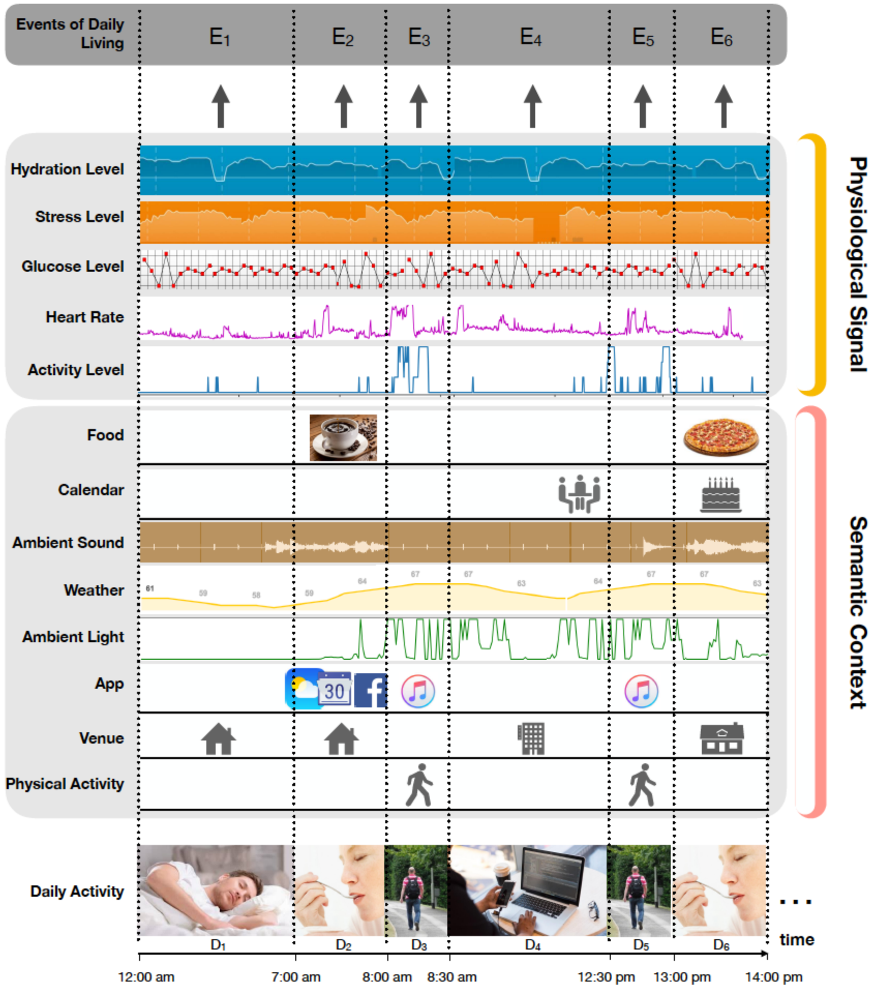


Figure 5.1: Enriching daily activity to daily event. D_1: sleeping, D_2: eating, D_3: commuting, D_4: working, D_5: commuting, D_6: eating

or working, cannot provide enough information about the person except for his ability or inability to perform the activity. On the other hand, an event, such as dinner at Osteria with Tom, provides a lot more information. Events are a common concept in human daily life that represents the aggregation of activities and other attributes into meaningful semantic entities. We believe that events provide better abstractions to correlate the current states of a human being with daily activity, semantic context, and physiological signal, as well as modeling the person, such as his health state, using learning techniques. In this chapter, to associate all of these multimodal data at a higher level, we use the events of daily living that can be considered as semantically enriched daily activities containing spatio, temporal, informational, experiential, structural, and causal aspects as the top data stream in Figure 5.1.

Each person behaves differently to different events in their life. Therefore, a goal of many emerging systems is to model a person to provide the right guidance and to help them use their unique attributes. The primary consideration in building a personal model is to recognize one's events of daily living by using his own physical, biological, social, and personal data. Recently the medical community has recognized this and started emphasizing that one should develop machine learning and AI techniques in the context of a person, rather than populations. This personalized design in clinical science is called N-of-1 in which a single person is the entire trial. In this design, a patient's time periods of treatment exposure are randomized rather than the number of patients, and therefore patient's response to each treatment is compared with each of his other responses. Thus, by collecting the single patient's data over a long period of time, N-of-1 trials can provide high-integrity and evidence-based information only relevant to the patient, as well as a deep assessment of treatment outcomes and adverse effects, a priori hypotheses, and statistical analyses [27]. We think that N-of-1 trials are the most indispensable approach to effectively estimate the uniqueness of each individual as well as correctly model the person based on his own experiences. One of the major technical challenges for N-of-1 trials in the events of daily living is how we can quantitatively and qualitatively measure one's life experiences without any intrusion in his regular life. In this chapter, we describe how to overcome this challenge, and then explore to what extent a daily event can be enriched from a daily activity.

Our approach based on N-of-1 trials begins with the analysis of long term self-tracked data focusing on a person's time periods of daily living exposure rather than the analysis of a group of people. We try to obtain the events of daily living through the Personicle data by relating the aforementioned six aspects of an event to a daily activity. We show this enrichment process through the eating activity, which is one of the most complex events and is central to human experiences and health. In this event, nutrients could be the most important information, especially from a health perspective. Currently, most food recognition approaches involve user interventions, because they require either manual recording of food information or taking photos of the food based on a person's initiative. We develop an unobtrusive self-labeling method for food consumption by focusing on each individual's physical response to different foods. In this approach, we analyze a series of heart rate values to find latent patterns under each consumed food, and therefore use features of the patterns to recognize the food consumption.

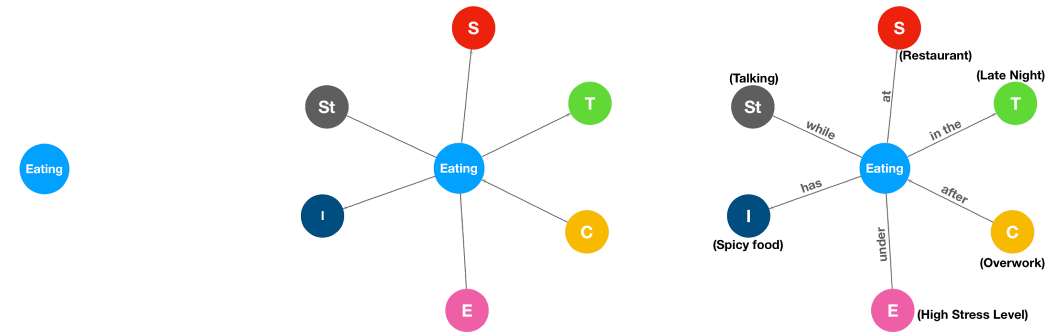
This chapter makes two contributions. First it introduces events in daily life as a more semantic construct than popular ADL, and shows how event knowledge graphs may be effective in populating all event fields. The second contribution is to use multimodal signals from common devices in detecting not only eating activity, but also classifying foods for a specific person in nutrition based groups. For our work, we also introduce the N-of-1 approach that is receiving increasing attention in the medical community for longitudinal study of a person for personalized approaches.

5.2 Related Works

There have been many studies tracking the life experiences of human beings. For example, Mann et al. devised wearable lifelogging devices that can collect visual data from an egocentric camera view to digitally collect one's life experiences [70, 71, 72, 73, 74]. Gemmell et al. developed a software named MyLifeBits, which can capture text, audio, and pictures of a person, in order to store all personal information found in PCs [40, 39, 13]. Gurrin et al. contributed to capture images of life experience along with sensor data, such as location, activity, and environmental information [41, 66, 29]. Aizawa et al. developed technologies to capture heterogeneous contexts in wearable videos [3, 47, 2]. However, much of this early research was focused more on collecting a low-level personal data and storing them in the system rather than recognizing his higher-level life experiences.

With the advancement of sensing technology, research on human activity recognition has become very active [64]. Wearable sensors like accelerometers, gyroscopes, and GPS, have been used to recognize physical activity, such as walking, running, or cycling. Machine learning algorithms like Decision tree [10, 77], Bayesian networks [103, 63], Neural networks [93], Markov models [118, 67], and Ensemble techniques [65] have increased the accuracy of recognition results. The external sensors have been used for more complex human activity recognition like ADL. Some of the approaches tried to employ cameras for identifying high-level activities through data models [88, 34]. Oh et al. proposed another approach based on a segmentation technique. They first segmented the sensor data streams by analyzing the pattern of an user's physical activity, and then recognized the segment as a high-level activity via machine learning algorithms [84, 86].

Another important approach of understanding a person's life experience involves a semantic enrichment of human activity. Riboni et al. described and recognized human social activities, such as a tea party, or meeting with a nurse, with a knowledge-based approach through web ontology language (OWL 2). They tried to accurately model the physical and social environment of users like location of persons, their role, their posture, and their used objects with knowledge engineering experts [94]. Helaoui et al. approached complex human activity recognition, such as cleaning up, by hierarchically decomposing the activity into simpler ones and recognizing its atomic features to get the complex activity in their ontological framework [44]. Meditsko et al. proposed a framework that analyzes object and place features from egocentric vision and accelerometer features from wearable devices in order to model ADL and fuse contexts with the activity [78]. However, most of these studies either remained activity recognition or focused on context enrichment for the purpose of specific research rather than understanding overall life experiences of human beings.



(a) Recognized Daily Activity (b) Related Event Aspect (c) Named Relationship

Figure 5.2: A sample of event knowledge graph for eating

To the best of our knowledge, however, there is no approach seeking to understand the current states of each human being at a higher level than activity. Moreover, we have not seen any approach to unobtrusively obtain semantic information without user interventions. The biggest difference of our work from prior research is that we try to closely tie daily experiences to the events of daily living by containing semantic knowledge found from each individual's regular life.

5.3 Toward Event Knowledge Graph

An event of daily living is a combination of daily activities and other related attributes into a semantically meaningful collection. There are diverse kinds of events of daily living according to one's culture, country, occupation, age and gender. For this reason, we believe that the structure of a daily event should be understood as a descriptive interpretation, rather than a finite number of named ones. We think that a good way to organize this type of data is a graph structure, which includes the relations between entities much like Google's use of knowledge graph (KG). The knowledge graph has completely changed the Google search because it semantically understands the relationships between entities, such as people, places, and things [101]. It helps to provide what it considers to be the most related information to the specific user's query from millions of other web sources. We desire to do the same thing for each individual's daily event in his daily life corpus, and therefore find and describe unique daily events as well as relevant information only for that specific individual.

As shown in Figure 5.2, we first recognize an activity of daily living as a subject entity by using the Personicle system. We then try to relate the subject entity to object entities based on a common event model. More specifically, according to Westermann et al., an event can consist of six aspects, namely temporal, spatial, experiential, structural, informational, and causal [113]. We utilize these event aspects as object entities to enrich the daily activity. The temporal aspect relates to time, such as starting time, ending time, and the length of the event. The spatial aspect provides geographic region of the event like GPS, and the type of location, such as restaurant, home, or work, where the event happened. The experiential aspect offers insights into how the events evolved via multimedia/sensor data, and the structural aspect specifies sub-activities or sub-events. The informational aspect provides further specific parameters that can enrich the event, and the causal aspect offers answers about an event's cause in the chronicle of daily life. Finally, in relating the subject entities to the object entities, we predicate the relationships between the daily activity and event aspects, and therefore build a triplet $T = (dailyActivity, predicate, eventAspect)$ as follows:

- <Eating><at><Home>
- <Eating><in the><Late night>
- <Eating><after><Overwork>
- <Eating><under><High stress level>
- <Eating><has><Spicy food>
- <Eating><while><Talking>

Table 5.1: Missing event aspect ratio when relating event aspects to an eating activity. T: Temporal, S: Spatial, E: Experiential, St: Structural, I: Informational

Missing Event Aspect Ratio (%)					
User List	T	S	E	St	I
User 1	0	6.06	0	0	**97.58**
User 2	0	2	0	0	**89**
User 3	0.72	24.64	0	0	**86.3**

Further steps, like querying or curation in the graph, will be handled in future research.

5.4 Eating Activity Enrichment

We first tried to build event knowledge graphs for eating by relating the event aspects to eating activities for the purpose of seeing to what extent they can be enriched. Table 5.1 shows the results obtained when exploring three users' Personicle data sets. The temporal aspect can be easily found since the time stamp and duration of the activity already existed in the Personicle dataset. The spatial aspect, such as venue name and type, was sometimes missing due to the lack of GPS signal, but normally could be obtained if the smartphone was connected to the Internet. We related sensor data, such as ambient light, ambient sound, or physiological signals, to the experiential aspect, and sub-activities, such as talking or physical activities, which was recognized by Personicle, to the structural aspect. However, for the informational aspect, we rarely found data except information included in the calendar application. Furthermore, the food item, which is one of the most significant pieces of information for eating, was completely missing. Therefore, we mainly focused on complementing the lack of the informational aspect, especially for food consumption, by suggesting an unobtrusive method.

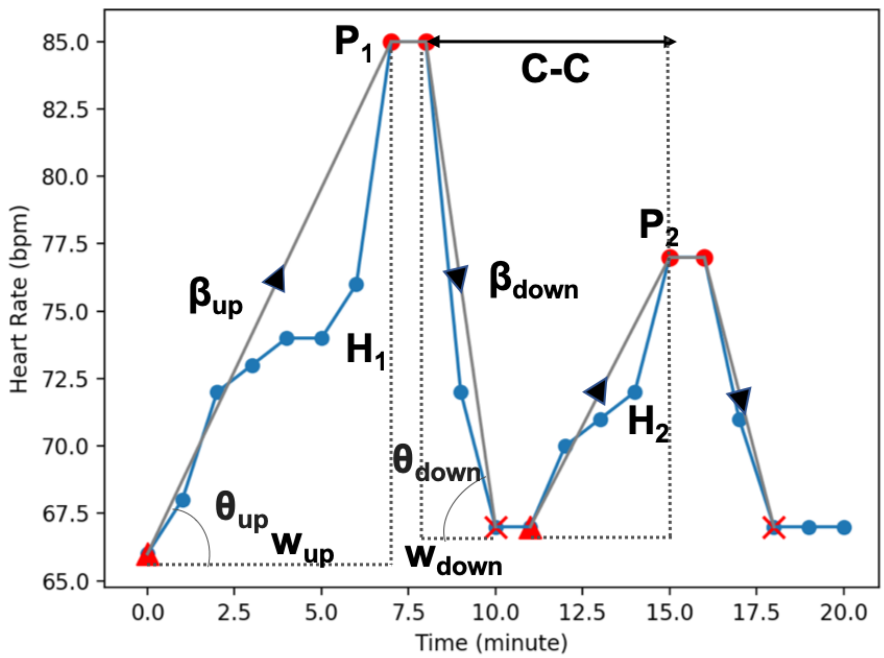


Figure 5.3: A sample of a unique heart rate cycle in response to food intake. This sample displays 21 median-filtered heart rate values and their structural features.

5.4.1 Feature Extraction

We hypothesize that the human body reacts differently to different types of food. This is based on an observation that heart rate forms a unique cycle in response to food intake. Heart rate data obtained by commercial wearable devices, such as Fitbit or Garmin, does not include original Photoplethysmogram (PPG) signals, which is fundamental raw data for feature extraction. We try to overcome this problem in feature extraction by suggesting a method based on the series of beats per minutes.

We first apply median filter to remove noise. Figure 5.3 shows a sample of the median-filtered series of heart rate values. Unlike PPG signals that can be characterized by sinus rhythm showing standard waves, segments, and intervals, the series of heart rate values, as in Figure 5.3, does not form these standard patterns. Moreover, there are a number of factors that can affect the change of heart rate, such as emotion, stress, or health, and thus even the number

Table 5.2: Definition of the structural features. HR = heart rate.

β = slope; regression coefficient between HR and time	P = peak heart rate of the main cycle
P_{mean} = average peak heart rate of all the cycles	P_{std} = standard deviation of peak heart rates of all the cycles
H = height	W = width
C_{std} = standard deviation of HR in the main cycle	C_{mean} = average HR of the main cycle
V = total variation of HR	$\theta = tan^{-1}(\frac{H}{W})$
$C - C_{mean}$ = average distance of P_is where i is ith cycle	$C - C_{std}$ = average standard deviation of distance between P_is

of cycles and the shape of the cycles are different each time. For this reason, we focus on what we termed the main cycle, which is affected the most by food intake, to extract the features for the model building.

Table 5.2 shows our defined structural features. We think that a physical response to food consumption can be explained by how fast one's heart rate reaches its highest value (P, W) and how different the heart rates are between the initial and peak values (H). In addition, the slope (β), which can be represented as a regression coefficient, and the angle (θ) of heart rate increase can define the body's reaction to food consumption. We also consider how fast the person's heart rate becomes stable by analyzing the width, slope, and angle of when the heart rate decreases. Additionally, the average (C_{mean}) and standard deviation (C_{std}) of heart rate in the cycle can describe one's physical reaction to different foods given that the heavier a food is, the more the body responds. When there are multiple cycles during a moment of food consumption, which may mean the person consumes many kinds of foods simultaneously, we consider the relationships between the cycles through $C - C_{mean}$ and $C - C_{std}$.

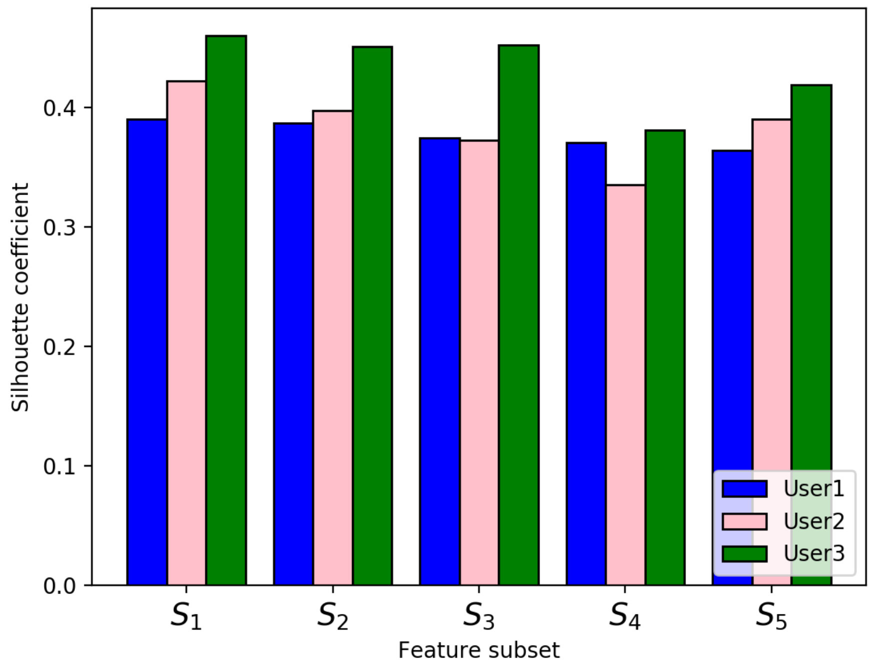


Figure 5.4: Silhouette coefficient on different feature subset. S_1=$\{P, C_{mean}, P_{mean}\}$, S_2=$\{P, C_{mean}, \theta_{down}, P_{mean}\}$, S_3= $\{W_{down}, P, P_{mean}\}$, S_4=$\{P, C_{mean}, \theta_{down}, P_{std}, P_{mean}\}$, S_5=$\{P, P_{std}, C_{mean}, P_{mean}\}$

5.4.2 Feature Selection

We observed from our labeled data sets that although there are foods with the same names, they can be very different across their food recipes and ingredients, and thus cause different physical responses. For that reason, we apply an unsupervised learning technique to cluster the series of heart rate values. To do this, considering that we cannot depend on the labels, we select the features by measuring the quality of a clustering structure.

We apply a partitioning technique via a graphical display, called silhouette, which measures how close each datum in one cluster is to other data in neighboring clusters. The silhouette coefficient for the selection of extracted heart rate features can be calculated as follows [95]:

$$s(i) = \frac{y(i) - x(i)}{max\{x(i), y(i)\}} \tag{5.1}$$

For each possible feature subset i, where $i = \{f_1, f_2, ..., f_n\}$, and n is the number of selected features, $x(i)$ is the average distance between i and all other feature subsets within the same cluster. $y(i)$ is the smallest average distance between i and all feature subsets in other clusters. The range of the silhouette coefficient is $[-1, 1]$. As the silhouette coefficient gets closer to $+1$, it shows the i lies well within its cluster, but a coefficient near -1 means that it has been assigned to the wrong cluster.

We try to select the features through this partitioning technique. We first make all the subsets with the extracted features in Section 5.4.1, and then run a clustering algorithm with each of them to see which feature combination returns the highest silhouette coefficient. Figure 5.4 shows the top-5 results obtained from three users' data. We can see from the results that P and P_{mean} positively affect the clustering result, and have a synergy effect when combined with C_{mean}. Therefore, among the subsets, we select S_1, which returns the highest silhouette coefficient of all.

5.4.3 Clustering

Spectral clustering is a graph partitioning technique to identify communities of nodes in a graph based on the edges connected to each other. A major advantage of spectral clustering is that it shows better performance than traditional clustering techniques, such as K-means, because it does not make assumptions on the form of the cluster [109]. For this reason, we apply this graph partitioning method on each person's heart rate data, and try to cluster the users' food eaten more efficiently than those of the convex clustering techniques.

To find the best number of clusters, K, we apply a modularity function, Q, developed by Newman and Girvan [82]. This function finds the optimal number of K by measuring the strength of division of a graph network into clusters. Thus, we choose the value of K that can maximize the modularity function Q [114]. The modularity function can be defined as follows [38]:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \tag{5.2}$$

where m is the number of edges, A is the affinity matrix, k_i is the degree of vertex i, and $\delta(c_i, c_j) = 1$ if i and j belonged to the same cluster, and otherwise $\delta(c_i, c_j) = 0$.

We first find the optimal number of K with the modularity function Q, and then run the spectral clustering with this K to cluster each of the sample sets into different clusters, $c_1, c_2, ..., c_k$. We will see to what extent similar kinds of foods according to the body responds can be grouped into a cluster in section 5.5 and lastly try to give a name to each of the clusters based on the experimental results.

5.5 Experimental Validation

In this section, we describe our experimental setting for clustering including experiment design and data collection. Then, we present the experimental results and discuss the effectiveness of the self-labeling technique by comparing it to another graph partitioning algorithm, Girvan Newman (GN), which is based on edge betweenness[1] of each vertex [82], and a convex clustering algorithm, K-means. Finally, we provide an event knowledge graph for the eating activity by relating each of the obtained data to the relevant event aspects.

[1] Number of shortest paths passing through the edge.

5.5.1 Experimental Setting

We recruited three participants who are highly motivated in food logging. They were males in their early 20's, early 30's, and late 60's who used Android smartphones, Huawei Mate 9, Samsung Galaxy S9 plus and Galaxy S8, respectively. We first asked each participant to install the Personicle application on his smartphone and then asked him to bring his phone with him as often as possible. We provided a heart rate tracker, an Fitbit Charge2 or Fitbit Blaze to each participant, and linked those devices to the Personicle application. We then encouraged the participants to make a food journal that contains at least food name and time of consumption through his preferred food logging tool. More importantly, we trained the participants that they have to start eating only after their heart rate becomes stable. This is because we had observed that people usually move before they start eating, such as walking towards the table, and this kind of movement can highly affect the change of heart rate. The length of the data collection period was 14 months, 11 months, and 6 months for each of the participants. Personicle data sets included heart rate, phone oriented sensor data and daily activity. For the heart rate data, we set the range for which data will be returned as 1 minute. Since the participants made food logs separately from Personicle, we matched up food logs with their Personicle data set through time stamps.

After making those data sets, we lastly filtered out the following noises for the self-labeling experiment. We expect that once our algorithm is trained, these noises will be handled by the system. One major noise was a sample set that starts from abnormally high heart rates. These kinds of patterns arose from either early moves in the eating activity or moves near food consumption. We also saw that there were many sample sets with no heart rate value because the participants did not wear the Fitbit when they had a meal. Furthermore, sometimes there were partial heart rate values because of system issues on either the Fitbit API or the Personicle application. From this preprocessing step, we received 270 sample sets for User 1, 49 sample sets for User 2 and 55 sample sets for User 3. The ratio of noise was

18%, 16.9%, and 9.8%, respectively.

5.5.2 Results and Evaluation

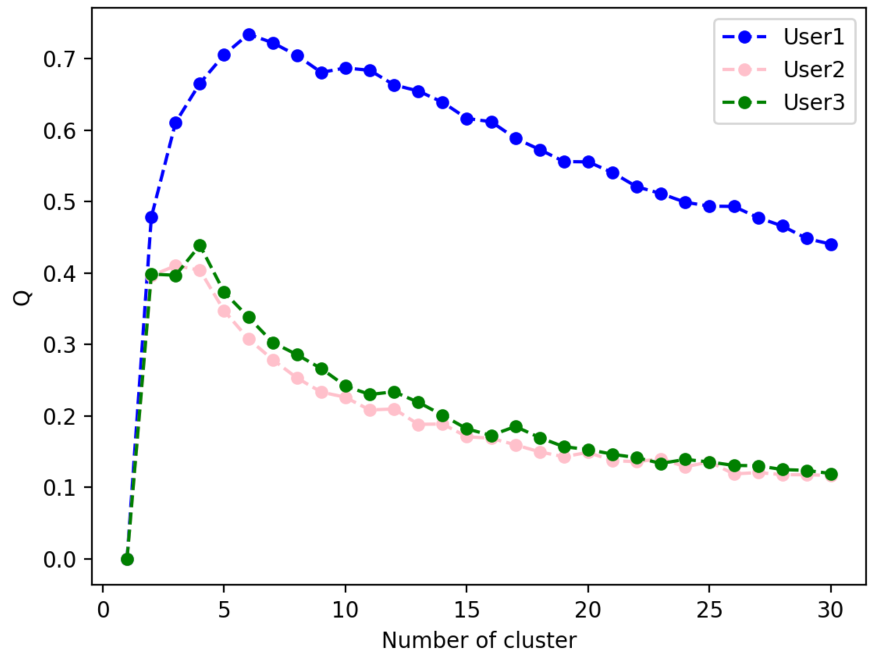
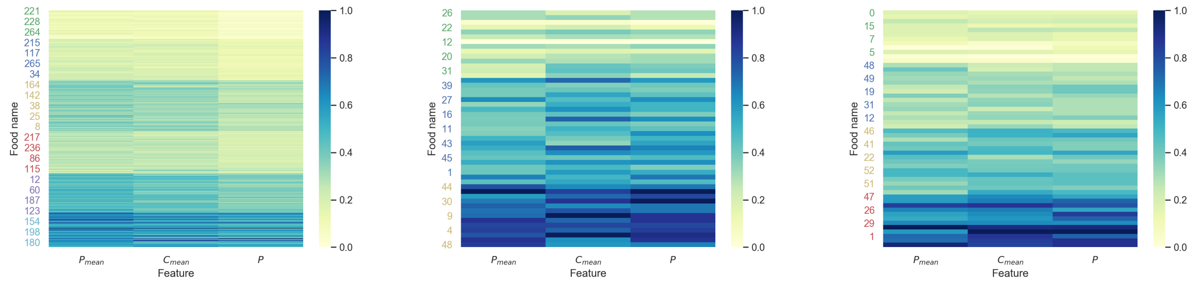


Figure 5.5: Q versus the number of clusters for the self-labeling experiment.

We first ran the modularity function Q to find the optimal number of clusters K. To do this, we created an undirected graph for each of the participants where vertices are foods consumed, and an edge is linked between vertices if they have relationships of P_{mean}, C_{mean}, and P. The graph of each participant contained 270 nodes, 49 nodes, and 55 nodes, respectively. We repeatedly ran spectral clustering algorithm on different numbers of K and tried to find when the highest modularity function Q could be obtained. Figure 5.5 shows how Q varies with the number of clusters on each user's graph. The peak for User 1, User 2, and User 3 are K = 6, Q = 0.7341, K = 3, Q = 0.4102, and K = 4, Q = 0.4387, respectively. We chose these results as the optimal k for each of the users. These results bring up an important point that each cluster should be labeled as a food group rather than an exact food name



(a) User 1's result (k = 6) (b) User 2's result (k = 3) (c) User 3's result (k = 4)

Figure 5.6: The heatmap of each user visualizing the measurements for a row over all the samples.

given that the k is a small number. The participants' food logs show that they often consume more than one food, such as pork with spicy sauce and brown rice, or a croissant sandwich with a cup of coffee. This supports the assumption that food consumption should be labeled as a food group in which each sample set considers the body's reactions to the sum of food consumed.

With the results obtained from Figure 5.5, we separately ran the spectral clustering algorithm with the optimal K (K = 6, K = 3, and K = 4 for User 1, User 2, and User 3) on the graph of each participant's sample sets. We then analyzed the clustering results by visualizing measurements for a subset of rows over all the samples. Figure 5.6 shows the heatmaps obtained from the clustering results. The column shows the features that we selected, and the row indicates the sample set labeled by the food name. Each number at the left corner stands for a sample set ID, and the color is for differentiating each cluster. The bar located at the right means that the darker the color the higher the feature value. We can see from the heatmap that User 1's sample sets were distinctly clustered into 6 groups, and User 2 and User 3's sample sets were clustered into 3 groups and 4 groups, respectively. These heatmap results bring up an another important point that each food group can stand for a level of heaviness since the visualization shows how much the food affects each participant's body. Based on this observation, we label each food group as a level of heaviness in the body where the higher the food group level the heavier in the body.

Table 5.3: Sample spectral clustering results obtained from User 1's data set.

Level 1	**Level 2**	**Level 3**	**Level 4**	**Level 5**	**Level 6**
sweet bread, oatmeal	stir-fried anchovy, spicy dumpling	pad thai	sushi	sweet garlic chicken	hot pot, beer
cold soba	poke (salmon and tuna)	chicken, rice	spicy ramen	chicken fried rice	pasta, wine
bagel	poke (salmon and tuna)	croissant sandwich	spicy ramen	spicy fried noodle	chips, sangria
oatmeal	strawberry smoothie	croissant sandwich	spicy ramen	spicy fried rice	chicken, beer
coffee, sweet potato	omelet, mushroom	croissant sandwich	sushi	beef, chicken, noodle	pork belly, beer

Table 5.4: Sample spectral clustering results obtained from User 2's data set.

Level 1	**Level 2**	**Level 3**
sandwich, coffee	seafood, salad, chips	spicy fried rice, noodle
sandwich, soda	dumpling, spicy chicken	seafood, wine
salad, sandwich	hamburger, fries, soda	spicy hot pot
sushi	salad, salmon, potato	steak, wine
soup, salad	curry, chicken, rice	spicy hot pot

Table 5.5: Sample spectral clustering results obtained from User 3's data set.

Level 1	**Level 2**	**Level 3**	**Level 4**
protein bar	muffin	beef rolls	boba tea
apple	donut	spicy hamburger	beef, ham, cheese
eggs	chicken, potato, rice, spinach	rice, hotdog, onion, egg	beef, chicken, pepper, rice
shrimp, broccoli	beef, carrot, onion, rice	chicken, pepper, onion, rice	steak, potato, burrito
banana	cake	pizza, cake	pork, rice

Finally, Table 5.3, 5.4 and 5.5 show five representative foods from each of the food level clusters. To evaluate these results, we compared them to those of two baselines, which are K-means and GN. For the K-means, we used elbow method, which draws values for K on the X axis and distortions on the Y axis, and chooses the K at an elbow, to find the optimal number of clusters. The higher the level number the heavier the food group is. The results are as follows:

User 1's result: As shown in Table 5.3, level 6 included all the alcoholic beverages, and level 1 and 2 mainly contain light foods, such as sweet bread, oatmeal and strawberry smoothie. We saw that similar kinds of foods like poke, croissant sandwich, sushi, and spicy ramen were clustered in the same level. The results obtained from baseline algorithms showed different outputs from those of spectral clustering. Here are some different results:

- `GN - Level 6`: chicken fried rice, sweet garlic chicken
- `GN - Level 4`: beef-chicken-noodle
- `K-means - Level 1`: croissant sandwich, sushi
- `K-means - Level 2`: french toast, bagel-banana, steak, pork

The results obtained from GN showed that level 6 did not only include alcoholic beverages, but also other heavy foods, which were originally clustered in level 5 at spectral clustering. Furthermore, most of the foods, which were clustered in level 5 at spectral clustering, were clustered in level 4 at GN. The results obtained from K-means (K = 3) had another outcome to the aforementioned algorithms. The level 1 contained some foods, which were clustered in level 3 and 4 at spectral clustering, and many different kinds of foods were clustered all together in the same level 2.

User 2's result: Table 5.4 shows that level 3 contained alcoholic beverages, spicy, and oily foods. Then, levels 1 and 2 split the rest of the foods according to their heaviness in the user's body. However, the baseline algorithms show different results. Here are some selected results from GN and K-means:

- `GN - Level 3`: hamburger-fries-soda, curry-chicken-rice
- `GN - Level 1`: seafood-salad-chips, salad-salmon-potato
- `K-means - Level 2`: soup-salad
- `K-means - Level 3`: curry-chicken-rice

The results obtained from GN showed that there was only one food in level 2. All the other foods, which were clustered in level 2 at spectral clustering, were clustered in either level 1 or level 3. In addition, the results obtained from K-means (K=3) showed that there were blurred boundaries between levels.

User 3's result: Table 5.5 indicates that the more the combination of heavy foods the higher the level, but his baseline results are different from spectral clustering. Here are some results obtained from GN and K-means:

- `GN - Level 4`: pizza-cake, chicken-pepper-onion-rice
- `GN - Level 1`: cake, beef-carrot-onion-rice
- `K-means - Level 2`: shrimp-broccoli, beef rolls

The results for GN showed that some of foods, which were clustered in level 3 at spectral clustering, were clustered in level 4. Furthermore, a couple of foods, which were clustered in level 2 at spectral clustering, were clustered in level 1. The K-means (K=4) results showed that many of foods, which were clustered in either level 1 or level 3 at spectral clustering, were clustered in level 2.

The above mentioned results show that the spectral clustering outperforms the other two baselines in terms of clustering the level of heaviness. The spectral clustering and the GN that we designed for this experiment start from the same graph construction process using a k-nearest neighbor graph, and then find the optimal number of K by using the same modularity function Q. The different process is that when finding the cluster in the graph, the GN starts with the full graph and then gradually removes the edges with the highest edge betweenness score up to find the K number of clusters. On the other hand, spectral clustering runs K-means algorithm in the end by using the eigenvectors obtained from its Laplacian matrix as features. We can predict from the results that embedding the vertices of a graph into a low-dimensional space through the eigenvectors works more than the edge betweenness of the GN in our data set. In addition, the baseline results using K-means show that the simple convex clustering based on Euclidean distance between each datum cannot properly cluster our sample sets, which are located in close proximity to each other.

Lastly, we built an event knowledge graph for eating by relating all the data we obtained to the daily activity and naming the relationships based on their event aspects. Figure 5.7 shows the descriptive interpretation of one of the events for eating obtained from User 1's data set. We can see from the event that the user ate heavy foods (Level 6) while talking with his wife at an Italian restaurant in the evening and this event lasted for 2 hours in celebrating their anniversary. We can also see from the stress level information that he was relaxed during the event.

5.6 Using Event Knowledge Graph

The event knowledge graph is an effective approach for understanding unique daily experiences. It captures and represents specific daily events with personalized semantic information. These graphs for specific events such as meetings, entertainment, family time, spiritual

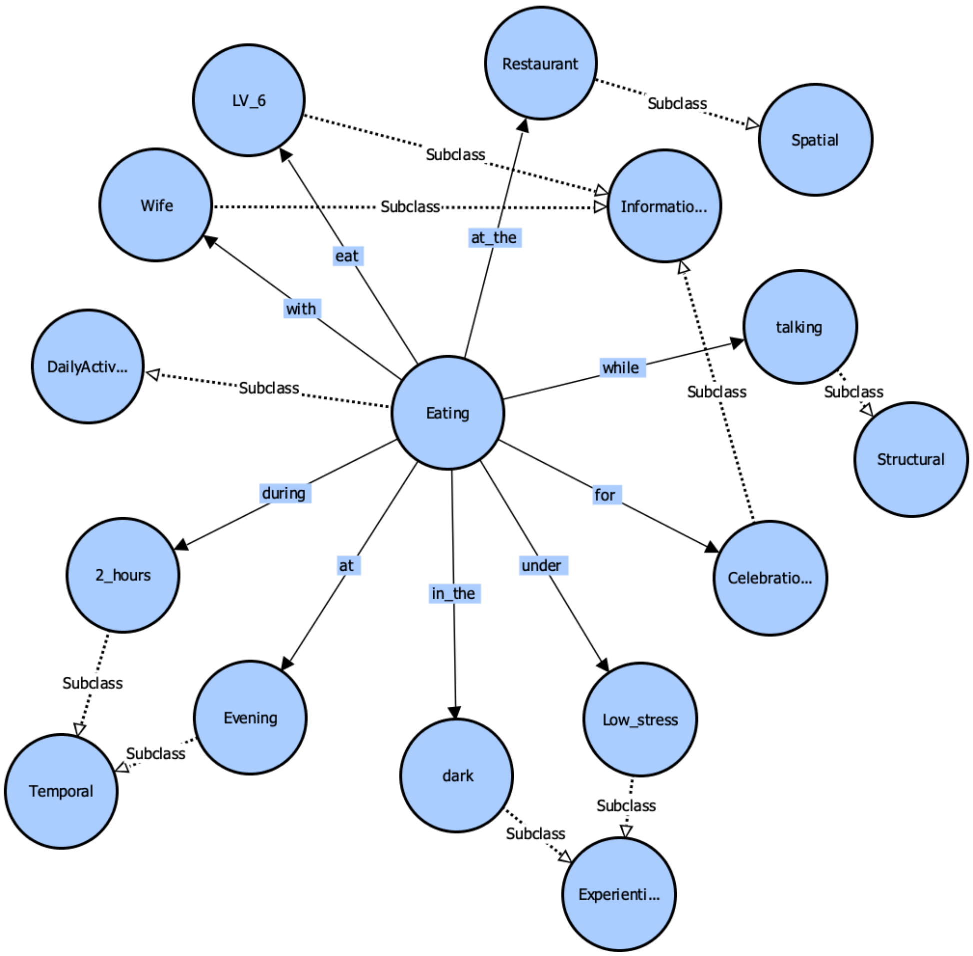


Figure 5.7: A visualization of the basic event knowledge graph for eating.

events, may be prepared and used to build a detailed event chronicle for a person. Analysis of such a chronological history of the enhanced daily events could reveal many latent correlations between lifestyle and health states of the person through causal analysis using event mining. Once the causal aspect can be unobtrusively obtained and related to the event, it will be possible to systematically find the reasons why a person behaves as he/she does in a specific situation. In addition, it is possible to find similar kinds of people through their events analysis and obtain better insights about current states of each individual. This could be used for understanding social behaviors of people. Furthermore, healthcare is one of the most natural applicable areas. For example, if the glucose level of a diabetic person suddenly shoots up, it can find which event and its aspects are involved in the changes. It can also find other diabetic patients who have similar culture, occupation, age, and lifestyle as well as having the similar symptoms to him, and therefore retrieve their solutions that had a positive effect on their health conditions. These findings could finally lead to a contextual recommendation system in the perpetual cycle of life events monitoring.

As such, event knowledge graphs provide a powerful representation and analysis approach to collect and aggregate contextually actionable information by organizing all knowledge sources related to a specific type of events. Like knowledge graphs in search engines, this helps to find relevant information to specific events that may be essential in particular applications. In this thesis, we demonstrated this in the context of dining events. Extensions to other types of events will require similar analysis and approaches to extract relevant facets of that event. The event knowledge graph could also be enlarged and updated according to the person's reactions to specific events.

5.7 Conclusion

Since daily experiences are closely tied to events, qualitatively and quantitatively recognizing daily events and their attributes is essential for analyzing current states of a person. This paper builds towards the emerging research area in medical and related disciplines to develop N-of-1 trials considering that each individual is unique. With this personalized trial deign, we tried to obtain each person's events of daily living and their attributes in unobtrusive ways through mobile devices. Specifically, this paper concentrates on finding latent semantic information from commonly used multimedia data and building an event knowledge graph that recognizes daily events through semantically enriching a low-level activity. It describes the methodology behind daily event recognition by showing a concrete example of the process in a dining event enrichment. We developed a self-labeling method of food consumption that focuses only on a physical response with the goal to unobtrusively obtain an important missing semantic context. Results obtained from the three subjects validate the potential of such an approach for recognizing daily events. Such enriched daily experiences could play a very important role in building a model of the person reflecting the dynamics of his reactions under specific conditions. Follow-up research would allow for enlarging the event knowledge graph by finding and relating more semantic event aspects, and thus revealing more daily events in each individual's daily life.

Chapter 6

Semantic Enrichment of Working Activities

In this chapter, we show another example of complementing the lack of the informational aspect for an event. This time, we attempt to recognize further specific information about the "working" activity by analyzing each individual's Personicle dataset. We first aim to diversify the type of "working" and then try to infer the users' job by clustering the types of their "working".

6.1 Diversifying the Type of Working

To semantically enrich a "working" activity, we first attempt to diversify the type of "working". With the method described in Chapter 3, we have only recognized one type of "working", which happens at the main workplace. Our previous method focuses on finding the main workplace to recognize the "working" activity and thus leads to a problem endemic that other types of working, such as working away from the office, cannot be recognized. Figure

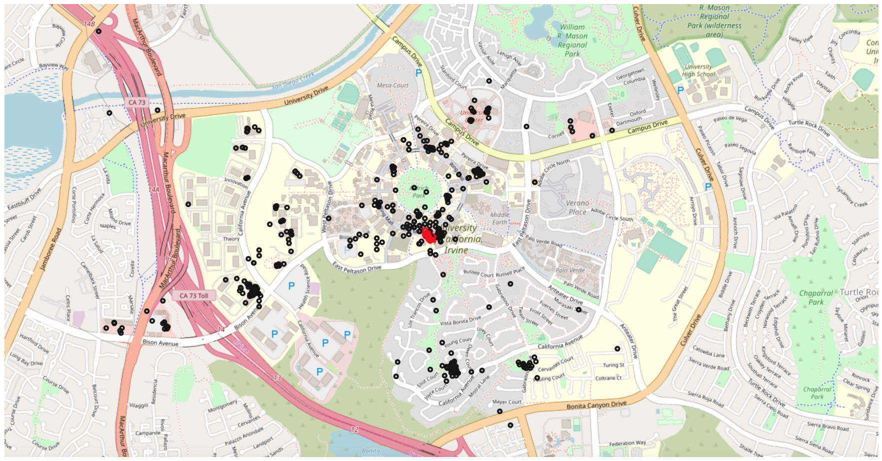


Figure 6.1: An actual example showing the limitation on the current recognition method that cannot recognize "working" if the user does not stay at the main workplace. Each dot represents a daily activity. The red dots are "working" and others are "unknown".

(a) Clustering result using K-means, Mean shift, Expectation Maximization (EM) and Hierarchical clustering.

(b) Clustering result using DBSCAN (Density-Based Spatial Clustering of Applications with Noise).

Figure 6.2: A comparison of different clustering algorithms in Figure 6.1's GPS dataset.

6.1 shows an actual example obtained from one of our Personicle system users. We drew all the daily activities, which happened during the most common working hours (8:00 am - 6:00 pm), as the colored dots on the map to understand the user's activities during the time range. We can notify from Figure 6.1 that there might be other types of "working" since all the black dots, which are currently "unknown" activities, also happened during the regular working hours. In this section, we try to reveal what these "unknown" activities can stand for.

We first make the best use of our commonsense to analyze the black dots on the map. Looking back through the days we work, we usually spend most of our time at the main workplace. Also, we sometimes have a meeting at other locations, such as the customer's office or cafe and may visit the same places multiple times for the work. This simple commonsense can indicate that the type of "working" could be separated mostly by the type of location. Therefore, we apply a clustering technique to classify each "unknown" activity into similar spatial groups. We then try to assign a spatial role to each of the spatial groups and then regard the type of "unknown" activities as their cluster's spatial role. To select a correct clustering algorithm to solve our problem, we train the different models using different clustering algorithms and compare the results to one another. In this section, we show the process by using one of our users' spatial data (e.g., latitude and longitude), which is displayed in Figure 6.1.

The baseline results obtained from K-means, Mean shift, EM and Hierarchical clustering show that these are not an ideal algorithm to cluster the latitude and longitude data since these algorithms not seem like reflect the geodetic distance as shown in Figure 6.2a. Although we tuned each of the parameters, such as the number of clusters, bandwidth, or the number of components, it still seemed that there is substantial distortion at latitudes. On the other hand, as shown in Figure 6.2b, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), which finds arbitrarily sized and shaped clusters based on the spatial density of data [32], showed a positive result that can be utilized for our clustering problem. It does not even require to specify the number of clusters to be generated, which is a frequent problem in data clustering and is a distinct issue from the process of actually solving the clustering problem. Based on this experiment, to classify each "unknown" activity into similar spatial groups, we use DBSCAN as our clustering algorithm and attempt to optimize its parameters.

DBSCAN clusters a spatial dataset based on two parameters: a distance between two points (Epsilon), and the minimum number of neighbors a given point should have (minPts). For the minPts, Ester et al. suggest to use the default value of minPts=4 for two-dimensional

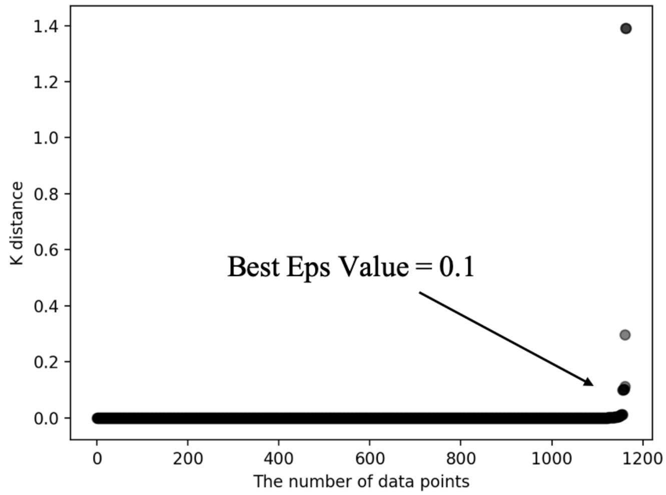
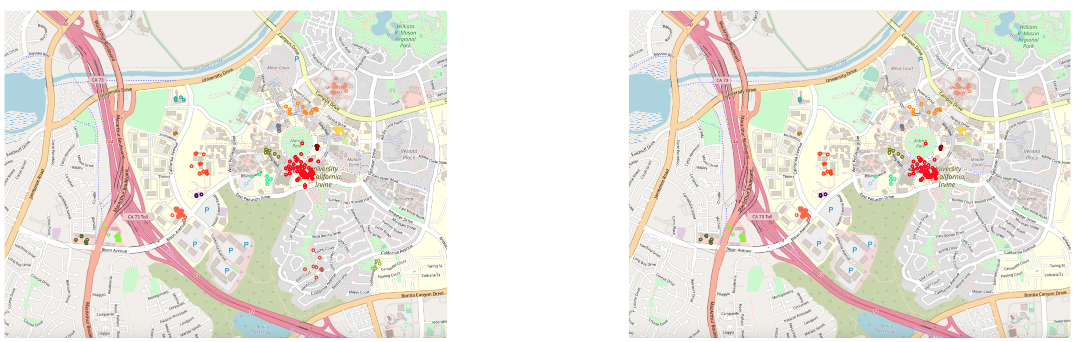


Figure 6.3: Points sorted by distance to the 4th nearest neighbor.



(a) Result using minPts=4, Epsilon=0.1.

(b) Result using minPts=4, Epsilon=0.1, and kernel density estimation.

Figure 6.4: Optimized clustering results using DBSCAN.

data [32] and minPts=2∗dimension for more than two-dimensional data [96]. We use the former value (minPts=4) given that our data set is consist of a pair of latitude and longitude and then determine Epsilon with the parameter by using the "elbow" method described in [32]. For using the "elbow" method, we plot the data points sorted by distance to the 4th nearest neighbor (minPts=4) as shown in Figure 6.3 and then choose the best Epsilon value at the "elbow" point, which is 0.1 in the Figure 6.1's dataset. To compare the optimized result to the baseline result (Figure 6.2b), we ran DBSCAN again with these newly found parameters (minPts=4, Epsilon=0.1) and then obtained the results as shown in Figure 6.4a.

Each color in Figure 6.4a can stand for different spatial roles given that it reflects the density of the user's spatial data during the regular working hours. Thus, we first assign a role of "main workplace" to the red cluster considering that the user spends most of his working hours in this area. After that, we assign a role of either "eating/working" or "outside work" to the rest of the clusters based on their venue type. More specifically, we assign the role of "eating/working", which might be something like "lunch meeting" or "meeting at cafe", if the venue type of the cluster belongs to cafe, bakery or restaurant and then assign the role of "outside work" to the rest of the clusters. Also, we attempt to enrich a "working" activity with calendar data, such as "eating/working; lunch meeting with Raj, Ping and Ramesh on Diabetes", if the user has the information in his/her calendar application. Finally, to further optimize the given results, we try to filter out noise from each of the clusters. From a commonsense point of view, if an activity only lasted for a short period of time (e.g., 5 minutes or 10 minutes) compared to those with other activities (e.g., 1 hour) in the same cluster and plus the relative frequency of the short-duration activity is even negligible, we could consider it as noise. To differentiate this noise from each of the spatial cluster, we use a 1-dimensional clustering method by applying a Kernel Density Estimation (KDE) technique and therefore try to cluster the activities in terms of the length of their duration. The KDE technique can help identify the density of a distribution of the activity duration since it finds where a lot of data is grouped together and where it isn't even in 1-dimensional data [100]. 1D clustering with KDE can be done in 5 steps:

1. Normalize data
2. Compute densities
3. Find local maxima
4. Find local minima
5. Cluster the data at each of the local minima

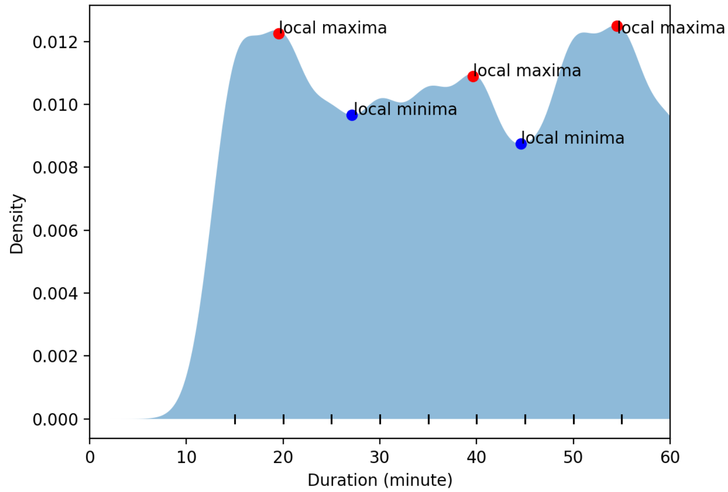


Figure 6.5: The kernel density estimation for the one-dimensional duration of the "main workplace" cluster.

We first normalize the duration of the activities in each spatial cluster and then compute their densities by using kernel density estimator:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h}) \tag{6.1}$$

where $(x_1, x_2, ..., x_n)$ are the duration of each activity, K is the kernel (e.g., Gaussian kernel) and $h > 0$ is a smoothing parameter called bandwidth. Figure 6.5 shows the kernel density estimation for the activity-duration data of the "main workplace" cluster. To cluster this data, we mark the local maxima and local minima on the graph and then cluster the data at each of the local minima. We then consider the lower bound, such as activities having a duration between 5 minutes and 25 minutes in Figure 6.5, as noise and remove them from the cluster. After applying all the aforementioned process, we finally obtain the end result as shown in Figure 6.4b.

Table 6.1: Personicle users' data description.

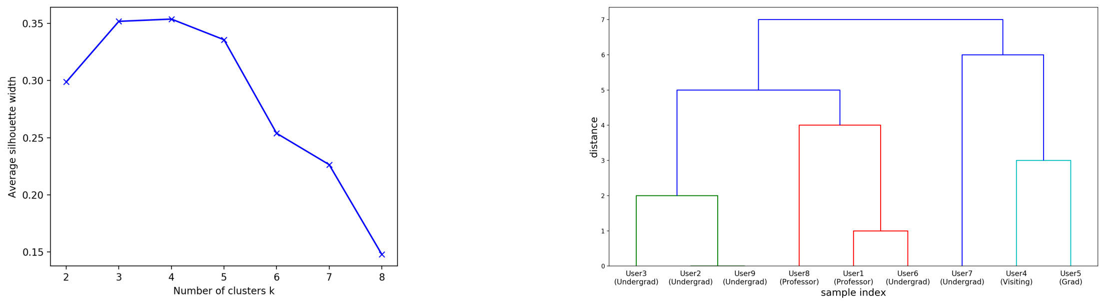
User ID	Job	Data Collection Period	# of "Working" samples
User1	Professor	212 days	4,075
User2	Undergrad	99 days	1,165
User3	Undergrad	10 days	191
User4	Visiting	20 days	224
User5	Grad	257 days	3,340
User6	Undergrad	20 days	236
User7	Undergrad	158 days	2,115
User8	Professor	145 days	1,320
User9	Undergrad	15 days	131

Table 6.2: Personicle users' average time usage (%) in each type of the "Working".

User	Job	Main Workplace	Outside Work	Eating/Working	Home
User1	Professor	47.84%	12.47%	6.7%	33%
User2	Undergrad	26.38%	3.8%	4.85%	65%
User3	Undergrad	10.5%	6.12%	0%	83.37%
User4	Visiting	90.87%	4.21%	0%	4.9%
User5	Grad	61.55%	0.4%	3.33%	34.7%
User6	Undergrad	26.6%	10.43%	8.12%	54.83%
User7	Undergrad	40.91%	1.72%	19.19%	38.17%
User8	Professor	15.61%	17.33%	2.74%	64.3%
User9	Undergrad	18.63%	4.4%	5.84%	71.16%

6.2 Clustering the Sort of Jobs

With the results obtained in section 6.1, we now attempt to infer the users' job so that we can enrich the "Working" activity further. To do this, we first hypothesize that the users' job would be analogous if their "Working" pattern is similar to one another. We then try to catch the pattern by analyzing the types of their "Working", which are the semantics obtained from section 6.1. More specifically, we attempt to cluster each individual into similar occupation groups by using the types of their "Working". Table 6.1 describes our experiment dataset including the users, their jobs, their data collection periods and the number of their "Working' samples. We collected these data from professors, visiting scholars, graduate students and



(a) Choosing optimal number of clusters with an average silhouette method.

(b) Dendrogram.

Figure 6.6: Optimized results using Hierarchical clustering.

undergraduate students who are at the University of California, Irvine (UCI) and University of Turku (UTU). In this section, we try to identify whether each of these occupations has a distinct "Working" pattern to be separated and thus see to what extent similar kinds of occupations can be clustered together. The data collection periods vary from 10 days to 257 days.

By using this dataset, we first obtain the types of each users' "Working" (e.g., "main workplace", "eating/working", and "outside work") and then calculate how much time each user spends for each type of the "Working" everyday. Table 6.2 shows the users' average time usage during their normal working hours (8:00 am - 6:00 pm). These features and values are used to cluster each individual into similar occupation groups. We apply a Hierarchical clustering method, which is generally applicable to a small set of data [6], given that we only try to cluster 9 samples (users). To determine the optimal number of occupation groups in our dataset, we apply an average silhouette method, which measures the quality of clustering by determining how well each data point lies within its cluster. This method finds a value of k that maximizes the average silhouette over a range of possible values for k [58]. The silhouette coefficient of each data point can be calculated as follows:

$$s(i) = \frac{b(i) - a(i)}{max(a(i), b(i))} \tag{6.2}$$

where $a(i)$ is a intra-cluster distance of i_{th} data and $b(i)$ is a nearest-cluster distance of i_{th} data. Figure 6.6a shows the plot that draws the average silhouette coefficient over different values of k. This plot indicates that k=4 could be the optimal number of clusters. With this parameter, we ran the Hierarchical clustering and obtained a dendrogram in which we represented each cluster as different colors as shown in Figure 6.6b. We can see from the dendrogram that most of the undergraduate students are grouped in the same cluster. This result means that the undergraduate students have different "Working" patterns from other occupation groups. Also, the visiting scholar and graduate student are grouped in the light-blue cluster. Considering that these two jobs are based on research and thus require spending most of their working time at the "main workplace", this clustering result can be explained on the similarity of the "Working" pattern. Lastly, the professors are grouped in the red cluster. Their average time usage during the working hours (Table 6.2) shows that they also spend some of their working time for "eating/working" and "outside work". This pattern reflects the nature of their job and the way that they work, such as researching at the main office, giving a lecture at classrooms and meeting people at other places (e.g., colleagues' office, company or cafe). User6 is also grouped in the red cluster. This might be explained by the fact that the students in undergraduate school usually take their classes at multiple buildings and study at a cafe or cafeteria, which is a similar spatial pattern to those of the professors' group. This result could be improved once we collect more of User6's spatial data given that his dataset is only 20 days worth of data.

Finally, we compare the result to those of other clustering algorithms as shown in Table 6.3. The value of each k was determined by each algorithm's parameter tuning process. In this comparison, since we only handle a very small number of samples, we focus more on seeing to what extent similar kinds of occupations can be grouped by different clustering techniques rather than evaluating the model itself. Viewed in this way, spectral clustering does not cluster similar jobs but rather to divide all the users separately. EM clustering only clusters the users into either Group 0 or Group 1, which is too simple. On the other hand, as

Table 6.3: A comparison of the results using different clustering algorithms.

User	**Job**	**Hierarchical (k=4)**	**K-means (k=4)**	**Mean-shift (k=5)**	**EM (k=2)**	**Spectral (k=6)**
User1	Professor	Group 0	Group 2	Group 3	Group 0	Group 1
User2	Undergrad	Group 3	Group 1	Group 0	Group 1	Group 4
User3	Undergrad	Group 3	Group 1	Group 0	Group 1	Group 0
User4	Visiting	Group 1	Group 0	Group 1	Group 0	Group 3
User5	Grad	Group 1	Group 0	Group 2	Group 0	Group 1
User6	Undergrad	Group 0	Group 2	Group 0	Group 1	Group 2
User7	Undergrad	Group 2	Group 2	Group 4	Group 0	Group 1
User8	Professor	Group 0	Group 1	Group 0	Group 1	Group 0
User9	Undergrad	Group 3	Group 1	Group 0	Group 1	Group 5

with the hierarchical clustering, the results of K-means and mean-shift clustering show that some of the occupation groups, such as the undergraduate student group, can be correctly clustered together. From these results, we can see the potential of using the pattern of each individual's "Working" to cluster the users into similar occupation groups. We could then use this semantic to enrich the "Working" activity further and thus have more chances to understand each individual's life experience in more detail.

Chapter 7

Conclusion and Future Work

Continuously monitoring a person's everyday life would provide a significant amount of valuable data to build personal models. Although the latest sensor technology would have enabled people to track parts of their own daily lives, much of the collectible data is scattered in isolated silos with different granularities and semantics, and thus the semantic gap becomes even more formidable. This poses a primary challenge to multimedia research: there needs to be an efficient way of bridging the semantic gap between the low-level multimedia logs and high-level events. Thus, in this thesis, we designed a chronicle of personal data, called Personicle, in which we aggregated, integrated, and synchronized different data streams to make sense of all the multi-modal correlated information.

This thesis focused on introducing how we can build the Personicle unobtrusively and accessibly by using the common device combination of a smartphone and wearable fitness device. To do this, we first defined an atomic interval to bring multi-modal data streams into the context of a daily life space and to abstractify the processing of a person's everyday life. We then developed a method to segment a day into similar patterns of atomic intervals.These segments were then used for the recognition of daily activities which provide much more meaningful human-related information than those of low-level streams of data (Chapter 3). After that, to build an enriched chronicle of the daily activities, we designed the event-triggered Ecological Momentary Assessment in which we maximize the chance of aggregating the semantic data. To do this, we trigger the EMA process at the right moments with the proper medium, such as a voice command logger (Chapter 4). We went on to suggest an unobtrusive approach to obtain latent semantics from heterogeneous signals, and thus an approach for enhancing events of daily living based on semantic context enrichment. With the enhancement method, we finally organized the chronicle of events of daily living and then built the Personicle with their various descriptive attributes (Chapter 5). We expect that such an enriched daily experience chronicle could play a very important role in building a model of a person which reflects their personal dynamics and tendencies under specific conditions.

In this thesis, we concentrated on completing the implementation of Personicle in order to introduce what Personicle is. Thus, there are many opportunities for improvement and future research.

- **Performance improvement**: In the current Personicle, we have been using a 5-minute long atomic-interval to recognize a person's daily activities. However, the optimal length of the atomic-interval might be different from activity or person. Therefore, research on personalizing the length of the atomic-interval could be important work in improving recognition performance. Furthermore, given that we have been using a common daily activity model, there needs to be reinforcement learning performed towards the personalization of each daily activity recognition. We also need an experimental validation of the food consumption methodology with more active users as well as studying how to exclude the impact of movement on heart rate patterns so that we can effectively apply the method in the Personicle.
- **Event graph**: The event knowledge graph could be an effective approach to under-

standing unique daily experiences. It may capture and represent specific daily events with personalized semantic information. These graphs for specific events such as meetings, entertainment, family time, and spiritual events may be prepared and used to build a detailed event chronicle for a person. Analysis of such a chronological history of the enhanced daily events could even reveal many latent correlations between the lifestyle and health status of the person through causal analysis using event mining. Once the causal aspect can be unobtrusively obtained and related to the event, it will be possible to systematically find the reasons why a person behaves as they do in a specific situation.

- **Finding more latent semantics**: To keep enriching the events of daily living, it is required to recognize high-level semantics for each of the events. In this thesis, we only focused on enriching “Eating” and “Working” activities by introducing how to unobtrusively recognize food consumption type and diversifying the working event by inferring the user’s job. Much more work must be done for finding latent semantics so that we can recognize more diverse events of daily living. For example, the analysis of ambient sound could be used to infer important semantic information, such as talking, watching TV, or even the emotion of the user, which can ultimately lead to recognizing more complicated events like “Socializing” or “Relaxing”.

- **Building a Personal Health Navigation (PHN)**: The goal-based guidance system, such as a navigation system, continuously estimates the current state, finds the best way to achieve the desired goal, and guides actions to make it happen. With the Personicle, the PHN could be designed in a way that perpetually collects measurements about the person’s health state. The PHN would then help an individual maintain his or her desired health state by providing situationally actionable and easy to follow guidelines. All the person’s following actions are continuously measured to provide a new estimation of his or her health state. If there is a change in the health state, it will be reflected in the next guidance. A cycle of these processes will move the person's health state closer to the goal.

- **Building personal models**: To understand the preferences and particularities of an individual, it is necessary to build personal models. The personal model could represent a personalized knowledge base of how a person reacts to different stimuli under specific conditions or how their physiology changes from an intervention. The primary consideration in building the personal model, which establishes the premise that each person is a unique entity, could be how various factors uniquely drive each individual's personal state. To extract this personalized knowledge, we would need to aggregate long and short-term information about a person and analyze how this data interacts with each other. Then, the models could be used to provide actionable guidance.

- **Event mining**: Since we started obtaining the events of daily living, we now need to combine the events and derive actionable causal relationships from the events. With the Personicle, we could start finding how different aspects of our lives are interconnected and how they are influencing our everyday life. Mining relevant patterns from the chronicle of daily events could discover causal relationships among the events, their attributes, and behaviors. Such research would allow us to derive rich insights about the person and ultimately result in building even more effective personal models.

Bibliography

[1] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles. Towards a better understanding of context and context-awareness. In *International Symposium on Handheld and Ubiquitous Computing*, pages 304–307. Springer, 1999.

[2] K. Aizawa, T. Hori, S. Kawasaki, and T. Ishikawa. Capture and efficient retrieval of life log. In *Pervasive 2004 workshop on memory and sharing experiences*, pages 15–20, 2004.

[3] K. Aizawa, K. Ishijima, and M. Shiina. Summarizing wearable video. In *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, volume 3, pages 398–401. IEEE, 2001.

[4] K. Aizawa, K. Maeda, M. Ogawa, Y. Sato, M. Kasamatsu, K. Waki, and H. Takimoto. Comparative study of the routine daily usability of foodlog: A smartphone-based food recording tool assisted by image retrieval. *Journal of diabetes science and technology*, 8(2):203–208, 2014.

[5] K. Aizawa and M. Ogawa. Foodlog: Multimedia tool for healthcare applications. *IEEE MultiMedia*, 22(2):4–8, 2015.

[6] O. Akman, T. Comar, D. Hrozencik, and J. Gonzales. Data clustering and self-organizing maps in biology. In *Algebraic and Combinatorial Computational Biology*, pages 351–374. Elsevier, 2019.

[7] R. Albatal, C. Gurrin, J. Zhou, Y. Yang, D. Carthy, and N. Li. Senseseer mobile-cloud-based lifelogging framework. In *Technology and Society (ISTAS), 2013 IEEE International Symposium on*, pages 144–146. IEEE, 2013.

[8] S. Alsubaiee, Y. Altowim, H. Altwaijry, A. Behm, V. Borkar, Y. Bu, M. Carey, I. Cetindil, M. Cheelangi, K. Faraaz, et al. Asterixdb: A scalable, open source bdms. *Proceedings of the VLDB Endowment*, 7(14):1905–1916, 2014.

[9] O. Amft and G. Troster. On-body sensing solutions for automatic dietary monitoring. *IEEE pervasive computing*, 8(2), 2009.

[10] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *International Conference on Pervasive Computing*, pages 1–17. Springer, 2004.

[11] O. Beijbom, N. Joshi, D. Morris, S. Saponas, and S. Khullar. Menu-match: Restaurant-specific food logging from images. In *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, pages 844–851. IEEE, 2015.

[12] C. G. Bell and J. Gemmell. *Total recall: How the e-memory revolution will change everything*. Dutton, 2009.

[13] G. Bell and J. Gemmell. A digital life. *Scientific American*, 296(3):58–65, 2007.

[14] G. Biegel and V. Cahill. A framework for developing mobile, context-aware applications. In *Pervasive Computing and Communications, 2004. PerCom 2004. Proceedings of the Second IEEE Annual Conference on*, pages 361–365. IEEE, 2004.

[15] L. Bossard, M. Guillaumin, and L. Van Gool. Food-101–mining discriminative components with random forests. In *European Conference on Computer Vision*, pages 446–461. Springer, 2014.

[16] A. Briassouli, J. Benois-Pineau, and A. Hauptmann. Overview of multimedia in healthcare. In *Health Monitoring and Personalized Feedback using Multimedia Data*, pages 1–6. Springer, 2015.

[17] B. S. Burke et al. The dietary history as a tool in research. *Journal of the American Dietetic Association*, 23:1041–1046, 1947.

[18] L. E. Burke, S. Shiffman, E. Music, M. A. Styn, A. Kriska, A. Smailagic, D. Siewiorek, L. J. Ewing, E. Chasens, B. French, et al. Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. *Journal of medical Internet research*, 19(3):e77, 2017.

[19] V. Bush et al. As we may think. *The atlantic monthly*, 176(1):101–108, 1945.

[20] S. Carter, J. Mankoff, and J. Heer. Momento: support for situated ubicomp experimentation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 125–134. ACM, 2007.

[21] Y.-W. Chang, C.-J. Hsieh, K.-W. Chang, M. Ringgaard, and C.-J. Lin. Training and testing low-degree polynomial data mappings via linear svm. *Journal of Machine Learning Research*, 11(Apr):1471–1490, 2010.

[22] L. Chen, C. D. Nugent, and H. Wang. A knowledge-driven approach to activity recognition in smart homes. *IEEE Transactions on Knowledge and Data Engineering*, 24(6):961–974, 2012.

[23] J. Cheng, B. Zhou, K. Kunze, C. C. Rheinländer, S. Wille, N. Wehn, J. Weppner, and P. Lukowicz. Activity recognition and nutrition monitoring in every day situations with a textile capacitive neckband. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 155–158. ACM, 2013.

[24] E. K. Choe, S. Abdullah, M. Rabbi, E. Thomaz, D. A. Epstein, F. Cordeiro, M. Kay, G. D. Abowd, T. Choudhury, J. Fogarty, et al. Semi-automated tracking: A balanced approach for self-monitoring applications. *IEEE Pervasive Computing*, 16(1):74–84, 2017.

[25] G. Ciocca, P. Napoletano, and R. Schettini. Food recognition: a new dataset, experiments, and results. *IEEE journal of biomedical and health informatics*, 21(3):588–598, 2017.

[26] F. Cordeiro, E. Bales, E. Cherry, and J. Fogarty. Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3207–3216. ACM, 2015.

[27] K. W. Davidson, Y. K. Cheung, T. McGinn, and Y. C. Wang. Expanding the role of n-of-1 trials in the precision medicine era: Action priorities and practical considerations. *NAM Perspectives*, 2018.

[28] A. K. Dey. Understanding and using context. *Personal and ubiquitous computing*, 5(1):4–7, 2001.

[29] A. R. Doherty, K. Pauly-Takacs, N. Caprani, C. Gurrin, C. J. Moulin, N. E. O'Connor, and A. F. Smeaton. Experiences of aiding autobiographical memory using the sensecam. *Human–Computer Interaction*, 27(1-2):151–174, 2012.

[30] A. R. Doherty and A. F. Smeaton. Automatically segmenting lifelog data into events. In *Image Analysis for Multimedia Interactive Services, 2008. WIAMIS'08. Ninth International Workshop on*, pages 20–23. IEEE, 2008.

[31] Y. Dong, J. Scisco, M. Wilson, E. Muth, and A. Hoover. Detecting periods of eating during free-living by tracking wrist motion. *IEEE journal of biomedical and health informatics*, 18(4):1253–1260, 2014.

[32] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.

[33] D. D. Farhud. Impact of lifestyle on health. *Iranian journal of public health*, 44(11):1442, 2015.

[34] A. Fathi, A. Farhadi, and J. M. Rehg. Understanding egocentric activities. *2011 International Conference on Computer Vision*, 2011.

[35] Y. Feng, C. K. Chang, and H. Ming. Recognizing activities of daily living to improve well-being. *IT Professional*, 19(3):31–37, 2017.

[36] D. Ferreira, V. Kostakos, and A. K. Dey. Aware: mobile context instrumentation framework. *Frontiers in ICT*, 2:6, 2015.

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

[38] S. Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

[39] J. Gemmell, G. Bell, and R. Lueder. Mylifebits: a personal database for everything. *Communications of the ACM*, 49(1):88–95, 2006.

[40] J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C. Wong. Mylifebits: fulfilling the memex vision. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 235–238. ACM, 2002.

[41] C. Gurrin, A. F. Smeaton, D. Byrne, N. O'Hare, G. J. Jones, and N. O'Connor. An examination of a large visual lifelog. In *Asia Information Retrieval Symposium*, pages 537–542. Springer, 2008.

[42] C. Gurrin, A. F. Smeaton, A. R. Doherty, et al. Lifelogging: Personal big data. *Foundations and Trends® in Information Retrieval*, 8(1):1–125, 2014.

[43] M. A. Hall. Correlation-based feature selection for machine learning. 1999.

[44] R. Helaoui, D. Riboni, and H. Stuckenschmidt. A probabilistic ontological framework for the recognition of multilevel human activities. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 345–354. ACM, 2013.

[45] K. Hnatkova, D. Kowalski, J. J. Keirns, E. M. van Gelderen, and M. Malik. Qtc changes after meal intake: Sex differences and correlates. *Journal of electrocardiology*, 47(6):856–862, 2014.

[46] J.-H. Hong, J. Ramos, and A. K. Dey. Toward personalized activity recognition systems with a semipopulation approach. *IEEE Transactions on Human-Machine Systems*, 46(1):101–112, 2016.

[47] T. Hori and K. Aizawa. Context-based video retrieval system for the life-log applications. In *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 31–38. ACM, 2003.

[48] S. Hotta, T. Mori, D. Uchida, K. Maeda, Y. Yaginuma, and A. Inomata. Eating moment recognition using heart rate responses. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 69–72. ACM, 2017.

[49] https://www.google.com/fit/. Google Fit.

[50] S. S. Intille, K. Larson, and C. Kukla. Just-in-time context-sensitive questioning for preventative health care. In *Proceedings of the AAAI 2002 Workshop on Automation as Caregiver: The Role of Intelligent Technology in Elder Care*. AAAI Press Menlo Park, CA, 2002.

[51] R. Jain and L. Jalali. Objective self. *IEEE MultiMedia*, 21(4):100–110, 2014.

[52] L. Jalali, D. Huo, H. Oh, M. Tang, S. Pongpaichet, and R. Jain. Personicle: personal chronicle of life events. In *Workshop on Personal Data Analytics in the Internet of Things (PDA@ IOT) at the 40th International Conference on Very Large Databases (VLDB), Hangzhou, China*, 2014.

[53] L. Jalali, H. Oh, R. Moazeni, and R. Jain. Human behavior analysis from smartphone data streams. In *International Workshop on Human Behavior Understanding*, pages 68–85. Springer, 2016.

[54] J. D. Johnston. Physiological responses to food intake throughout the day. *Nutrition research reviews*, 27(1):107–118, 2014.

[55] H. Junker, O. Amft, P. Lukowicz, and G. Tröster. Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recognition*, 41(6):2010–2024, 2008.

[56] H. Kagaya, K. Aizawa, and M. Ogawa. Food detection and recognition using convolutional neural network. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 1085–1088. ACM, 2014.

[57] D. Kahneman, A. B. Krueger, D. A. Schkade, N. Schwarz, and A. A. Stone. A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306(5702):1776–1780, 2004.

[58] L. Kaufman and P. J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons, 2009.

[59] Y. Kawano and K. Yanai. Foodcam: A real-time food recognition system on a smartphone. *Multimedia Tools and Applications*, 74(14):5263–5287, 2015.

[60] N. Kleitman. *Sleep and wakefulness*. University of Chicago Press, 1963.

[61] F. Kong and J. Tan. Dietcam: Automatic dietary assessment with mobile camera phones. *Pervasive and Mobile Computing*, 8(1):147–163, 2012.

[62] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.

[63] O. D. Lara and M. A. Labrador. A mobile platform for real-time human activity recognition. In *2012 IEEE consumer communications and networking conference (CCNC)*, pages 667–671. IEEE, 2012.

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

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

[66] H. Lee, A. F. Smeaton, N. E. OConnor, G. Jones, M. Blighe, D. Byrne, A. Doherty, and C. Gurrin. Constructing a sensecam visual diary as a media process. *Multimedia Systems*, 14(6):341–349, 2008.

[67] S. Lee, H. X. Le, H. Q. Ngo, H. I. Kim, M. Han, Y.-K. Lee, et al. Semi-markov conditional random fields for accelerometer-based activity recognition. *Applied Intelligence*, 35(2):226–241, 2011.

[68] J. Lin, E. Keogh, L. Wei, and S. Lonardi. Experiencing sax: a novel symbolic representation of time series. *Data Mining and knowledge discovery*, 15(2):107–144, 2007.

[69] M. Luštrek, B. Cvetković, V. Mirchevska, Ö. Kafalı, A. E. Romero, and K. Stathis. Recognising lifestyle activities of diabetic patients with a smartphone. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2015 9th International Conference on*, pages 317–324. IEEE, 2015.

[70] S. Mann. Wearable computing: A first step toward personal imaging. *Computer*, 30(2):25–32, 1997.

[71] S. Mann. Continuous lifelong capture of personal experience with eyetap. In *Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, pages 1–21. ACM, 2004.

[72] S. Mann, J. Fung, C. Aimone, A. Sehgal, and D. Chen. Designing eyetap digital eyeglasses for continuous lifelong capture and sharing of personal experiences, 2005.

[73] S. Mann, J. Huang, R. Janzen, R. Lo, V. Rampersad, A. Chen, and T. Doha. Blind navigation with a wearable range camera and vibrotactile helmet. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 1325–1328. ACM, 2011.

[74] S. Mann, R. C. H. Lo, K. Ovtcharov, S. Gu, D. Dai, C. Ngan, and T. Ai. Realtime hdr (high dynamic range) video for eyetap wearable computers, fpga-based seeing aids, and glasseyes (eyetaps). In *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–6. IEEE, 2012.

[75] J. W. Marr. Individual dietary surveys: purposes and methods. In *World review of nutrition and dietetics*, volume 13, pages 105–164. Karger Publishers, 1971.

[76] Y. Matsuda, H. Hoashi, and K. Yanai. Recognition of multiple-food images by detecting candidate regions. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, pages 25–30. IEEE, 2012.

[77] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE, 2006.

[78] G. Meditskos, P.-M. Plans, T. G. Stavropoulos, J. Benois-Pineau, V. Buso, and I. Kompatsiaris. Multi-modal activity recognition from egocentric vision, semantic enrichment and lifelogging applications for the care of dementia. *Journal of Visual Communication and Image Representation*, 51:169–190, 2018.

[79] S. Mittal, K. Gopal, and S. Maskara. A versatile lattice based model for situation recognition from dynamic ambient sensors. *International Journal on Smart Sensing and Intelligent Systems*, 6(1):403–432, 2013.

[80] M. E. Mlinac and M. C. Feng. Assessment of activities of daily living, self-care, and independence. *Archives of Clinical Neuropsychology*, 31(6):506–516, 2016.

[81] N. Nag and R. Jain. A navigational approach to health: Actionable guidance for improved quality of life. *Computer*, 52(4):12–20, 2019.

[82] M. E. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

[83] H. Oh and R. Jain. From multimedia logs to personal chronicles. In *Proceedings of the 2017 ACM on Multimedia Conference*, MM '17, pages 881–889, New York, NY, USA, 2017. ACM.

[84] H. Oh and R. Jain. From multimedia logs to personal chronicles. In *Proceedings of the 2017 ACM on Multimedia Conference*, pages 881–889. ACM, 2017.

[85] H. Oh and R. Jain. Detecting events of daily living using multimodal data. *arXiv preprint arXiv:1905.09402*, 2019.

[86] H. Oh, L. Jalali, and R. Jain. An intelligent notification system using context from real-time personal activity monitoring. In *Multimedia and Expo (ICME), 2015 IEEE International Conference on*, pages 1–6. IEEE, 2015.

[87] H. Oh, J. Nguyen, S. Soundararajan, and R. Jain. Multimodal food journaling. In *Proceedings of the 3rd International Workshop on Multimedia for Personal Health and Health Care*, pages 39–47. ACM, 2018.

[88] H. Pirsiavash and D. Ramanan. Detecting activities of daily living in first-person camera views. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[89] P. Pouladzadeh, S. Shirmohammadi, and R. Al-Maghrabi. Measuring calorie and nutrition from food image. *IEEE Transactions on Instrumentation and Measurement*, 63(8):1947–1956, 2014.

[90] P. E. Puddu and A. Menotti. The impact of basic lifestyle behaviour on health: how to lower the risk of coronary heart disease, other cardiovascular diseases, cancer and all-cause mortality. *Lifestyle adaptation: a global approach. Esc Council for Cardiology Practice*, 13:32, 2015.

[91] M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney. Recognition and volume estimation of food intake using a mobile device. In *Applications of Computer Vision (WACV), 2009 Workshop on*, pages 1–8. IEEE, 2009.

[92] T. Rahman, M. Czerwinski, R. Gilad-Bachrach, and P. Johns. Predicting about-to-eat moments for just-in-time eating intervention. In *Proceedings of the 6th International Conference on Digital Health Conference*, pages 141–150. ACM, 2016.

[93] C. Randell and H. Muller. Context awareness by analysing accelerometer data. In *Digest of Papers. Fourth International Symposium on Wearable Computers*, pages 175–176. IEEE, 2000.

[94] D. Riboni and C. Bettini. Owl 2 modeling and reasoning with complex human activities. *Pervasive and Mobile Computing*, 7(3):379–395, 2011.

[95] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[96] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data mining and knowledge discovery*, 2(2):169–194, 1998.

[97] K. A. Sauder, E. R. Johnston, A. C. Skulas-Ray, T. S. Campbell, and S. G. West. Effect of meal content on heart rate variability and cardiovascular reactivity to mental stress. *Psychophysiology*, 49(4):470–477, 2012.

[98] E. Sazonov, S. Schuckers, P. Lopez-Meyer, O. Makeyev, N. Sazonova, E. L. Melanson, and M. Neuman. Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior. *Physiological measurement*, 29(5):525, 2008.

[99] A. J. Sellen and S. Whittaker. Beyond total capture: a constructive critique of lifelogging. *Communications of the ACM*, 53(5):70–77, 2010.

[100] B. W. Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.

[101] A. Singhal. Introducing the knowledge graph: things, not strings. *Official google blog*, 5, 2012.

[102] E. Stellar and E. E. Shrager. Chews and swallows and the microstructure of eating. *The American journal of clinical nutrition*, 42(5):973–982, 1985.

[103] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *2007 11th IEEE international symposium on wearable computers*, pages 37–40. IEEE, 2007.

[104] E. M. Tapia, S. S. Intille, and K. Larson. Activity recognition in the home using simple and ubiquitous sensors. In *International Conference on Pervasive Computing*, pages 158–175. Springer, 2004.

[105] E. Thomaz, I. Essa, and G. D. Abowd. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1029–1040. ACM, 2015.

[106] C. C. Tsai, G. Lee, F. Raab, G. J. Norman, T. Sohn, W. G. Griswold, and K. Patrick. Usability and feasibility of pmeb: a mobile phone application for monitoring real time caloric balance. *Mobile networks and applications*, 12(2-3):173–184, 2007.

[107] T. Van Kasteren, A. Noulas, G. Englebienne, and B. Kröse. Accurate activity recognition in a home setting. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 1–9. ACM, 2008.

[108] G. Villalobos, R. Almaghrabi, P. Pouladzadeh, and S. Shirmohammadi. An image procesing approach for calorie intake measurement. In *Medical Measurements and Applications Proceedings (MeMeA), 2012 IEEE International Symposium on*, pages 1–5. IEEE, 2012.

[109] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[110] P. Wang and A. F. Smeaton. Semantics-based selection of everyday concepts in visual lifelogging. *International Journal of Multimedia Information Retrieval*, 1(2):87–101, 2012.

[111] P. Wang, A. F. Smeaton, Y. Zhang, and B. Deng. Enhancing the detection of concepts for visual lifelogs using contexts instead of ontologies. In *Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on*, pages 1–6. IEEE, 2014.

[112] P. Wang, L. Sun, S. Yang, A. F. Smeaton, and C. Gurrin. Characterizing everyday activities from visual lifelogs based on enhancing concept representation. *Computer Vision and Image Understanding*, 148:181–192, 2016.

[113] U. Westermann and R. Jain. Toward a common event model for multimedia applications. *IEEE MultiMedia*, 14(1), 2007.

[114] S. White and P. Smyth. A spectral clustering approach to finding communities in graphs. In *Proceedings of the 2005 SIAM international conference on data mining*, pages 274–285. SIAM, 2005.

[115] W. Willett. *Nutritional epidemiology.* Oxford University Press, 2012.

[116] J. Yang. Toward physical activity diary: motion recognition using simple acceleration features with mobile phones. In *Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics*, pages 1–10. ACM, 2009.

[117] K. Yatani and K. N. Truong. Bodyscope: a wearable acoustic sensor for activity recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 341–350. ACM, 2012.

[118] C. Zhu and W. Sheng. Human daily activity recognition in robot-assisted living using multi-sensor fusion. In *2009 IEEE International Conference on Robotics and Automation*, pages 2154–2159. IEEE, 2009.

[119] F. Zhu, M. Bosch, I. Woo, S. Kim, C. J. Boushey, D. S. Ebert, and E. J. Delp. The use of mobile devices in aiding dietary assessment and evaluation. *IEEE journal of selected topics in signal processing*, 4(4):756–766, 2010.

[120] E. Ziglio, C. Currie, and V. Rasmussen. The who cross-national study of health behavior in school-aged children from 35 countries: findings from 2001–2002. *Journal of School Health*, 74(6):204–206, 2004.