早稲田大学審査学位論文
博士（人間科学）

Personal Data Analysis for Well-Being Oriented Living Support

ウェルビーイング志向生活支援のためのパーソナルデータ分析

２０２１年7月

早稲田大学大学院　人間科学研究科

粕谷清治

KASUYA, Seiji

研究指導担当教員：　金　群　教授

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research Background

With the rapid development of information and communication technology, such as artificial intelligence (AI), big data and Internet of Things (IoT), personal data relevant to an individual can be collected from a variety of sensors embedded in smartphones and wearable devices. In addition, location information can be obtained from GPS and WiFi, and new value can be expected to be created from location information containing various contextual information. For example, tweets of Twitter, a popular SNS, also contain content with location information. We post tweets and pictures with location information even when you go out.

On the other hand, well-being oriented living support is widely expected to meet individual needs. Nowadays people have a variety of health-minded lifestyles and living environments in comfortable places, and the sense of value has become diversified. For example, if the activity level in a day is low, it is good to give a personalized suggestion, such as an advice on more exercise. It can be said that the importance of personal data analysis is increasing.

Personal data generally consists of the following three categories: 1) Volunteered data, such as user profiles; 2) Observed data collected by sensors and from SNS, Internet and online services; 3) Inferred data extracted and analyzed from the volunteered or observed data. Especially, data related to places contain various contextual information, so they can be said to be an important part of personal data. To deeply understand the living of an individual, it is important to further analyze the living habits and locations.

We may have had such an experience of having a positive feeling or emotion about a place sometimes and having a negative feeling or emotion about the same place some other times. For example, one day, when you go home unsatisfied with your work, on your way home, you walk into a cafe and spend some time to talk with some people. When you walk out of the cafe, you are refreshed with a different mood and feeling. From this experience, if going to this cafe becomes a habit, you can think that this is a place of attachment and reliance on your heart, where you can go without being aware of it. Although many studies use social data for sentiment analysis, few studies focus on emotion analysis and quantification of individual feelings about a place.

1.2 Purpose of This Study

The purpose of this study is to explore and verify effective personal data analysis methods to provide well-being oriented living support for individuals. This thesis consists of three studies and describes the model, framework, and approach to personal data analysis for the purpose of well-being oriented living support. The protocol of this study was approved by the Ethics Review Committee on Research with Human Subjects of Waseda University (Approval No.: 2017-224).

In Study 1, an integrated framework for personal data analysis to provide well-being oriented living support is newly proposed, and the feasibility and effectiveness of the proposed approach is verified by the statistical method.

Study 2 is to build a new personal model for a place, use the method of decision tree to extract the relationship between transportation means and destinations, and finally grasp the living habits and patterns.

In Study 3, three sentiment analysis methods are used to measure and quantify individual emotions and feelings on places, and experiments are designed to evaluate which method is more appropriate.

1.3 Thesis Organization

The rest of this thesis is organized as follows.

Chapter 2 describes and discusses the basic concept, model, and framework for well-being oriented personal data analysis to offer suggestions and advices to improve the living quality of an individual. A feasibility study with an application scenario is given to demonstrate the effectiveness of the proposed approach.

In Chapter 3, a personal model for a place, namely personal whereabouts model, is introduced, and experiments are designed to extract factors related to transportation means and destinations by decision trees and capture the living patterns of an individual for better personal data analysis.

Chapter 4 describes the approach to measuring and quantifying an individual's feelings about a place using three representative methods in sentiment analysis: emotion dictionary, personalized dictionary, and Bayesian classification. Experiments and protocols are designed to evaluate these methods using tweet data including locations and an individual's emotion changes with regard to these locations before entering and, after exiting a location.

Chapter 5 summarizes this study by discussing the major contributions, features with comparison to Research with First-Person's View, and academic and social significances. Finally, future work on improvement of personal data analysis for well-being oriented living support is addressed.

Chapter 2

Well-Being Oriented Living Support Based on Personal Data Analysis†

2.1 Background

In recent years, with the rapid development of cyber computing technology, larger amount of individual related data is generated and collected from the digital society. Application and utilization of this kind of big data has become increasingly important for fields ranging from personal education to public health. More recently, personal data has been considered as a promising component of Internet of Things (IoT), and personal data analytics has been proposed and applied with reference to 1) healthcare and well-being, 2) life logs and citizen services, and 3) wearable devices [1].

According to WHO, health is generally defined as "a state of complete physical, mental, and social well-being, and not merely the absence of disease or infirmity" [2]. With the help of life logs and personal information management by wearable devices, more and more people are seeking to live in a more comfortable living environment with various senses of value. Thus, to pursue a better and healthy life in the integrated cyber-physical-social world, it is essential to understand an individual's well-being status based on the organization and analysis of personal data.

In a previous study, we proposed a conceptual framework to facilitate well-being oriented living support, which aims at providing personalized recommendations based on users' classified daily living data [3]. Furthermore, we have analyzed and reported results of experiments on organizing and analyzing personal data to help understand users' life styles [4]. In this study, we focus on cyber-enabled well-being oriented daily living support for an individual based on personal data analysis. It is important to note that we did not seek to generalize the results of personal data analysis to other persons. In this chapter, our vision, basic idea, and model for well-being oriented living support based on personal data analysis have been addressed. We propose an extensive framework of personal data analytics for well-being oriented living support in terms of behavioral analysis based on the data collected from an individual's daily activities. Additionally, in this chapter, we demonstrate the feasibility of this framework through a study that used and analyzed a set of personal and environmental data.

2.2 Related Work

2.2.1 Personal data analysis for daily life support

Gemmel, et al. [4] introduced the design and implementation of a system named MyLifeBits, which aimed at storing all of one user' s digital data based on a set of principles. Lou et al. [5] proposed the concept of intelligent personal health record (iPHR), and designed a health and medical information system to provide users with personalized healthcare information to facilitate their activities of daily living. Estrin [6] discussed personalized data analysis, which can be used to infer health status and well-being from digital behaviors.

Dobrinevski [7] identified categories of the collected raw personal data, and presented evidences of measurable changes in the personal capability by focusing on personal analysis, especially on patterns of communication and collaboration between individuals. Epstein et al. [8] presented a model of personal informatics used by self-trackers, in which personal informatics is considered in four types of lapses; forgetting, upkeep, skipping, and suspending. Teraoka [9] introduced an organizing structure with a zooming user interface in an interactive system, which enabled users to recall the collected personal data from several viewpoints, and further helped them to find various related information.

Mun et al. [10] developed a current awareness system named Personal Data Vaults, in which a privacy architecture was designed for individuals to retain their ownership of personal data.

2.2.2 Daily activity classification and personalized recommendation

Schuldhous et al. [11] evaluated four inertial sensors that were placed on the wrist, chest, hip, and ankle of 19 subjects, who performed activities of daily living. Xu et al. [12] provided a context-aware personalized activity classification system based on the concept of context specific activity classification. They designed and implemented sensor fusion algorithms that were involved in personalized activity monitoring and activity classification. Li et al. [13] presented and extracted personalized fitting patterns to predict missing ratings based on the similarity score set, which combines both the user-based and item-based collaborative filtering. Moreover, they proposed algorithms to increase the recommendation accuracy based on the traditional collaborative filtering. Minor et al. [14] developed two algorithms for learning activity predictors, in which the Independent Predictor was used as a simple baseline approach, and the recurrent activity predictor was introduced to improve the baseline model. Aissi et al. [15] proposed enhanced spatial data warehouse exploitation by recommending personalized MDX queries to the users while taking into account their preferences and needs.

2.2.3 Behavior analysis for daily life management

Chowdhury et al. [16] designed three different recommendation algorithms and described a pattern weaving approach, which effectively provided users with contextual and interactive recommendations of composition knowledge and usable model patterns. Morsel and Kerschberg [17] presented a personal health explorer as a semantic health recommendation system, which allowed users to perform ontology guided semantic search for relevant information. Benlamri and Zhang [18] proposed a knowledge-driven recommender for mobile learning on the Semantic Web, which is an approach for context integration and aggregation in an upper ontology space.

Consolvo et al. [19] proposed the concepts and strategies based on behavioral and social psychological theories, to design and build a system that encourages people to live in physically active lifestyle. Jalali and Jain [20] proposed the ideal healthy life based on several personal life events, which were collected from asynchronous data streams with wearable sensors from heterogeneous sources. Bentley et al. [21] built a health mash-up system as an individually focused platform to discover the trends over time from the multiple aspects of well-being data, and discussed the behavior changes among the participants within the mobile-based environment. The observation results cannot only promote the development of well-being related systems, but can also benefit users' general well-being in their daily lives. Bogomolov et al. [22] proposed an alternative approach to providing evidences to reliably recognize daily stresses based on behavioral metrics with additional indicators, such as weather conditions and personality traits. Mafrur et al. [23] proposed an approach to modeling human behavior based on users' smartphone data logs by combining a variety of sensor data rather than only focusing only on one sensor. Castro et al. [24] designed and implemented a framework named InCense, to analyze the frequent activities performed by elder adults and the conditions related to their habits or symptoms. McNaull et al. [25] proposed a system which can provide feedback to highlight the important criteria for sleep quality during the night. Jin et al. [26] proposed a human-centric safe and secure framework for ubiquitous living environments, which aimed at providing holistic and integrated living support, such as accident prevention, wandering detection, and health control.

2.2.4 Position of this study

Personal data analysis is playing an important role in facilitating daily living support. Many studies have focused on analyzing personal behaviors and daily activities in order to provide personalized recommendations. Frameworks and platforms have been developed for health monitoring and daily life management. As compared with these related works, the present study focuses on analyzing

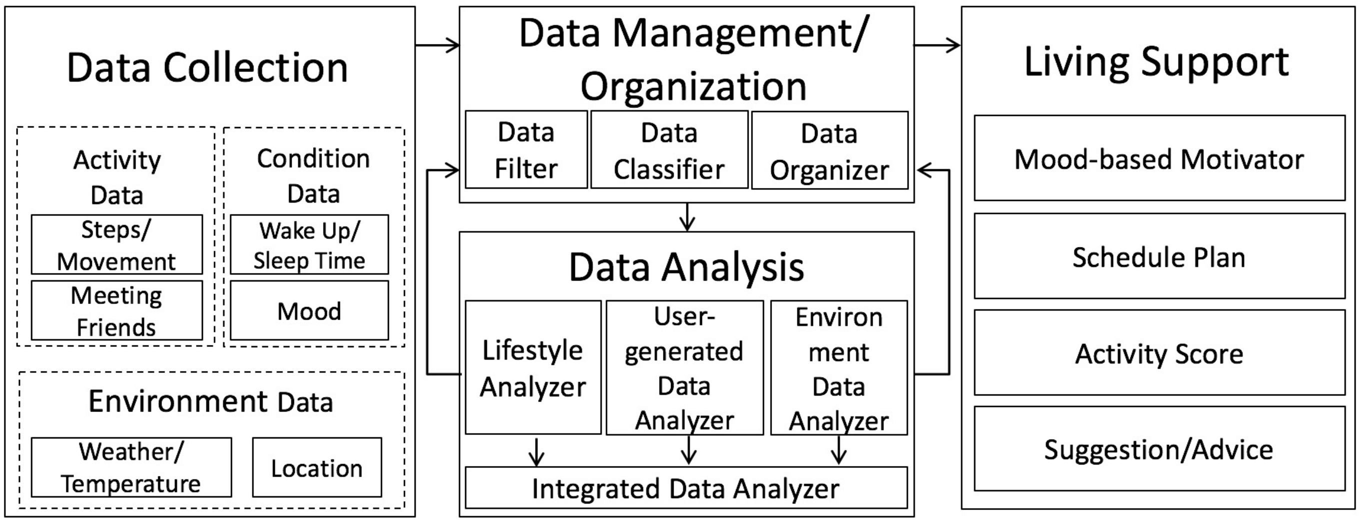

Figure 2.1 Framework of Well-Being Oriented Living Support

individuals' daily personal data from their life events and local environments, to provide them with well-being oriented daily living support according to their different lifestyles.

2.3 Concept of Well-Being and Living Support

2.3.1 Definition of Well-being

Generally, well-being is defined as maintaining optimal health and social connections, and the elements of well-being include self-acceptance, positive relations with others, autonomy, environmental control, life goals, and personal growth [27]. Human well-being concerns the provision of personal safety and secure life, as well as the minimum of supplies for a good life. To this purpose, it is necessary to maintain good living and to be able to choose social connections freely [2].

In this study, we focus on providing well-being oriented support for an individual based on personal data that represents an individual's personal experiences in daily living. In particular, it would associate with several small successful experiences, which can result in personal satisfaction, a sense of fulfillment, and a sense of achievement. To achieve this, the extraction and analysis of the daily activities related to such successful experiences can facilitate the understanding and sharing of information on a well-being oriented human life, and can provide an individual with adequate and sustainable personalized living support for the enhancement of social well-being and health.

2.3.2 Framework of Well-Being Oriented Living Support

Considering the above discussion, in this study, living support aims to continuously improve an individual's quality of life (QoL) to achieve well-being. Contrary to other studies, in addition to deriving a multi-faceted understanding of both mental and physical health, our study tries to understand well-being oriented life based on the accumulation of a series of small successful experiences. That is, based on the analysis of an individual's daily activities, it is attempted to provide adaptive support to each person with reference to his/her unique lifestyle. Furthermore, based on extractions of the features of well-being and life environments, both mental and physical health can be improved based on a comprehensive analysis of their associations.

As shown in Figure. 2.1, four major modules, i.e., Data Collection, Data Management/Organization, Data Analysis, and Living Support, are proposed and designed to provide well-being oriented living support.

Firstly, three basic kinds of living data are collected from the daily life of individuals, namely, activity data, which includes an individual's behavioral habits with reference to some common purposes; condition data, which includes an individual's daily routines and moods; and environment data, which includes the weather/temperature information and the geographical location data.

Then, the collected data are pre-processed for further analysis in the Data Management/Organization module, which includes a Data Filter, Data Classifier, and Data Organizer.

Using these modules, we analyze an individuals' daily data with reference to three different aspects. Specifically, the Lifestyle Analyzer is used to analyze the user's different lifestyles, aiming at identifying the diversified features in the user's daily lives. The User-generated Data Analyzer is employed to extract a user's behavioral features, which can provide the user with personalized recommendations.

Finally, the Environment Data Analyzer is used to analyze the corresponding data from different environments, which can provide a user with timely support according to the dynamical detection of the changed environments. These analysis results are integrated and comprehensively considered by the Integrated Data Analyzer.

Finally, in the Living Support module, the integrated mechanisms are developed to provide a specific individual user with personalized suggestions, such as a schedule plan or activity score, to support his/her well-being oriented life.

2.3.3 Living Support Based on Personal Data Analysis

In this study, to analyze personal data and provide well-being oriented living support, we define and categorize a set of personal data as follows.

Step (ST): ST is one kind of user action or the activity parameter for a single day. For instance, we count and record a user's daily walking steps, which can be used for the action analysis.

Sleep Time (SL): SL is used to describe one kind of condition data for a single day. SL can be utilized to monitor a user's daily sleep cycle and to further infer the user's mood and personal satisfaction.

Deep Sleep Time (DS): DS is one kind of special SL data for a single day. DS is used to describe the sleep quality, and should be viewed as an important factor related to a user's mood and activity.

Tweet (TW): TW indicates the number of posts of a user in social media (i.e., Twitter in this study), and it is related to the user's activities during a single day. We collect the tweets from Twitter to analyze the contents of the user's posts. Weekday or Weekend/Holiday (WH): WH indicates day-related information, such as weekday, or weekend/holiday. This kind of data will strongly influence the user's daily schedule or activities.

Current Weather (CW): CW is one kind of external environment data. For instance, CW can describe the current weather, such as "Rain" or "No Rain," which will influence the changes in the activities included in the daily schedule.

Location (LO): LO is used to describe a special place or venue. It is one kind of external environment data.

The basic procedure to provide well-being oriented living support based on personal data analysis

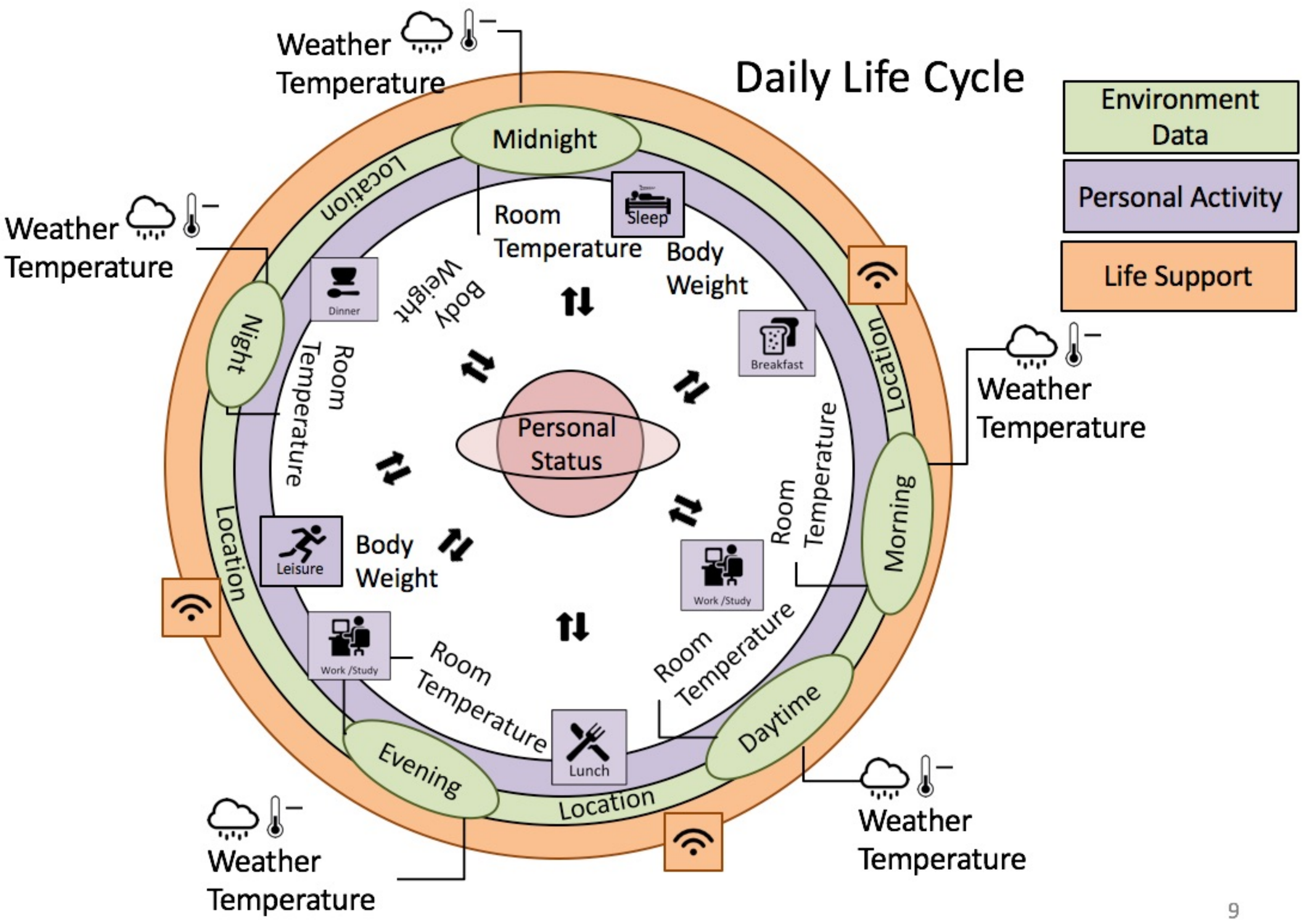

Figure 2.2 Conceptual Image of the Data Collected in the Daily Life Cycle

is shown as follows.

Step 1.Collect personal data, including an individual's behavioral habits, daily routines, and activity related environment data;

Step 2.Pre-process the collected data to filter the noise data, and label them as ST, SL, DS, TW, WH, CW, and LO. Subsequently, classify the data into three major categories: volunteered data, observed data, and inferred data;

Step 3.Analyze these categorized data to extract the behavioral features, to detect information on the dynamic changes in the environment, and to infer different lifestyles;

Step 4.Compare the analysis results in different situations, and provide a specific user with personalized recommendations, such as a schedule plan, activity score or mood based suggestions.

2.4 Data-Driven Living Support

2.4.1 Data Collection in Daily Living

Generally, to analyze personal data for the provision of well-being oriented living support, the data collected from an individual's daily life can be classified into three basic types [28]: volunteered data, observed data, and inferred data. Volunteered data refers to the open data that is created and shared by a group of individuals directly, such as the user profile data. Observed data refers to the data that is collected from the record/history of actions of an individual, such as the location check-in data. The inferred data refers to the data that is extracted and analyzed from the volunteered or observed data, such as the derived economic condition. All these kinds of data comprise the so-called personal data on our daily life, and they can be viewed as an important part of the presented big data. The observation, collection, and analysis of this kind of personal big data benefits the understanding of an individual's unique lifestyle, and can be finally used to provide him/her feedback and well-being oriented living support services. In other words, the data observed and collected from wearable devices, smart phones, and web services can be stored and organized on the cloud, and the integrated data, along with the extracted features and patterns, can be analyzed to provide an individual personalized feedback to support well-being oriented living.

Figure 2.2 presents a conceptual image of the process of collection of heterogeneous data in the daily life cycle. Specifically, the personal data of a specific individual user can be collected from morning to midnight, through a variety of wearable devices. All activity data, condition data, and environment data are detected and selected, including the user's movements, the weather and room temperature, and the temporal and location data related to both the work and private life of the user. In addition, personal information such as body weight, mood, and sleeping time can be recorded. All these data can be integrated and organized to provide an individual with the personalized daily living support.

2.4.2 Framework of Personal Data Analytics

Based on the above discussion, here we demonstrate how the collected personal data can be analyzed to provide an individual user with well-being oriented living support. The framework for the personal data analysis has been presented in Figure. 2.3.

As shown in Figure. 2.3, after data filtering and classifying, the three basic data, i.e., activity data, condition data, and environment data, can be extracted as the observed data, while the user profile data can be classified as volunteered data.

Then, four components, namely User Feature Extractor, Lifestyle Analyzer, Location Detector, and Situation Analyzer, are used to analyze the observed and volunteered data, to further obtain the inferred data. Specifically, the activity data and users' profiles can be utilized to analyze and extract users' activity related features to refine their profiles.

The activity and condition data can be used to analyze an individual user's lifestyle. The environment data can be utilized to automatically detect the current information in terms of location and situation, respectively. Thus, three kinds of inferred data can be obtained, that is, the data related to a user's activity-based daily routines, dynamical situation, and classified lifestyle. Further, these diversified data can be utilized to provide the user well-being oriented living support from the Activity Score Based Adviser, Lifestyle Based Recommender, Schedule Plan Generator, and Mood Based Monitor.

2.5 Feasibility Study Based on Statistical Analysis

2.5.1 Data Sets

In this study, we obtained different kinds of personal data to conduct a feasibility study. Specifically, the activity-related data collected from wearable devices (such as Jawbone [29]) includes the number of walk steps, sleep time, and deep sleep time. The environment-related data in our living area, including the atmospheric pressure, precipitation in one hour or 10 minutes, minimum and maximum temperature, average humidity, wind direction with the maximum wind speed, sunshine

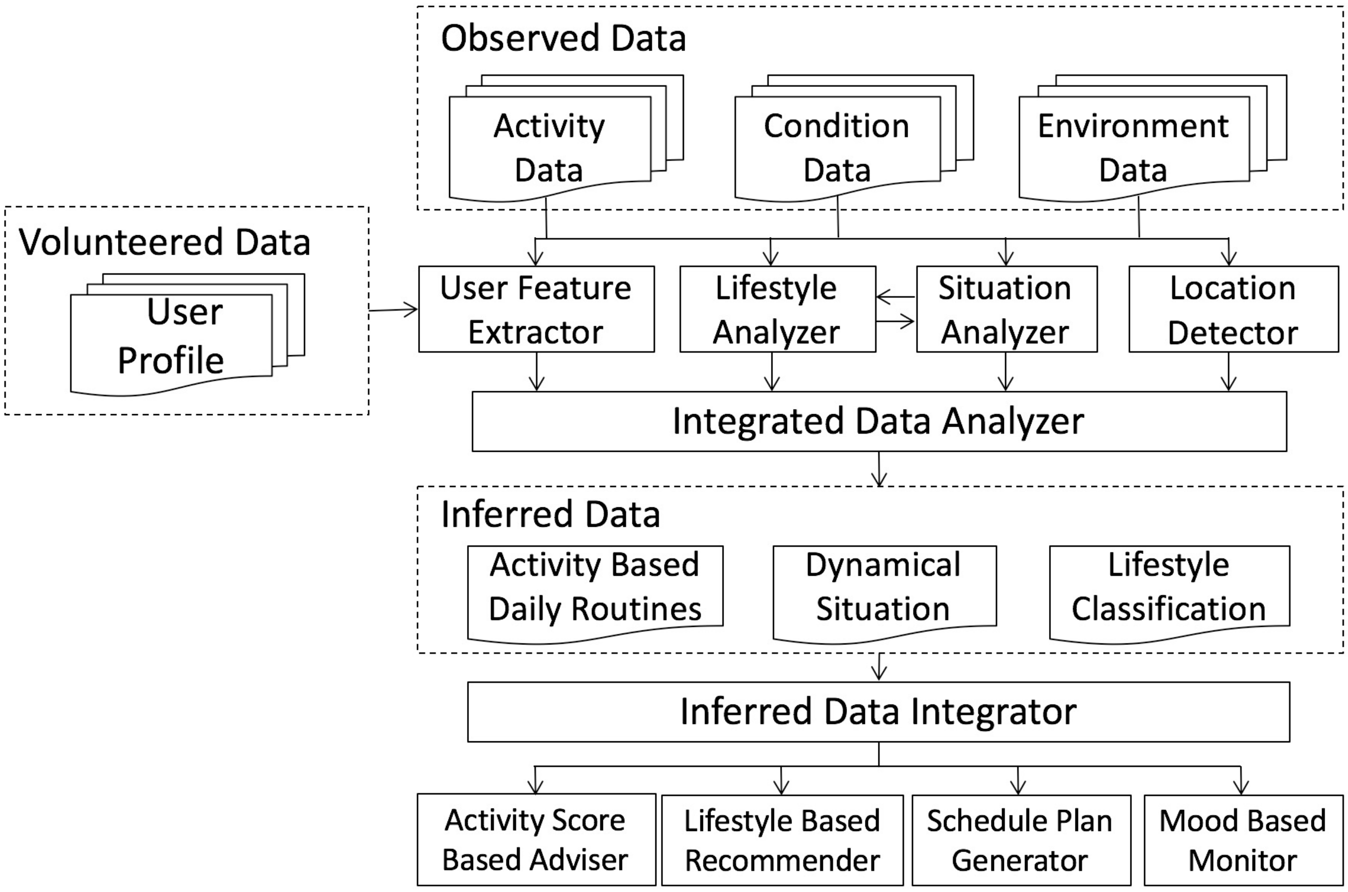

Figure 2.3 Personal Data Analysis for Cyber-Enabled Living Support

Table 2.1 Data Types, Units, and Sources of the Collected Personal Data

Data Type	*Unit*	*Data Source*
Sleet time	h	Wearable device
Deep sleep time	h	Wearable device
Number of walk steps	Step	Wearable device
Total number of Tweets	Tweet	Twitter
Precipitation of one hour	mm	Website

hours, and daytime and night weather were collected from the Japan Meteorological Agency (JMA) [30]. Besides, Twitter [31] data was utilized as a type of condition-related data.

Table 2.1 presents the detailed information pertaining to the collected personal data with different data types, units, and data sources.

For a specific individual user, a business man in this study, we collected the activity, environment, and condition-related data for 205 days (from July 4, 2014 to January 24, 2015). Furthermore, we classified this data set into two subsets marked as "weekday" and "weekend/holiday," to conduct further comparisons. We used the collected data for the specific user to conduct the feasibility study based on the statistical analysis.

2.5.2 Basic Properties for Statistical Analysis

As discussed above, after filtering the noise data and missing data from the raw activity and environment-related data, we obtained experiment data for 112 weekdays and 64 weekends/holidays. For each kind of data in these two sub-sets, we calculated the Arithmetic Mean, Standard Deviation (SD), Median, Minimum, and Maximum for the statistical analysis.

The arithmetic mean [32] was calculated using Equation. (2.1).

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.1}$$

where n denotes the size of data set, i.e., the number of elements in the data set, and x_i denotes the value of each element in the data set.

The Standard Deviation (SD) [32] was obtained using Equation. (2.2).

$$SD = \sqrt{\frac{1}{n}\sum_{n=1}^{n}(x_i - \overline{x})^2} \tag{2.2}$$

The Median $Q_{\frac{1}{2}}(x)$ was calculated using Equation. (2.3).

$$Q_{\frac{1}{2}}(x) = \begin{cases} x'_{\frac{n+1}{2}}, & if\ n\ is\ odd \\ \frac{1}{2}\left(x'_{\frac{n}{2}} + x'_{\frac{n}{2}+1}\right), & if\ n\ is\ even \end{cases} \tag{2.3}$$

where x'_j is the value of the element in the data set, sorted in an ascending order.

Table 2.2 and 2.3 present these descriptive statistics for the data on weekdays and weekends/holidays, respectively.

2.5.3 Statistical Analysis for Different Data Sets

We conducted several comparison analyses under different conditions. In addition to the classification based on weekdays and weekends/holidays, we consider the weather data, i.e., "Rain" and "No Rain," to conduct the comparative analysis.

Table 2.2 Basic Properties of Personal Data on Weekdays.

Weekday(N=112)	*Mean*	*SD*	*Median*	*Minimum*	*Maximum*
Precipitation (mm)	3.17	9.18	0	0	57.5
Average temperature (℃)	16.89	9.15	19.3	0.7	30.2
Sleep time (h)	5.39	1.44	5.14	1.16	10.27
Deep sleep time (h)	3.07	1.05	3.04	0	5.92
Ratio of deep sleep versus sleep time (%)	56.81	12.95	58.16	0	82.94
Number of walk steps (step)	6775.85	2183.81	6850.5	1156	11597
Number of tweets in AM (0:00-11:59)	2.31	1.44	2	0	14
Number of tweets in PM (12:00-23:59)	1.23	2.38	0	0	17
Total number of tweets (tweet)	3.54	2.72	3	1	20

Table 2.3 Basic Properties of Personal Data on Weekends/Holidays.

Weekday(N=112)	*Mean*	*SD*	*Median*	*Minimum*	*Maximum*
Precipitation (mm)	4.14	15.87	0	0	109.5
Average temperature (℃)	15.82	9.56	16.8	-0.2	30.2
Sleep time (h)	6.25	1.77	6.11	1.9	10.78
Deep sleep time (h)	3.48	12.5	3.65	0.88	6.3
Ratio of deep sleep versus sleep time (%)	55.28	13	55.68	25.83	100
Number of walk steps (step)	4866.66	3158.01	4002	320	13101
Number of tweets in AM (0:00-11:59)	3.33	1.95	2	0	10
Number of tweets in PM (12:00-23:59)	2.61	2.97	2	0	17
Total number of tweets (tweet)	5.94	3.78	5	2	26

The experiment data were examined using the following four steps:

(1) We classified the whole data set in terms of "weekday" and "weekend/holiday."

(2) Using the precipitation data, we further divided the data set into "Rain" and "No Rain."

(3) We calculated the mean and standard deviation for each data set.

(4) We applied the Welch' s t-test [33], and calculated the corresponding values for t-value, Degrees of Freedom (DF), and p-value for each data set, to identify the significance of the difference between two data sets.

Generally, it depends on the normality and/or equality of variance to apply a statistical method for a specific data set. The Welch' s test can be used when the data is normally distributed, but not necessary to guarantee the equality of variance. The normality is the feature of data distribution with a top value in the center of the distribution curve. To verify the normality, several basic tests, such as the Shapiro-Wilk test and the Kolmogorov-Smirnov test, can be used. However, it will cause the so-called multiplicity problem, which occurs when another statistical test is conducted again for a data set after a basic test has already been applied for the same data set. For instance, different p-values may be obtained. Multiple uses of statistical tests should be avoided. Therefore, in this study, we decided to directly apply the Welch' s test for the data of walk steps, sleep time, deep sleep time and number of tweets. As is well known, the data related to a human (such as the height, weight, sleep time, walk steps) is of the normality. We confirmed this by plotting the data distribution graphs for these data of walk steps, sleep time and deep sleep time. On the other hand, the number of tweets could not be guaranteed to have the normality. In this case, the post hoc power, defined by the sample size, effect size and significance level (in this study, they are 176, 0.8 and 0.05, respectively, since we defined the significance level as 0.05 for two-sided tests and the effect size as 0.8 [34]), can be used to verify the correctness of the applied statistical test. Generally, the statistical test is regarded valid and meaningful if the post hoc power is larger than 0.8 [35]. We will calculate the post hoc power for the number of tweets after the statistical test.

To apply the Welch' s test, specifically, t-value [36] can be calculated by Equation. (2.4), and the Degree of Freedom (DF) [33] can be obtained by Equation. (2.5).

$$t = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{2.4}$$

$$DF = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2^2})^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1^- 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2^- 1}} \tag{2.5}$$

where $s1 = \sqrt{\sum_1 / n_1(n_1 - 1)}$ and $s2 = \sqrt{\sum_2 / n_2(n_2 - 1)}$, which denote the variance of each data set. Σ_1 and Σ_2 are the sums of squares of the elements in each data set from their mean, respectively.

Moreover, p-value is the probability for a given statistical model in statistical hypothesis testing, which can be obtained by Equation. (2.6), or more generally, by the so-called t-distribution table [37].

$$t_a\left(DF\right) : \int_{t_a}^{\infty} \frac{1}{\sqrt{DF}\ B\left(\frac{1}{2}, \frac{DF}{2}\right)\left(1 + \frac{t^2}{DF}\right)^{\frac{DF+1}{2}}} dt = \alpha \tag{2.6}$$

Applying these, we can obtain the results for the number of walk steps, sleep time, and deep sleep time from the activity-related data, and the number of tweets from the condition-related data. As for the null hypothesis, we assumed that there is no difference between the mean of two observed data according to "Rain" or "No Rain" days on the weekdays and weekends/holidays. The null hypothesis can be rejected when the p–value is lower than the significance level. As is known, the p–value represents the probability that the difference occurs by chance. If the p–value is lower than 0.05, it means the difference between two data sets is confirmed. If the p–value is lower than 0.01, it indicates there is a difference almost certainly. Otherwise, the null hypothesis will be accepted.

The corresponding results are shown in Figures 2.4 to 2.7 and Table 2.4 as well. From Table 2.4, we can see that the test items, the number of walk steps, sleep time, and deep sleep time, on weekdays, show the significant difference between "Rain" or "No Rain" days (walk steps at the significance level of 0.01, and sleep time and deep sleep time at the significance level of 0.05). On the other hand, for the number of tweets, no significant difference between "Rain" or "No Rain" days is observed. Moreover, the null hypothesis for all the test items on weekends/holidays cannot be rejected. In other words, there is no difference between two observed data on "Rain" or "No Rain" days for weekends/holidays. From the result given and discussed above, we can conclude that for a specific individual, the weather factor of "Rain" or "No Rain" can be viewed as an important influence on his/her walk steps, sleep time and deep sleep time on weekdays.

In addition, as mentioned above, for the data of the number of tweets, we calculated the post hoc power, and the results were 0.9664 for weekdays and 0.7788 for weekend/holidays, respectively. Particularly, the latter was smaller than 0.8. Since the post hoc power is defined by the sample size, effect size and significance level, we tried to increase the sample size of the data for the number of tweets in order to confirm the correctness of the applied test, i.e., the post hoc power is larger than 0.8. For the data set used above, if one or more items the data were missed for a day, all data for this day would be simply removed to guarantee the integrity of the data set. However, for partly missing data, other methods can be applied, such as to use the arithmetic mean of available data to substitute the missing item(s). In this way, we increased the sample size of the tweet

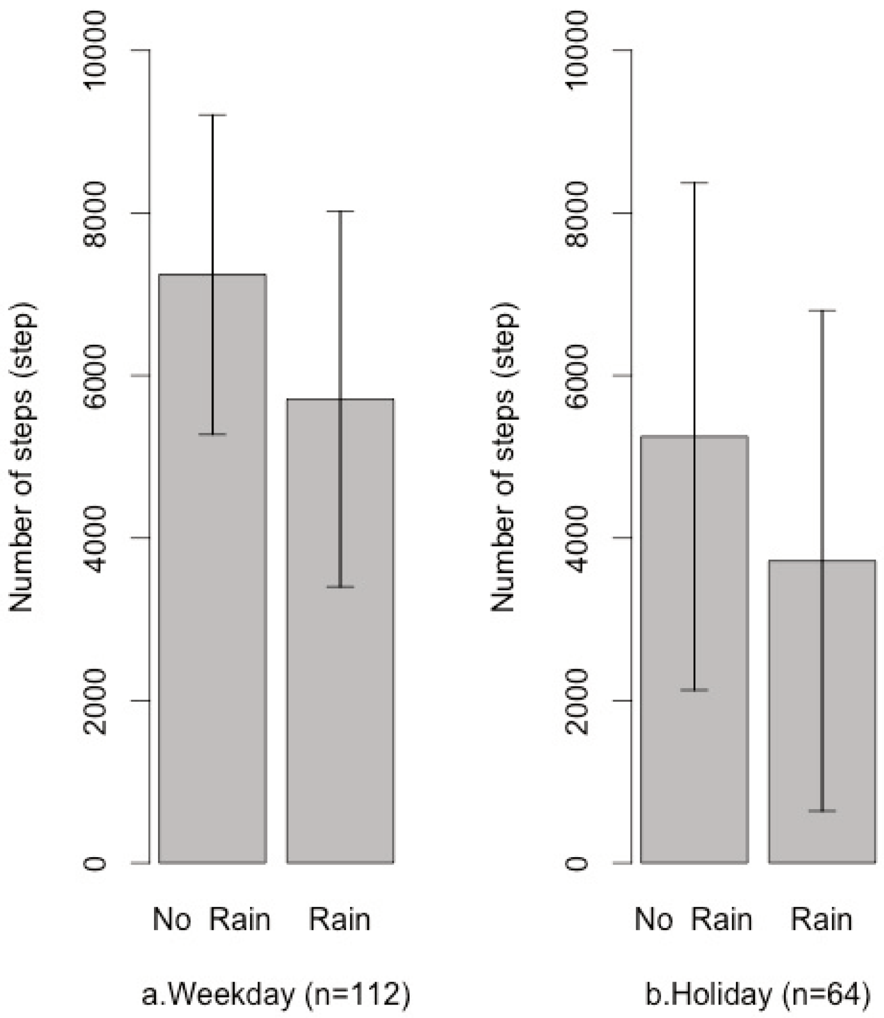

Figure 2.4 Average of Walk Steps in According with Weather on Weekday and Weekend/Holiday

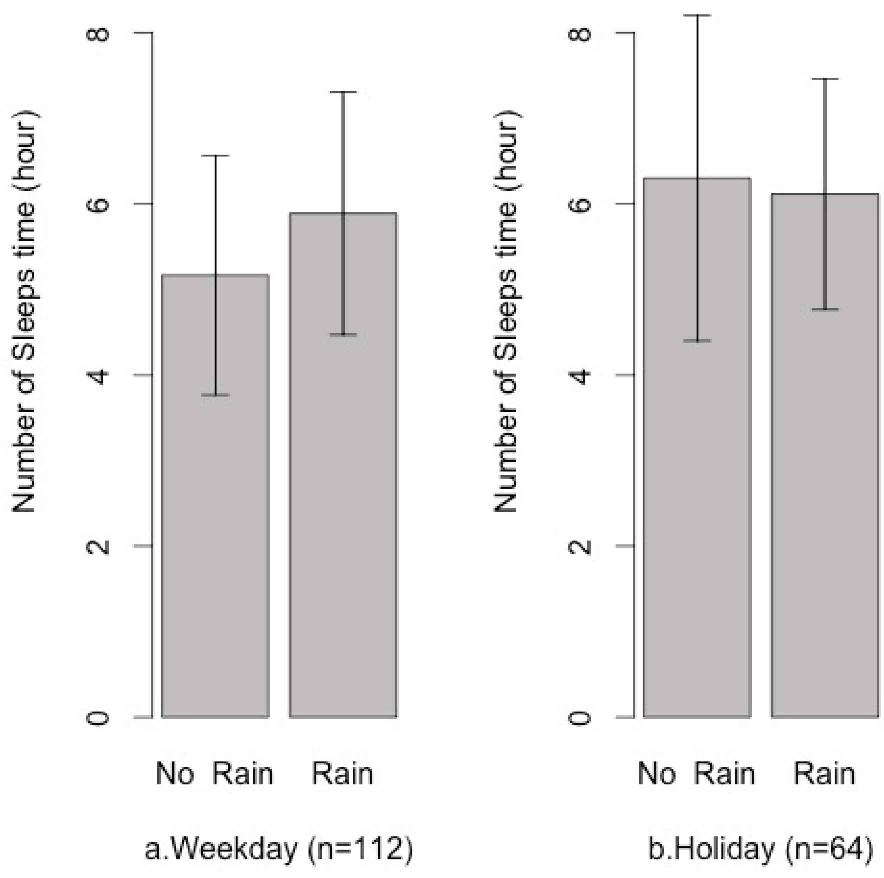

Figure 2.5 Average of Sleeps Time in According with Weather on Weekday and Weekend/Holiday

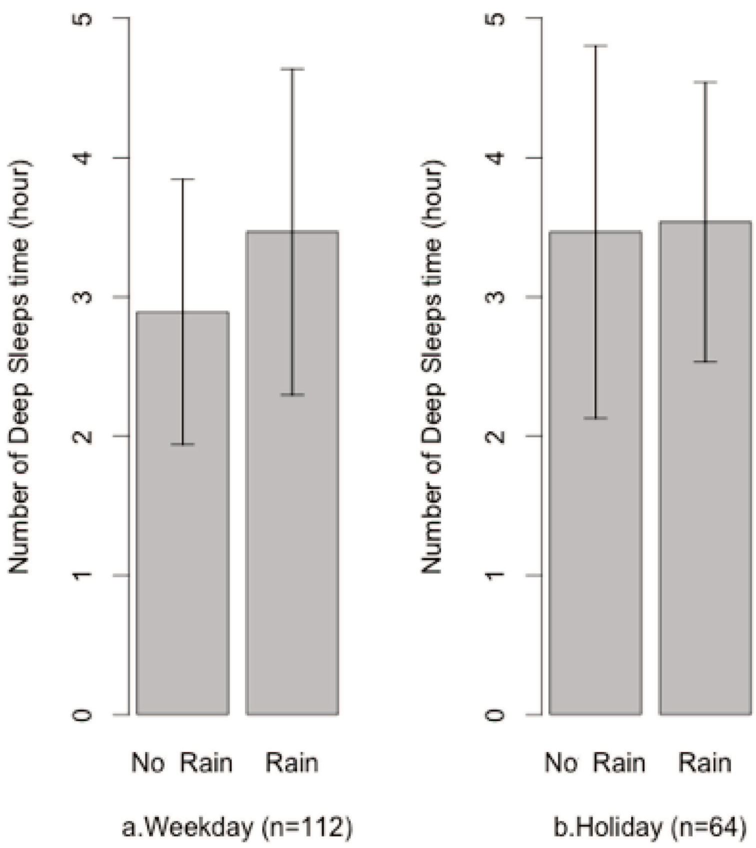

Figure 2.6 Average of Deep Sleeps Time in According with Weather on Weekday and Weekend/Holiday

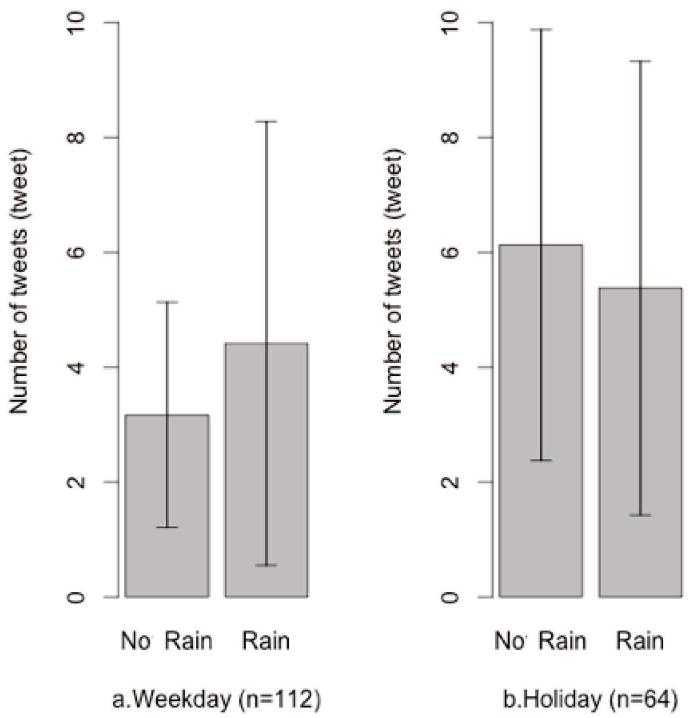

Figure 2.7 Average of Tweets in According with Weather on Weekday and Weekend/Holiday

Table 2.4 Statistic Results According to the Weather on Weekdays and Weekends/Holidays.

Action	*Situation*	*Condition*	*Mean*	*SD*	*DF*	*t-value*	*p-value*
Walk Steps	Weekday (a)	No Rain	7,240.80	1,966.12	54.78	3.3699	0.001 **
		Rain	5,709.21	2,310.47			
	Weekend/Holiday (b)	No Rain	5,249.27	3,121.51	26.06	1.7158	0.098 n.s.
		Rain	3,718.81	3,079.30			
Sleep Time	Weekday (a)	No Rain	5.17	1.40	62.12	2.4831	0.016 *
		Rain	5.89	1.42			
	Weekend/Holiday (b)	No Rain	6.30	1.90	36.28	0.4267	0.067 n.s.
		Rain	6.12	1.35			
Deep sleep time	Weekday (a)	No Rain	2.89	0.95	52.88	-2.5177	0.014 *
		Rain	3.47	1.17			
	Weekend/Holiday (b)	No Rain	3.46	1.34	34.17	-0.227	0.822 n.s.
		Rain	3.54	1.00			
Number of Tweets	Weekday (a)	No Rain	3.17	1.96	40.58	-1.783	0.082 n.s.
		Rain	4.41	3.86			
	Weekend/Holiday (b)	No Rain	5.53	2.35	31.63	1.2145	0.234 n.s.
		Rain	4.74	2.45			

data to 193, and the values of the post hoc power became 0.9721 for weekdays and 0.8036 for weekends/holidays, respectively. And the p-values were changed from 0.082 to 0.631 for weekdays, and from 0.234 to 0.253 for weekend/holidays, but the results of n.s. (not significant) were kept unchanged.

2.5.4 Comparison with Related Work

Existing works have focused on providing a variety of personalized life support considering the following four major features: (1) individual behavior related features, (2) life event based features, (3) personal trait based features, and (4) local environment related features.

We summarize these features of several representative related works and compare them with the present work. As shown in Table 2.5, our proposed framework comprehensively took all of these features into consideration while providing well-being oriented living support.

It is however important to acknowledge that our study did not seek to generalize the results of the personal data analysis to other persons. If fact, it provides an individual with well-being oriented living support based on evidence obtained specially for that specific individual.

Table 2.5 Comparison of This Work with Related Works on Living Support

Research	*Individual Behavior*	*Life Event*	*Personal Trait*	*Local Environment*
J Gemmel et al., Cosolvo et al.	✓			
Jalali and Jain,		✓		
Mun et al., Xu et al	✓		✓	✓
G. Luo et al,		✓	✓	✓
Teraoka, Schuldhaus et al	✓	✓	✓	
This work	✓	✓	✓	✓

2.5.5 Application Scenario

Based on the findings related to the user John, who is a young business man, we assume that he works on weekdays (from Monday to Friday), and rests on Saturday, Sunday, and holidays. Basically, he carefully records his personal data in different environments through several wearable devices. For instance, he uses the wearable device Jawbone on his wrist almost all the time, records his walk steps during the day, and measures the sleep time. Such data can be regarded as observed data for the personal data analysis.

As shown in Figure 2.8, two functional modules: the Lifestyle Analyzer and Integrated Data Analyzer, were utilized to analyze the observed data, and to provide him suggestions for improving his QoL. For instance, based on the automatically recorded sleeping time data for a specific lifestyle in terms of his sleep time, and the Integrated Data Analyzer compares the results with John's previous data, to detect if there is any difference or irregular change in his daily life. If it is found that John is short of sleep, a notice message, such as "Too little sleep this week," will be sent to him, reminding him to improve his lifestyle. Furthermore, two other functional modules, Inferred Data Integrator and Mood-based Motivator, will work together to provide the user with well-being oriented suggestions. For example, based on the analysis results of John's current lifestyle, the Inferred Data Integrator will integrate his current lifestyle, situation, and daily routine related data to analyze his current life quality. If he continues to exhibit irregular living habits, the Mood-based Motivator will send him a suggestion message such as "Go to bed by 11pm tonight," to help him change his lifestyle.

On the other hand, it is assumed that John has a weekly habit of jogging every Wednesday to

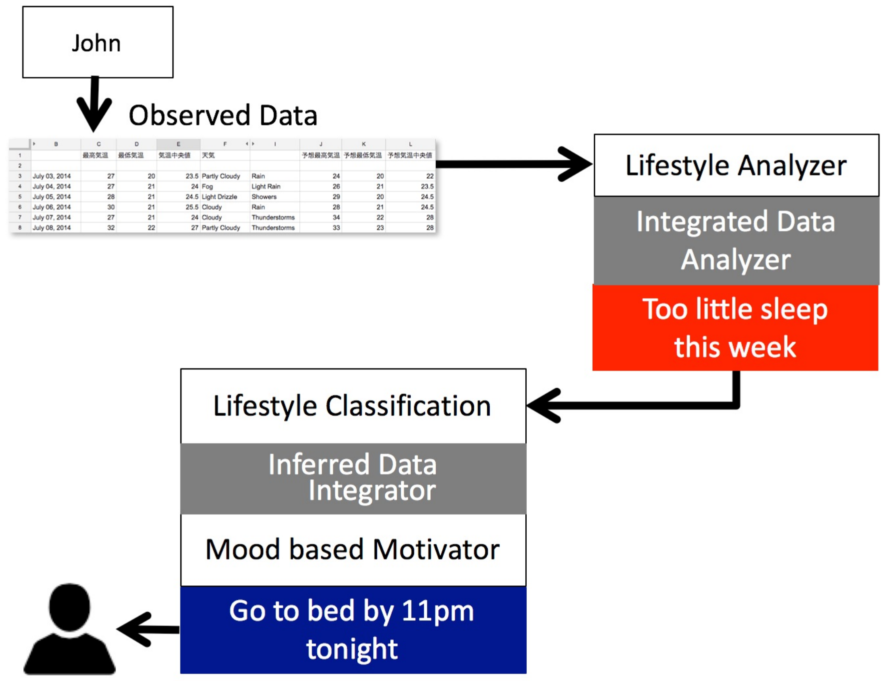

Figure 2.8 Application Scenario for the Mood-Based Motivator

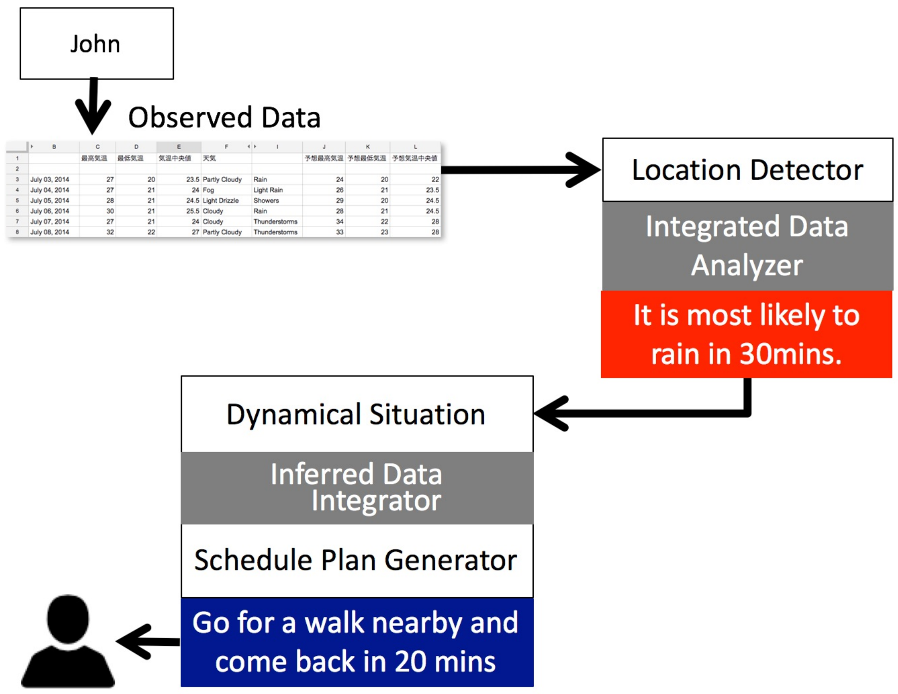

Figure 2.9 Application Scenario for the Schedule Plan Generator

maintain good health. As shown in Figure 2.9, the Location Detector will work with the Integrated Data Analyzer to process the detected location data and the collected weather data simultaneously, and will send him a message such as "It is most likely to rain in 30 mins." Furthermore, considering the dynamic changes in situations, the Inferred Data Integrator and Schedule Plan Generator will be utilized to develop a plan based on a timely changed schedule, such as "Go for a walk nearby and come back in 20 mins," if it is most likely to rain in 20 to 30 mins.

2.6 Summary

The purpose of this study was to provide an individual with well-being oriented living support based on the results or evidence obtained from personal data analysis, especially for this specific individual. In this chapter, we proposed an integrated framework for personal data analysis to provide an individual with individualized living support.

Following the introduction of the basic definition of well-being, we designed and proposed a framework with four major function modules for well-being oriented living support. Then, based on the classification of a user's daily living data collected from the cyber-social-physical environment, the work flow was presented to demonstrate how well-being oriented living support can be provided for an individual based on personal data analysis. Specifically, based on the framework for living support oriented personal data analysis, three kinds of data, namely, volunteered data, observed data, and inferred data, were analyzed, and the statistical analysis technique was demonstrated. Additionally, an integrated data set including five types of personal data collected over 200 days was utilized to conduct a feasibility study. Based on the results of statistical analyses among different data sets, an application scenario was presented to demonstrate the effectiveness of our proposed method.

As for our future work, we aim to improve the design and implementation of our proposed framework with more functional modules. We will also develop corresponding algorithms to aid the provision of well-being oriented living support. Experiments will be conducted to evaluate the proposed framework and application system.

Chapter 3

Personal Whereabouts Model and Extraction of Living Patterns[†]

3.1 Background

Personal data can be recorded using various sensors on smartphones and wearable devices, including locations, and activities. It becomes possible to acquire location data with GPS so as to collect personal data accurately and efficiently, which makes the analysis of behavior patterns and whereabouts important for personal data analysis. However, there are not many works focusing on the study of the individual's whereabouts.

Using applications on smartphone and acceleration sensors, personal data can be briefly summarized and provided visually in a cycle of a week, month, year. However, the aggregated data does not include some environment data such as the weather and temperature where a person lives. By using environment data, we can understand personal behavior more effectively.

In our previous study, we have proposed a well-being oriented framework for effective use of personal data [38]. We conducted an experimental analysis based on people's daily activity data to demonstrate the feasibility of our proposed approach. As a result, according to the analysis of user's daily activities on weekdays and holidays, we found several different lifestyles [39].

In this study, we propose a personal whereabouts model in which environment factors are taken into account. To construct personal whereabouts model, we consider four components: situation analysis, pattern extraction, user profile, and whereabouts attributes. The collected personal data, including location data, such as transportations and destinations, and environment data, such as weather, temperature, and humidity, is organized and analyzed to extract the living patterns according to the different transportations and destinations.

3.2 Related Work

3.2.1 Use of Personal Location Data from Wearable Devices

The methods of recognition of current locations have been studied for a long time using position detection technology.

The position detection technique performs based on either relative positioning or absolute positioning. In various related studies, the false recognition is reduced to predict and verify the transition of the position where a person moves. Guidotti et al. [40] proposed two clustering methods to automatically detect the individual movement based on position information. Their results showed a better accuracy than other existing mobility algorithm. Abe et al. [41] developed a system to collect the location information and a user' s impression data according to different places. Viel et al. [42] proposed a method to detect a user's location by fingerprint, beacons and wifi. They evaluated the attribute of position information using machine learning techniques and applied an evaluation model to estimate the position.

3.2.2 Personal Data and Personalization

Reiss et al. [43] improved the performance of an existing classifier by personalization of activity recognition. Moreover, they built a mechanism that can be used in the mobile system. Guo et al. [44] presented a mechanism to collect various kinds of personal data from multiple wearable devices. They adopted three kinds of context inference methods for the mechanism. Shakiba et al. [45] presented a recommendation method of places and activities by improving location-based collaborative filtering methods. The method consisted of two models. One focused on the temporal feature, and the other focused on the correlation between activities. They improved existing collaborative filtering. Karkar et al. [46] proposed a framework to manage personal health data and showed a process flow for a new therapy. This framework could improve the efficiency to use human health data. Esptein et al. [47] analyzed personal data and helped to identified meaningful and individual results. They prepared a dataset of location information and activities, and visualized various records. Their results showed that it would be useful to specifying an improved point using personal data. Pärkkä et al. [48] built an online system that automatically collected five daily activities (lying, sitting/standing, walking, running, and cycling) by sensors. They improved their personalization algorithm which performed with higher classification accuracy than other existing algorithm's.

3.2.3 Position of this study

In this study, we consider two aspects of feature extraction with regard to the individual ' s whereabouts. First, we try to understand personal living patterns from behavior patterns. Second, we aim to analyze the location (e.g., transportations and destinations) and environment data (e.g., weather, temperature, and humidity) to construct the whereabouts model.

3.3 Personal Whereabouts Model

In this study, we propose a personal whereabouts model focusing on analysis of location data and associated contexts including a variety of information such as purposes, situations, events, personal preferences and subjective impressions/feelings. In the model the term whereabouts is used, by meaning, to express *Ibasho* in Japanese, a wider concept beyond a geographical location. Take a train/subway station as an example. Someone may work at the station. Another one may go by for commuting almost every weekday. Someone else may just go for a trip occasionally. Due to different purposes and situations, their impressions and feelings may be quite different.

Fig. 3.1 shows four major components of personal whereabouts model: situation analysis, pattern extraction, user profile, and whereabouts attributes. First, the situation analysis is carried out to extract an individual's life pattern using four types of data: environment, schedule, transportation, and destination. The environment data, such as weather, has the relation with transportation. The schedule represents a predetermined schedule, and affects the transportation and destination. The transportation, such as train, represents what kind of vehicles a user uses when he/she moves to a destination. The situation is analyzed based on these data, and the result is sent to the pattern extraction module.

To extract patterns, the user profile is used. Attributes of personal data influence the characteristics of the personal life, such as occupation. Characteristics of an individual are used as the items related to behavior patterns of a person. The pattern extraction utilizes the results of situation analysis and attributes of user profile. The module of whereabouts attributes is based on the extraction of behavior patterns, such as contexts and events. Contexts refer to the set of facts or circumstances that surround a situation or event[49]. Events represent things which happened in a given place and time, such as exhibitions, sports contest, music concert, festival, and gathering with friends, etc.

3.4 Experiment and Discussion

3.4.1 Data Acquisition

We adapted the decision tree which is a typical method of machine learning. Recent studies explained that the decision tree can be used as an efficient way to classifying user's actions. Similar to these studies that use sensors or smartphones to collect user's activity data[40, 48], we collected the data using a smartphone application, Moves [50], to construct a data set.

The data collection period is 1417 days from April 25, 2014 to March 12, 2018. The subject is one male, and his age is 40's. He is an office worker and generally works from 8 to 17 o'clock on

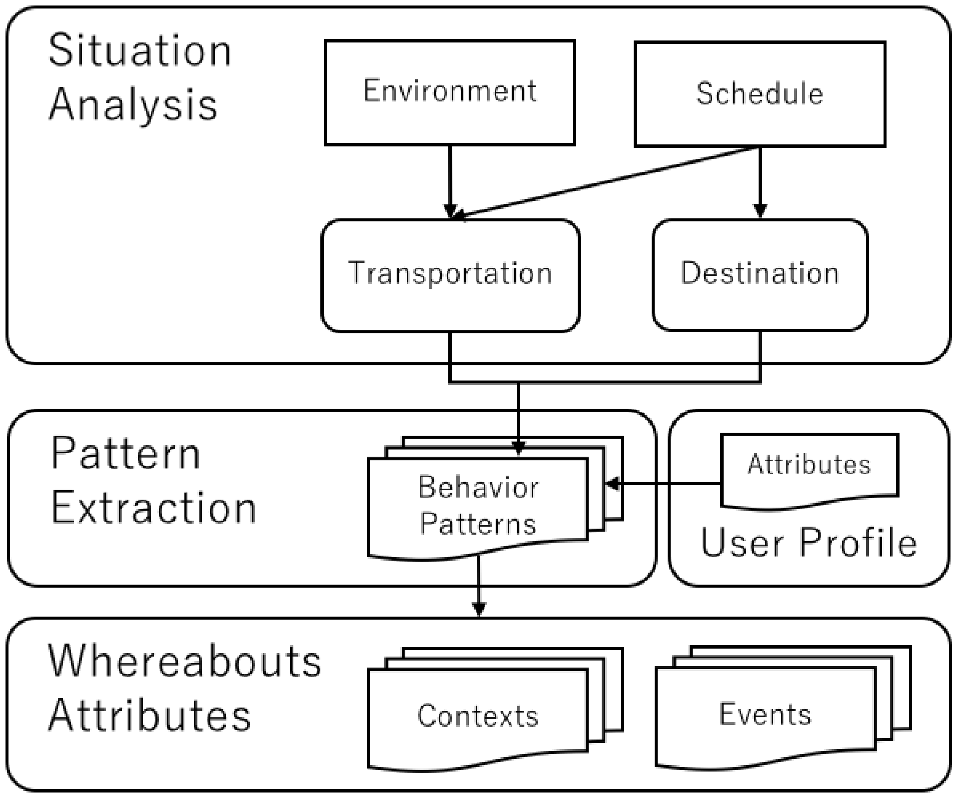

Figure 3.1 Major Components of Personal Whereabouts Model

Table 3.1 Location Data

Data	Data Type
Location Category	Text
Location Name	Text
Starting Staying Time	Timestamp
Ending Staying Time	Timestamp
Latitude	Numerical
Longitude	Numerical

Table 3.2 Transportation Data

Data	Data Type
Status Type	Text
Transportation Name	Text
Starting Moving Time	Timestamp
Ending Moving Time	Timestamp

weekdays. He sometimes goes on a business trip to the branch office and works overtime. His way of commuting is mainly a bicycle on sunny days.

Table 3.1 shows the attributes of whereabouts collected by Moves. Table 3.2 shows the attributes of transportation collected using Moves. We got the weather data [30] in his residential area from the meteorological institute. Table 3.3 shows the attributes of weather collected from the meteorological institute.

Fig. 3.2 shows the storyline which represents the movements in one day. In this storyline,

Table 3.3 Weather Data

Data	Data Type
Date	Date
Humidity	Numerical
Average temperature	Numerical
Precipitation	Numerical
Sunshine hours	Numerical
Average wind speed	Numerical
Average vapor pressure	Numerical
Weather overview (Day)	Text
Weather overview (Night)	Text

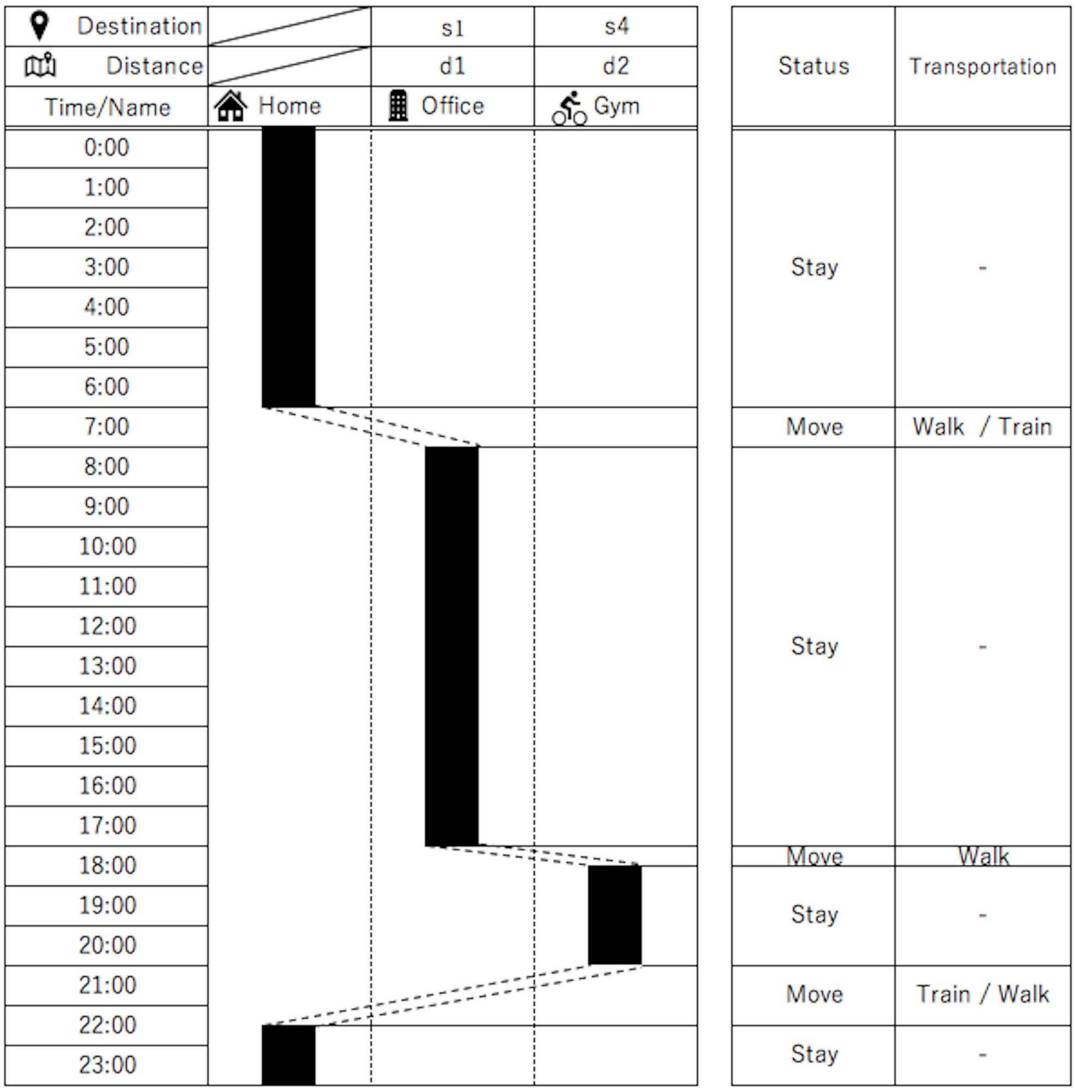

Figure 3.2 Storyline for the Movements of a Person One Day

the start time and staying time of movements are recorded in minutes. In the case of stay, the destination data is acquired. In the case of move, the transportation data is acquired. The transportation has some labels, such as walking, cycling, car, train, bus, etc. In addition, transportation is automatically recognized and recorded by Moves. The destinations have their location names. Moreover, building names and landmark names based on GPS are automatically recognized.

3.4.2 Preprocessing and Extraction of Factors Related to Transportation and Destination

The measurement of the traveling time starts from the end of the staying time and stops when it arrives at the destination. In this study, the dataset was created by dividing one day into one-hour units. The missing value was 145 days from January 18, 2016 to June 11, 2016. We removed the missing values. As a result, we got 1370 records in 688 days for the experiment.

We recorded only the last destination even if the person visited several places within an hour.

We attached two labels: weekdays and holidays. The weather data of the one person's residential area was acquired, and we labeled it with sunny, cloudy, and rainy. The numerical data of average temperature and humidity of the day was obtained.

After preprocessing, the dataset is organized into four types of data: transportation on weekdays, transportation on holidays, destination on weekdays, and destination on holidays. We use Weka* to create a decision tree[51] and analyze this dataset.

3.4.3 Analysis Result Based on Decision Trees

Table 3.4 shows the results of transportation classification. It revealed that transportation was used at 6 o'clock for commuting and used from 17 to 23 o'clock after leaving work. On holidays, we found that several transportations were used.

Table 3.5 shows the results of destination classification. The destination was confirmed at 6 o'clock in commuting and used from 17 to 23 o'clock after leaving work. It is not clear that the destination was associated with a periodical place on holidays. At a destination was according to the schedule for each holiday, respectively.

* https://www.cs.waikato.ac.nz/ml/weka/

Table 3.4 Factors of Transportation

Time Label	1st Factor		2nd Factor		3rd Factor		4th Factor	
	Weekday	Holiday	Weekday	Holiday	Weekday	Holiday	Weekday	Holiday
t6	Temperature	-	-	-	-	-	-	-
t7	Humidity	-	-	-	-	-	-	-
t8	Temperature	Temperature	-	-	-	-	-	-
t9	-	-	-	-	-	-	-	-
t10	Humidity	Temperature	Temperature	Humidity	-	Temperature	-	-
t11	-	Temperature	-	Humidity	-	-	-	-
t12	-	Temperature	-	Weather	-	Temperature	-	-
t13	Temperature	Temperature	Weather	Weather	Humidity	Humidity/Temperature	-	-
t14	Weather	Temperature	Humidity	Humidity	Temperature	Weather	-	Temperature
t15	-	Temperature	-	Humidity	-	Weather/Temperature	-	-
t16	-	Temperature	-	Weather	-	Humidity	-	Temperature
t17	Weather	Temperature	Humidity	Weather/Humidity	Temperature	Humidity/Weather/Temperature	Humidity/Temperature	-
t18	Weather	Temperature	Humidity	Weather/Humidity	Temperature	Humidity	Humidity	-
t19	Weather	Temperature	Humidity	Weather/Humidity	Humidity/Temperature	Humidity/Temperature	Humidity/Temperature	-
t20	Humidity	Temperature	Weather/Temperature	Humidity	Temperature	-	-	-
t21	Temperature	Temperature	Weather	-	Humidity	-	Temperature	-
t22	Temperature	Temperature	Weather	-	Temperature	-	Humidity	-
t23	Humidity	-	-	-	-	-	-	-

Table 3.5 Factors of Destination

Time Label	1st Factor		2nd Factor		3rd Factor		4th Factor	
	Weekday	Holiday	Weekday	Holiday	Weekday	Holiday	Weekday	Holiday
t6	Temperature	-	-	-	-	-	-	-
t7	-	-	-	-	-	-	-	-
t8	Temperature	Temperature	-	Humidity	-	-	-	-
t9	Temperature	Weather	Temperature	Temperature	-	-	-	-
t10	-	Weather	-	Humidity/Temperature	-	Humidity/Temperature	-	-
t11	-	Weather	-	Humidity/Temperature	-	Humidity/Temperature	-	-
t12	-	Weather	-	Humidity	-	Temperature	-	-
t13	-	Weather	-	Humidity/Temperature	-	Humidity/Temperature	-	-
t14	-	Weather	-	Humidity	-	Temperature	-	-
t15	Weather	Weather	Temperature	Humidity/Temperature	Temperature	Humidity/Temperature	-	-
t16	Weather	Weather	Temperature	Temperature	Humidity/Temperature	Humidity/Temperature	Temperature	Humidity/Temperature
t17	-	Weather	-	Humidity/Temperature	-	Humidity/Temperature	-	-
t18	Humidity	Weather	Humidity/Weather	Humidity/Temperature	Humidity/Temperature	Humidity/Temperature	Humidity/Temperature	Humidity/Temperature
t19	Weather	Weather	Humidity/Temperature	Temperature	Humidity/Temperature	Humidity/Temperature	Temperature	Temperature
t20	Weather	Weather	Humidity/Temperature	Humidity/Temperature	Humidity/Temperature	Humidity/Temperature	Humidity/Temperature	-
t21	Weather	-	Humidity/Temperature	-	Humidity/Temperature	-	-	-
t22	Temperature	Weather	Humidity/Temperature	-	Humidity/Temperature	-	-	-
t23	Temperature	-	-	-	-	-	-	-

Table 3.6 Accuracy of Personal Data Classification

	Transportation		*Destination*	
Day of week	*Weekday*	*Holiday*	*Weekday*	*Holiday*
Correctly classified days	51.52%	41.20%	40.70%	5.59%
Incorrectly classified days	48.48%	58.80%	59.30%	94.41%
Total number of days	887	483	887	483

3.4.4 Discussion

In this study, we carried out investigation using the environment data such as weather, temperature, and humidity. Furthermore, we analyzed and extracted living patterns in each hour of the day to understand a specific user' s living habits.

Table 3.4 represents the living patterns of transportations on weekdays and holidays. The results demonstrate that the commute time on weekdays appeared to be regular, which resulted in a small number of living patterns according to the transportation on weekdays. For instance, we found a regular pattern that this subject goes home during 18 to 22 o'clock on weekday. On the other hands, the results demonstrate that transportations are utilized in different time period on holidays according to different schedules, which resulted in various living patterns according to the transportation on holidays.

Table 3.5 represents the living patterns of destinations on weekdays. Similar to the living patterns of transportations, the results demonstrate that regularity on weekdays and diversity on holidays. Furthermore, it is noted that to extract more personal living patterns, we need to pay more attentions to activities which are in the early morning or after leaving work on weekdays, and during the whole day on holidays.

Table 3.6 shows the accuracy of classifications of transportation and destination on weekdays and holidays, respectively. The results demonstrate that the classification of transportation on weekdays has the highest accuracy, while the classification of destination on holidays has the lowest accuracy.

3.5 Summary

In this study, we proposed a personal whereabouts model to analyze personal data more effectively. The location and environment data was collected and analyzed to extract personal living patterns.

We introduced and constructed a personal whereabouts model with four components, including situation analysis, pattern extraction, user profile, and whereabouts attributes. A decision tree based method was used to organize and analyze the collected personal data, in order to extract living patterns. We presented the experiment design, and discussed the observations based on the extracted living patterns according to different transportations and destinations. The results demonstrated the feasibility of our proposed method.

For our future work, we will analyze and evaluate the behavior patterns based on personal whereabouts model from various perspectives. We will develop a system to analyze the lives of individuals for better understanding.

Chapter 4

Measurement and Quantification of an Individual's Feelings about a Place[†]

4.1 Background

With the widespread use of smartphones and tablet computers, people often search for nearby places from their current location. Such devices not only search for places, but also record locations and activity history, and even share real-time emotions using social networking services (SNS). Therefore, there are many kinds of location information in daily life.

An individual may have special emotions for a place. For example, cafés are associated with relaxing such as in one's own home; restaurants are perceived as places where meals can be eaten with family members or friends; terraces evoke reminiscences of natural light and breeze. In this way, when we imagine a place, we associate it with certain scenes and impressions.

In Japanese, the word "Ibasho" that includes not only location information but also one' s feelings about a place. For example, one day person may have had dissatisfaction because of work. On the way home, that person may enter a bar and talk with people. By the time they leave the bar, their feelings are refreshed. Many people have such an experience and, if such an experience

[†] ©2021 John Wiley & Sons, Inc. Reprinted, with permission, from S. Kasuya, K. Tago and Q. Jin, "Measurement and quantification of an individual's feelings for a place in personal data analysis," Human Behavior and Emerging Technologies, Vol. 3, No. 5, pp. 739–749, 2021.

leads a person to go to a particular place where they can change their feelings, that place develops a special association in that person’ s mind.

From a sociological viewpoint, where a person builds and maintains the connection with others can be divided into two categories: social and personal. Ibasho has demonstrated location information including social and personal feelings. However, a conceptual definition has not been established. Kasuya et al. [39] proposed a framework of well-being-oriented life support to analyze personal living habits. We used the data obtained from a single person and multiple the data related to that person.

In this study, we used data obtained from a person and multiple data related to that person to show that the data from a single person can be used for meaningful analysis. For example, personal data such as location information, meteorological data, daily schedule, physical condition are not all the same. By analyzing these day-to-day differences and relationships, a more personalized analysis can be conducted.

Many approaches have been proposed to understand the emotions in a tweet. However, it is difficult to quantify emotions, particularly those related to a place. If we can quantify such emotions, we can better understand the lives of individuals and provide more personalized services. In addition, by introducing the metaphor of place, support becomes more active, and interaction is enhanced.

In this study, we measure emotions in tweets related to a place and evaluate the measurement methods. To find an appropriate method for the emotion quantification, an evaluation model is created. Furthermore, we design four protocols in the evaluation model that provide results representing the feature of the place. Subsequently, we evaluate how well the results match the poster’s subjectivity and identify suitable methods for quantifying feelings about a place. The matching results are verified by statistical methods.

The contributions of this study are summarized as follows. We identify an appropriate method for quantifying the feelings of a place using three methods: emotion dictionary, personalized dictionary, and Bayesian classification. This allows us to improve the analysis of personal data including locations.

4.2 Related Work

4.2.1 Study of a place

It is a general observation that a person regularly visits approximately 25 places every day [52]. It is shown that the places a person prefers to visit are related to social interactions. Hansan and Ukkusuri [53] modeled selection patterns for an individual' s activity. Their proposed system complements the geographic location information using social media data.

"Meaning place" research focuses on places of personal significance. One of the challenges of this research area is to associate special feelings with a physical location and its features. Zhou et al. [54] demonstrated that their algorithm discovers places with reasonable accuracy and outperforms the well-known K-Means clustering algorithm for place discovery. In addition, they provided evidence that shapes more complex than "points" are required to represent the full range of people' s everyday places.

Most of the previous studies in this area generally evaluated behaviors occurring at home [55]. Li and Tong focused on behaviors outside the home and created a new model that can evaluate activities related to places with multipurpose positions. Bentley et al. [56] used SNS check-in services, which can record when a person enters and leaves a place, to collect data about where people go on a daily basis such as shops, food stores, and schools. They found that parks and outdoor locations are important places for families. However, their study did not clarify how to capture the changing emotions associated with a place . They also investigated the categories of important places and proposed a video-sharing system to be used for meaningful places.

To predict the popularity trend of newly opened places, D'Silva et al. [57] analyzed the relations between the place and the temporal characteristics of nearby locations using SNS data. Their results show that the concept of locality is helpful for predicting trends. Their approach counts the number of people stopping at a place, then predicts the number of people based on the data. However, their study focused only on the concentration of people at different times of the day throughout the week, and thus, the kind of emotional changes occurring at that location were not investigated.

Herckis et al. [58] verified a design using three virtual spaces to realize the place to be a house in the virtual space. The results clarified that the feature ”space” is important in the communication environment, and the designer supports the interaction through the spatial model. This information provides a starting place to build meaningful interpersonal relationships in virtual space. These results suggest that a location can have a specific role for an individual. Therefore, a place not only shows information about locations and coordinates, but also has special roles and causes certain feelings. Harrison and Dourish [59] stated that the actual framework of interaction lies in the ”field”. Subsequently, the environment is divided and considered in view of the “field” and “space” [60].

4.2.2 Research on feelings and sentiment analysis

Kumarasiri and Farook [61] proposed a mobile-based intelligent system that combines natural language processing (NLP) and aspect-based opinion mining (ABOM) techniques to assist user-centric decision-making in restaurant selection.

They measured decision-making accuracies and variations over three main areas: aspect extraction, aspect classification, and sentiment allocation. The overall results were acceptable . However, unsupervised methods always outperform supervised methods during sentiment allocation. Hollis et al. [62] proposed a support system that promotes well-being and analyzed interview responses about feelings. They showed the importance of self-recognizing negative moods using visualized past mood data and thinking of ways to improve the mood.

It has been reported that urban regeneration within traditional settings has transformed places and constructed meanings that were embedded in the existing social and cultural settings [63]. From the results of a questionnaire, they concluded that the importance of place in urban regeneration for psychological well-being lies in its psychological aspect and identity. In another study by Ujang and Zakariya [64], they stated that an attachment to a place occurred based on the connection between people and the environment. The results of an interview survey supported their findings. The survey reported that the role of a user and their ethnic background influence their response to a place.

Dragut et al. [65] found a complex polarity mismatch in the polarity classification in many emotional word/semantic dictionaries. They proved that this mismatch is an NP-complete problem. Based on their finding, they proposed the concept of polar consistency of emotional word dictionaries and showed that inconsistencies between multiple dictionaries and within dictionaries satisfy polar consistency.

Despite having five emotional word dictionaries and WordNet when evaluating emotions, inconsistencies occur in which evaluations are divided, making it an NP-complete problem. López and Cuadrado-Gallego [66] compared the use of various supervised learning methods for sentiment analysis. They found that TF-IDF, which is a method for evaluating the importance of words, is often used in sentiment analysis. Their results also showed that the best method was naive Bayes, followed by perceptrons. Furthermore, they reported that neural networks can perform accurate binary prediction, although neural networks are computation-intensive for parameter processing, training, and fitting.

Jia and Li [67] calculated a vector of emotional words from a large microblogging text and classified emotions. To improve the classification accuracy, they changed the calculation method of the intensity of emotion words in the emotion dictionary and applied the degree of negative words and the influence of adverbs to the rules of semantic calculation. Tago et al.[68] and Tago & Jin [69] analyzed positive and negative users' tweets. They used naive Bayes to solve the problem of emotion scores not being measured for words that are not included in the emotion dictionary. They also stated that positive users' follow and follower relationships increase just as in the real world.

Mu et al.[70] implemented a system that efficiently detects 5W1H (Who, When, Where, What, Why, and How) as events from microblogging. The system performs step-by-step sentiment analysis for events, making it possible to know the change in people's emotional tendencies toward the event.

In recent years, emotion prediction of multimodal microblogging consisting of text, images, and pictograms has attracted more attention. Predicting emotion changes caused by different events involves label noise, thereby causing problems. Chen et al.[71] introduced a weakly supervised multimodal deep learning (WS-MDL) scheme for robust and scalable sentiment prediction. Quantitative evaluation demonstrated excellent performance of this method compared with the latest and alternative methods.

4.2.3 Position of this study

Most of the findings using location information focus on users and capture the characteristics of the specified location on the time axis. Various sentiment analysis studies have been conducted. However, only a few have analyzed the attachment quantitatively. Ujang and Zakariya[63, 64] studied place attachment, but their study only involved a survey of regional attachment concerning urban planning.

In this study, we quantitatively analyze the feelings toward a place using several previously proposed sentiment methods. In addition, as the methods appropriate for quantifying the emotion toward a place are not clear, we evaluate and identify an appropriate method.

4.3 Quantification of Personal Feelings for a Place

In this section, we first describe the definition and concept of *"Ibasho"*. Next, our experimental design to quantify the feelings about an individual's *"Ibasho"* is explained. Then, we present our method of calculating emotional values from tweets. Furthermore, in this study, we adopt three methods: Emotion dictionary, Personalized dictionary, and Bayes classification. We compare the three methods and evaluate which method is more appropriate to quantify the feelings for a place.

4.3.1 Definition of a place

In Japanese, the word *"Ibasho"* is an abstract concept that means not only physical location information but also one' s feelings associated with a place. In this study, the definition of *"Ibasho"* is a place, space, and a place of mind for oneself.

The concept of “*Topophilia*” was introduced as the affective bond with one's environment—a person's mental, emotional, and cognitive ties to a place, which is related to the spatial segmentation of consciousness development and self-realization [72, 73]. In addition, he noted that humans tend to segment, systematize, and organize phenomena into opposition terms. For example, in traffic lights red signals danger, and green signals safety. *"Topophilia"* as referring to the human spirit and emotional and cognitive relationships between the structures of the human mind, has been studied as a potential human component. The potential components are abstract concepts consisting of psychological attitudes or intellect, which can be measured. Tuan originally defined *"Topophilia"* as the affective bond between people and place or environmental settings.

In this study, the feelings for a place indicates the special emotions when a person visits or associates with the place. The place that leads to the feeling is a space to achieve one' s purpose or for one' s peace of mind. The difference between *"Topophilia"* and *"Ibasho"* as defined in this study is as follows: *"Topophilia"* is broad and comprehensive while *"Ibasho"* is evaluated considering an individuals' feelings of place concretely and quantitatively.

4.3.2 Quantification methods and evaluation model

To extract feelings for a place from a tweet, we calculate emotional values. In this study, we evaluate which method is more appropriate for quantification. The quantification methods are an emotion dictionary [74], a personalized dictionary, and the Bayesian classification, which are the representative approaches for sentiment analysis. In this study, a personalized emotional word dictionary is created by using individual tweets as a corpus. For the emotion dictionary and the Bayesian classification, we apply the same approaches used in previous related studies [68, 69].

Next, we collect tweets when a person enters and leaves a place. Then, the difference between the emotional values is verified based on four protocols using the emotional values calculated by each of the three methods. In protocol one, raw emotion scores are used. In the other three protocols, emotion scores are normalized. If the emotional value of a new place is provided, we can compare it to previous places by normalizing the emotion scores. Finally, subjective evaluation is carried out for each result of all the protocols, and we investigate which method performs better at quantifying the feelings for a place. Figure 4.1 shows the experiment design.

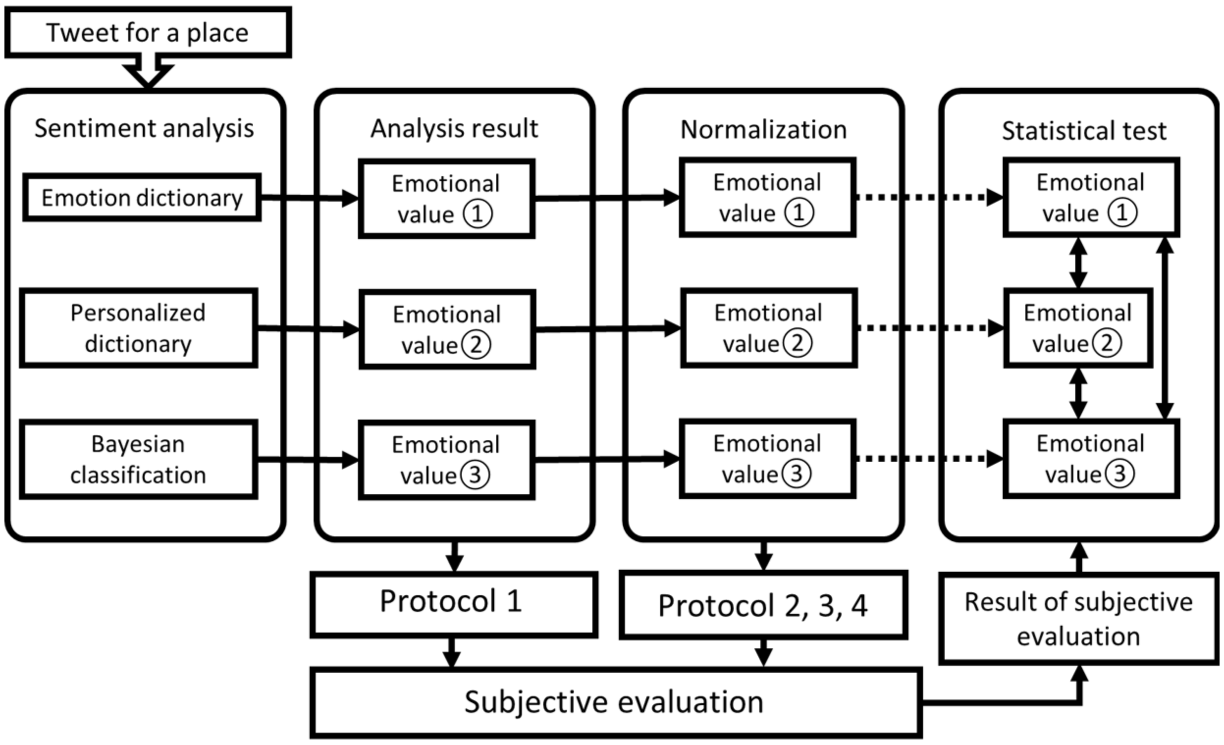

Figure 4.1 Experimental Design

```
優れる(sugureru):すぐれる:動詞:1^  superior:superior:verb:1
良い(yoi):よい:形容詞:0.999995^ good:good:adjective:0.999995
喜ぶ(yorokobu):よろこぶ:動詞:0.999979^  rejoice:rejoice:verb:0.999979
褒める(homeru):ほめる:動詞:0.999979^compliment:compliment:verb:0.999979
めでたい(medetai):めでたい:形容詞:0.999645^ congratulate:congratulate:adjective:0.999645
賢い(kashikoi):かしこい:形容詞:0.999486^smart:smart:adjective:0.999486
```

Figure 4.2 Part of the Dict Ed

4.3.3 Emotion dictionary

To evaluate emotions, many emotion dictionaries have been proposed. In this study, we use the emotional polarity dictionary [74] (hereafter, Dict Ed). The emotional polarity dictionary has 55,124 words (49,983 negative words, 5,122 positive words, 20 neutral words). Using the Dict Ed, Tago et al. [68] and Tago & Jin [69] carried out keyword matching. Using MeCab [75], a morphological analysis application, a sentence is separated on a word by word basis. If the words match the words in the emotional polarity dictionary, the score in the dictionary is used for a calculation. Figure 4.2 shows a part of Dict Ed.

The calculation of emotion scores is expressed by Equation (4.1)

$$
ESTEd_i = \begin{cases} s \left| \begin{array}{l} \\ -1 \leq s \leq +1 \\ \end{array} \right. & \\ \quad \text{if a tweet } T_i \in \text{Dict Ed} & \\ & \\ null \quad \text{otherwise} & \end{cases} \tag{4.1}
$$

where *ESTEd* represents the emotion score of word *i*. If the word is in the emotion dictionary, a score is given.

The emotion score of a tweet, *ESTEd*, is the sum of all the emotional words in the tweet. *ESTEd* can be calculated by using the following Equation (4.2):

$$
\begin{gathered} ESTEd = \sum_{i=1}^{|ESTEd_i|} ESTEd_i \\ (\text{If } ESTEd_i \text{ is null, it is regarded as } 0) \end{gathered} \tag{4.2}
$$

If there are no emotional words in a tweet, the emotion score *ESTEd* equals zero.

4.3.4 Personalized emotion dictionary

The personalized dictionary (hereafter, Dict Pd) is created based on one's tweets. To create the Dict Pd, morphological analysis is performed. Nouns, adjectives, verbs, adverbs, interjections, and conjunctions are selected. Then, these words are registered to the dictionary. The selected words are scored by the person who posted the tweets on an 11-point scale. The most negative word is scored as -5 and the most positive is scored as +5. The set of the word and score pairs are used as a Dict Pd.

In this study, we used a single person's tweets for this evaluation. The person evaluated the words in his tweets and created the dictionary. Subsequently, the Dict Pd had 1,460 words. Table 4.1 shows the scores and numbers of personalized dictionaries, and Figure 4.3 is part of a Dict Pd.

Table 4.1　Number of scored words

Score	Number
-5	13
-4	28
-3	47
-2	54
-1	47
0	275
1	247
2	382
3	223
4	109
5	35
Total	1,460

```
あえなく(arienaku)^ -1^ dare -1
ありがたい(arigatai)^   5^  thanksful^   5
いい(ii)^    5^  good^    5
うまく(umaku)^   4^   well^    4
じれったい(jirettai)^    -4^ ^    impatient -4
```

Figure 4.3　Part of the Dict Pd

The emotion scores are calculated by matching the keyword to the same word in the Dict Ed in the Dict Pd. The emotion score of the word $ESTPd$ can be calculated by using Equation (4.3):

$$ESTPd_i = \begin{cases} s \,\Big|\, (-5, -4, -3, ..0, .., 3, 4, 5) \\ \qquad \text{if a tweet } T_i \in \text{Dict Pd} \\ \\ null \quad \text{otherwise} \end{cases} \tag{4.3}$$

where $ESTPd$ shows the sum of all the emotional word scores in the Dict Pd.

The emotional score of the tweet $ESTPd$ is the sum of all the emotional words in the tweets in the Dict Pd. $ESTPd$ is expressed by Equation (4.4):

$$ESTPd = \sum_{i=1}^{|ESTPd_i|} ESTPd_i \tag{4.4}$$

$$(\text{If } ESTPd_i \text{ is null, it is regarded as } 0)$$

If there are no emotion words in a tweet, the emotion score $ESTPd$ equals zero.

4.3.5 Bayesian classification

Training data are needed to perform Bayesian classification. To perform naive Bayes, the training data were created by separating tweets obtained from a person who posted tweets about a location. A sentence was categorized as positive (P), negative (N), or neutral (Nt). During the training data-creation process, the poster of the tweets and another person labeled the tweets as P, N, or Nt independently. If a sentence had different labels, a third party labeled it, and the category was decided by a majority vote.

Using a trained naive Bayes model, the category scores of words for each category (P, N, and Nt) were calculated. Subsequently, the scores were summed for each category. Finally, the category with the maximum score was selected.

4.4 Experiment and Discussion

4.4.1 Experiment Design

We investigated the appropriate method for evaluating the feelings associated with a place. The experiment was designed based on multiple conditions to evaluate the three sentiment analysis methods that we used in this study. After carefully examining these methods and conditions, we conceived four protocols:

Protocol 1 focuses on the evaluation based on the emotional score for a location when a person enters and leaves. Protocol 2 uses the maximum and minimum ratio of emotional scores after being normalized to a range from zero to one. Protocol 3 focuses on the ranking of 13 selected locations based on the normalized emotional scores. Protocol 4 uses the top 10 locations in descending order in terms of the difference of normalized emotional scores between entering and leaving a location. Further details of the four protocols are as follows.

Protocol 1

In this protocol, the feeling score of a place (FSP) is obtained using the three measurement methods stated previously. Using FSP, the change in the emotion score at a place is evaluated by the following method.

We represent the emotion score when entering a place as FSPIn, and that when exiting as FSPOut. If FSPIn is higher than zero and FSPIn is higher than FSPOut, we denote the number as FSPH. In contrast, if FSPIn is less than zero and FSPOut is less than FSPIn, we denote the number as FSPL.

Protocol 2

First, emotion scores obtained by the three methods are normalized between zero and one, and the emotional values at 13 locations are targeted. Second, we determine how much the minimum and maximum emotion scores of 13 locations account for the minimum and maximum emotion scores of each location. The emotional score when entering a place is less than the score at the time of exit, $(FSPIn) < (FSPOut)$.

We then give the higher label ($FSPH \leftarrow FSPOut - FSPIn$>0). Next, we let *max (emotion score on the higher label (H)) - min (emotion score on the higher label (H))* be the denominator.

For the emotional score of each place, the higher label is assigned to $(FSPIn)$ < $(FSPOut)$. Max (emotion score on the higher label (H)) - min (emotion score on the higher label (H)) is used as a numerator, and the ratio is calculated. The emotion score of tweets from a higher label is denoted by FSPH.

Equation (4.5) is the emotional value of the higher label (FSPH) of Protocol 2.

$$FSPHratio = (max(FSPHp) - min(FSPHp))/(max(FSPHP) - min(FSPHP)) \tag{4.5}$$

For the emotional scores of all places, the lower label ($FSPL \leftarrow FSPOut - FSPIn < 0$) is given to the emotional value when $(FSPIn) > (FSPOut)$. We then let *max(emotion score of the lower* *label (L)) - min(emotion score of the lower label (L))* be the denominator.

Next, for the emotional score of each place, $(FSPIn)$ > $(FSPOut)$ is assigned a lower label. *Max (emotion score of the lower label (L)) -min (emotion score of the lower label (L))* is the numerator and its ratio is calculated. Accordingly, the emotion score of tweets from a lower label is denoted by FSPL.

Equation (4.6) is the emotional value of the lower label (FSPL) of Protocol 2.

$$FSPLratio = (max(FSPLp) - min(FSPLp))/(max(FSPLP) - min(FSPLP)) \tag{4.6}$$

Protocol 3

In this protocol, the ranking of 13 places is created. First, the emotional values obtained by each of the three methods are normalized between zero and one and the higher label (H) is given to $(FSPIn) < (FSPOut)$.

The emotional values are obtained as the higher labeled values. The protocol then finds the average value (FSPOutH- FSPInH) of the difference between the time of entering the place and the time of leaving the place. The average value, *Average (FSPOutH - FSPInH)*, is ranked from 1st to 13th in descending order. Similarly, the lower label ($FSPL \leftarrow FSPOut{-}FSPIn < 0$) is assigned to $(FSPIn) > (FSPOut)$. The protocol obtains the emotional values as the lower labeled values (FSPL), and finds the average value *(FSPOutL - FSPInL)* of the difference between the time of entering the place and the time of exiting. The obtained average value *Average (FSPOutL - FSPInL)* is ranked from 1st to 13th in ascending order. Next, the average value of *(FSPOut - FSPIn)* is calculated without considering the higher label and the lower label. The difference, *Average (FSPOut - FSPIn)*, obtained from the obtained average value is ranked from 1st to 13th in descending order.

Protocol 4

In this protocol, the top 10 places in which the difference in emotion scores is in descending order are created. First, the emotional values obtained by the three methods are normalized between zero and one. $(FSPIn)$ < $(FSPOut)$ is labeled as $(FSPH \leftarrow FSPOut - FSPIn\ >\ 0)$. Among the emotional values labeled above (FSPH), the top 10 locations are listed in descending order based on the difference between scores when entering the location and when exiting.

Expression (4.7) shows that the set S is sorted in descending order,

$$S_i \geq S_i + 1\ for\ all\ 1 \leq\ i < n \tag{4.7}$$

In the same way, the lower label $(FSPL\ \leftarrow\ FSPOut\ -\ FSPIn\ <\ 0)$ is assigned to $(FSPIn)$ > $(FSPOut)$.

Among the emotional values labeled below (FSPL), the top 10 locations are listed in ascending order of the difference between (FSPIn) and (FSPOut).

Expression (4.8) shows that the items are sorted in ascending order,

$$S_i \leq S_i + 1\ for\ all\ 1 \leq\ i < n \tag{4.8}$$

Using the results obtained by the four protocols, the three methods of emotion score extraction (Dict Ed, Dict Pd, and Bayesian classification) are evaluated. Then, a subjective evaluation is carried out for the results to investigate which method is suitable for representing the feelings for a place.

4.4.2 Experimental data

We used the iPhone app Moves to collect personal data about locations. The test subject posting the tweets was a 45 year old man working full time at a manufacturing company. He commuted by car from Monday to Friday and by bicycle twice a week for maintaining his health. His tweets were collected for nearly five months (145 days) from March 1st to July 31st, 2018, with 545 records

Table 4.2 Target locations for extracting and quantifying emotion scores.

No	Category of places	Visited count
1	Home	147
2	Office A	89
3	Family home	66
4	Super market A	18
5	Office B	15
6	Shopping center A	13
7	Super market B	9
8	Hospital A	7
9	Hospital B	6
10	Coffee shop	6
11	Shopping center B	6
12	Train station	6
13	Super market C	6

Table 4.3 Summary of the calculated emotion scores

	Higher label (FSPH)			Lower label (FSPL)		
	Min.	Max.	Their Difference	Min.	Max.	Their Difference
Emotion dictionary	0.000	0.602	0.602	0.000	0.597	0.596
Personalized dictionary	0.020	0.540	0.520	0.020	0.720	0.700
Bayesian classification	-0.824	-0.002	0.822	0.000	0.646	0.645

and 71 unique locations.

As the original Twitter dataset was not for evaluation of feelings about a place, we asked the poster of the tweets to recall and add his subjective experience about these locations to create a test dataset for this experiment. In addition, some of the acquired data did not record the name of the place. Thus, the locations were named as unknown. We targeted places visited more than five times because our focus is on frequented places. Consequently, we obtained 394 tweets and 13 places. The target locations for extracting and quantifying emotion scores are shown in Table 4.2.

We quantify emotional tweets posted when entering and leaving a place. Figure 4.4 illustrates the distributions of the emotion scores obtained by each of the three methods.

4.4.3 Emotion score calculation and protocol results

Based on the four protocols, we verified the difference between the methods. Table 4.3 shows the normalized maximum score, maximum score, and difference of the maximum and minimum values of the 13 locations.

To create the naive Bayes model, we first created a training dataset from the tweets. The overall

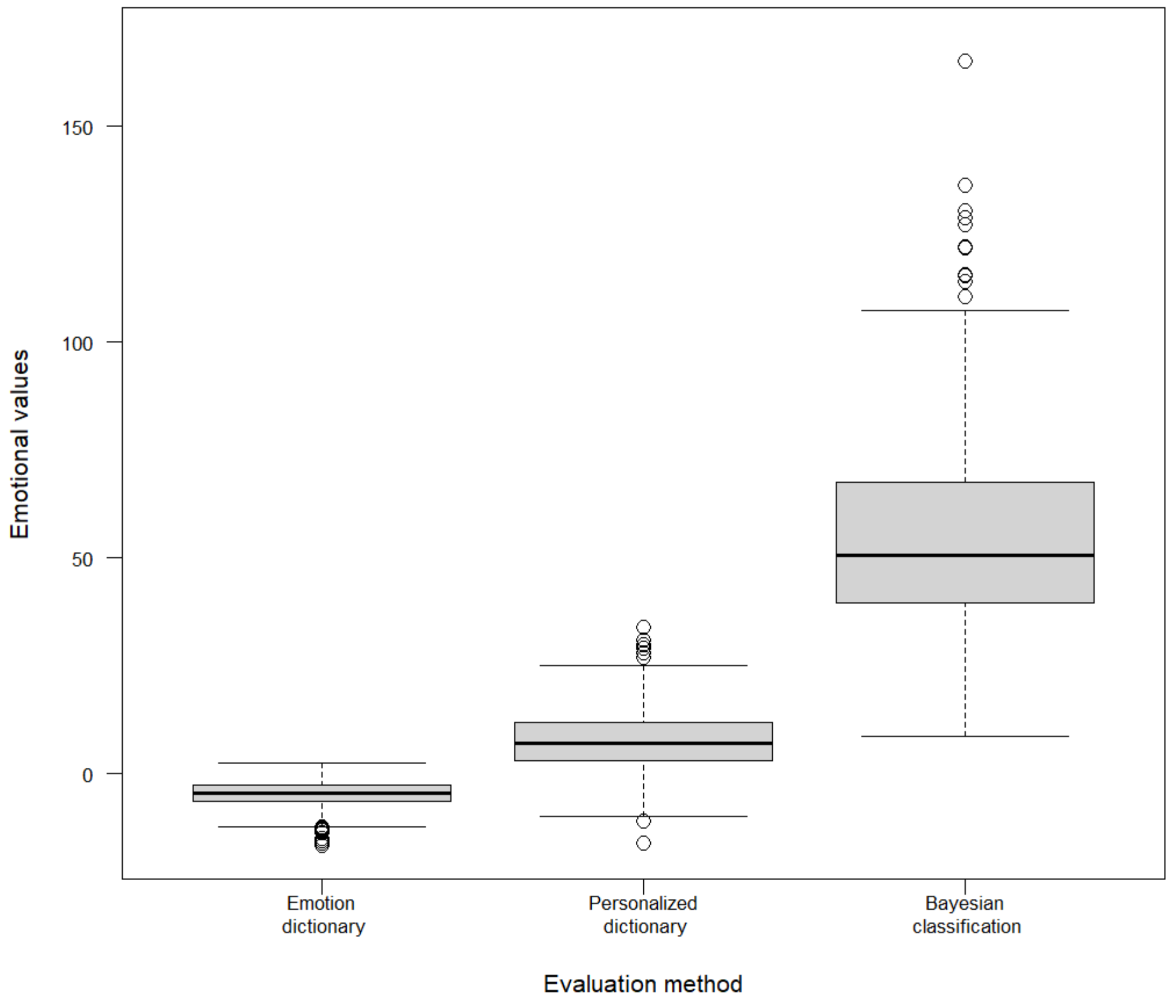

Figure 4.4 Distribution of emotional values of 13 locations

sentiment labels in the Bayesian classification were mostly Nt. Therefore, the values for the Nt data were selected such that the number of Nt cases would be half of P and N [68, 69]. Table 4 shows the number of tweets of each category and its percentage of the total. The number 909 represents the total number of sentences tweeted when entering and leaving the location. The result of five cross-validation is shown in Table 5.

To create the naive Bayes model, we first created a training dataset from the tweets. The overall sentiment labels in the Bayesian classification were mostly Nt. Therefore, the values for the Nt data were selected such that the number of Nt cases would be half of P and N (Tago et al., 2019; Tago and Jin, 2018) [68, 69]. Table 4.4 shows the number of tweets of each category and its percentage of the total. The number 909 represents the total number of sentences tweeted when entering and leaving the location. The result of five cross-validation is shown in Table 4.5.

Table 4.4 Number and percentage of Bayesian training data records

Emotion	Selected number	Ratio
P	427	47%
N	183	20%
Nt	299	33%
Total	909	100%

Table 4.5 Result of five cross-validation

		By Bayesian Classification			Total	Matching Rate
		P	N	Nt		
By the Evaluator	P	351	16	60	427	82.2%
	N	40	104	39	183	56.8%
	Nt	63	29	207	299	69.2%
	Total	454	149	306	909	69.4%

4.4.4 Comparative analysis of results

Subjective evaluation

The subject evaluated the results of the emotional values obtained by the four protocols subjectively. The evaluation was carried out on a five-point scale (5: strongly agree, 1: strongly disagree). The scores were summed for each emotion evaluation method. The higher the score, the more appropriate the method is for the person's subjective evaluation.

Chi-square test

To investigate whether the results of the three emotion evaluation methods have a significant difference, we adopted the chi-square test. The chi-square test is a nonparametric test that can verify whether the emotion score extraction method and the subjective evaluation are independent. If they are not independent, we can conclude that the result of subjective evaluation depends on an emotion score extraction method.

In the chi-square test, a chi-square value and a p-value are calculated. If the p-value is less than 0.05, it can be judged that the result of the emotion score extraction method is related to the subjective evaluation. Conversely, if the p-value is higher than or equal to 0.05, it can be judged that the result of the subjective evaluation and emotion evaluation methods are independent. Table 4.6 shows the results of the chi-square test for the four protocols of subjective evaluation based on the three extraction methods of emotion scores.

Table 4.6 Chi-square test in subjective evaluation and emotion score extraction method

Protocol	Evaluation method	n	Strongly agree	Agree	Nether agree nor disagree	Disagree	Strongly disagree	chi-sq	DF	p-value
	Emotion dictionary	81	12.35	34.57	37.04	14.81	1.23			
Protocol 1	Personalized dictionary	103	29.13	54.37	14.56	1.94	0	105.277	8	0.00000**
	Bayesian classification	115	69.57	20.87	7.83	1.74	0			
	Emotion dictionary	122	8.2	29.51	54.1	8.2	0			
Protocol 2	Personalized dictionary	123	8.85	21.24	50.44	17.7	1.17	15.109	8	0.004351**
	Bayesian classification	137	7.87	34.65	49.61	7.87	0			
	Emotion dictionary	80	12.5	20	52.5	15	0			
Protocol 3	Personalized dictionary	84	5.95	38.1	46.43	9.52	0	21.274	8	0.009505**
	Bayesian classification	81	0	44.44	44.44	9.88	1.23			
	Emotion dictionary	76	19.74	52.63	27.63	0	0			
Protocol 4	Personalized dictionary	78	25.64	51.28	23.08	0	0	NaN	8	0.003956**
	Bayesian classification	88	51.14	45.45	3.41	0	0			

Note: ** $p < .01$.
Abbreviations: DF, degree of freedom; NaN, not a number.

As shown in Table 4.6, in the subjective evaluation of the number of +In++Out and -In—Out (Protocol 1), a significant difference was found at the 5% level (x2(8)=105.277, p<.05). Table 4.6 shows that there is a significant difference in the subjective evaluation of the higher and lower proportions (Protocol 2) (x2(8)=15.109, p<.05). In the subjective evaluation of the higher and lower rankings (Protocol 3), we find that there is a significant difference in the 5% level (x2(8)=21.274, p<.05). Finally, the top 10 (Protocol 4) subjective evaluations show a significant difference (x2(8)=NaN: Not a Number, p<.05). In the subjective evaluation by the four protocols, the p-value is less than 0.05, hence we can conclude that the results of the emotion score extraction method are related to the subjective evaluation.

The results of the subjective evaluation in Table 4.6 are listed in order from left to right. The findings show that Bayesian classification is the best overall. However, the results differ depending on the protocol (from left to right). Protocol 1 is ranked on Bayesian classification, Dict Pd, and Dict Ed. Protocol 2 is ranked on Bayesian classification, Dict Ed, and Dict Pd. Protocol 3 is ranked on Dict Pd, Bayesian classification, and Dict Ed. Protocol 4 is ranked on Bayesian classification, Dict Pd, and Dict Ed.

4.4.5 Discussion

Related work does not clearly indicate which emotion evaluation method is appropriate for the special feelings at a particular location. Therefore, we evaluated the feelings for the location through our evaluation model.

Bentley et al. [76], who developed a system of personally important places, defined them as home, workplace, grandma' s house, and public venue. Ujang and Zakariya [63, 64] also aggregated the degree of attachment to a place using a questionnaire.

Compared with these studies, we calculate emotion scores quantality and evaluate emotion evaluation methods considering whether the subjective evaluation and the emotion are consistent. Regarding the extraction of emotion scores by the Dict Ed, the number of correspondences between the corpus of the Dict Ed and the words of the personal dictionary was fewer. This confirms that our vocabulary was fewer. In addition, as most of the emotional scores in the Dict Ed were negative,

negative emotional values appeared more often in the results. If we use more positive words in the Dict Ed, we can expect the score to improve for the Dict Ed.

Regarding the Dict Pd, a person who posts tweets scores emotional words for nouns, adjectives, verbs, adverbs, interjections, and conjunctions. We confirmed that when we use our words evaluated on an 11-ranges scale, the results are more applicable than those of the Dict Ed. The results differ depending on the protocol. However, Bayesian classification is respected on subjective evaluation overall. Using Bayesian classification, the emotion of each sentence was evaluated as P, N, or Nt. The words contained in the person' s own tweets were used for training. Therefore, we consider that the emotion evaluation was performed effectively. As a result of the subjective evaluation, the Bayesian classification showed good results.

The following points justify the efficacy of the subjective evaluation:

1. The person's own words were used as a corpus.
2. The categories of P, N, and Nt were evaluated for each sentence of the tweets.
3. The location on the higher label and the location on the lower label matched the person's own intuition.

The results are supported by the chi-square test. Regarding the subjective evaluation, when FSPH and FSPL were evaluated on a 5-point scale, the memories of the relevant tweets and their locations were retroactively evaluated.

4.5 Summary

In this study, we measured the feelings a person had for certain locations and quantitatively evaluated emotion quantitative methods. We evaluated the feelings for a place with various personal experiences and memories. *Ibasho* is defined as a place with one's special feelings or for achieving a certain purpose, similar to the concept of *"Topophilia"* given by Tuan.

To evaluate an appropriate quantification method for the feelings about a place, we created an evaluation model and verified three quantification methods: Dict Ed, Dict Pd, and Bayesian classification. In our experiment, we collected tweets when entering and leaving a place and calculated emotion scores. Four protocols were executed based on the evaluation model, and a subjective evaluation was carried out on those results. The results indicated that the emotion scores obtained by Bayesian classification were most consistent with the subjectivity.

Our study contributes an improvement on the analysis of personal data including locations. However, there are limitations. First, the Twitter dataset was created from the subjective experience of one subject when he visited certain locations. The emotional scores of landmark names could not be calculated because the names of landmarks supporting the link between the location information and places where the subject visited are needed and their names could not be ascertained despite visiting them frequently. In addition, we analyzed the tweets relevant to a place by the meanings of separate words to determine the emotional scores, but not by the meanings of a full sentence.

For future work, we will use datasets from more subjects with necessary information to solve the lack of location information for landmarks. Moreover, we will try to make improvements by introducing dependency analysis and utilizing the techniques of NLP. We will also further explore the evaluation method to obtain the individual's subjective emotion score and improve the quantification method to quantify the feelings associated with a place more accurately.

Chapter 5

Conclusion

5.1 Summary

This study aims to explore effective methods of personal data analysis for well-being oriented living support. In this thesis, we have presented three studies, which are closely related each other, and evaluation experiment results as well. Major contributions of this thesis are summarized as follows.

In Study 1 (Chapter 2), a framework of personal data analysis toward well-being oriented living support was proposed. In addition, feasibility study with a statistical method was conducted, and application scenarios were demonstrated. In Study 2 (Chapter 3), personal whereabouts model was presented for better analysis of personal data including location information to deeply understand individual living and behavioral patterns, and an experiment was designed to extract factors and patterns related to transportation means and destinations by decision trees. Finally, in Study 3 (Chapter 4), measurement and quantification of an individual's feelings and emotions about a place by three sentiment analysis methods, i.e., emotion dictionary, personalized dictionary, and Bayesian classification, have been explored. Experiments with four different protocols were designed and conducted using tweet data including location information and a subject's emotion changes about the locations. The results of evaluation experiments have shown that the proposed approaches and models for analysis of personal data including location information are valid and effective to offer well-being oriented living support.

The object of this study is personal data of an individual. In view of this, it is relevant to the approach of Research with First-Person's View. However, Research with First-Person's View is limited to analysis and evaluation from the subjective perspective of a researcher himself or herself. In contrast, the proposed approach in this study uses machine learning and statistical methods for quantitative analysis and evaluation in an objective manner, based on multiple types of datasets collected from an individual over a long period of time, in addition to qualitative analysis and evaluation of Research with First-Person's View.

Personal data analysis methods and models proposed in this thesis can ensure the objectivity of the research results and are of universality. Therefore, they can be generalized and provide a new perspective to existing research methods. In this sense, this study contributes to academic development and has academic significance. By implementing the integrated framework of personal data analysis for well-being-oriented living support proposed in this study as a system and applying it to personalized healthcare, it can be expected to promote people's health and quality of life. Therefore, it is considered to have social significance. Based on the concept of human sciences, this study examines and validates the exploration and effectiveness of personal data analysis for well-being oriented living support, which will contribute to the development of human sciences research.

5.2 Future Work

Future work for the three studies has been addressed in Chapters 2, 3 and 4, respectively. As a whole, the future work for personal data analysis toward well-being oriented living support, comprehensively considering these three studied presented in this thesis, includes improvement of personal whereabouts model (Study 2) by taking into account of an individual's feelings about a place (Study 3), and improvement of the proposed framework for personal data analysis (Study 1) by integrating the personal whereabouts model (Study 2) and an individual's feelings about a place (Study 3). In addition, we will develop algorithms to implement each functional module, construct a prototype system, and conduct experimental evaluation.

Acknowledgments

I would like to express my sincere gratitude to all the professors, staff, and students of Waseda University who were involved in my thesis. I entered e-School as a working student, and was impressed by Professor Qun Jin’ s teaching, and selected his laboratory to do my graduation thesis and master’ s program. Later, I was honored to be elected as the graduate representative at the graduation ceremony. During my enrollment, sometimes the work to implement my idea did not make the progress as I expected, and my heart was almost broken. However, Prof. Jin was always kind and attentive to me and gave me detailed advice. I would like to express my deepest gratitude to Prof. Jin. I was impressed by the lectures given by Professor Shoji Nishimura in the master’ s course, and he was kind to me since then, who serving as a deputy referee. I would like to express my sincere gratitude for Prof. Nishimura’ s candid guidance on experiment design and statistical methods for data analysis when I was writing my thesis. Professor Shigeto Ozawa, who was also a deputy referee, guided me to study based on the same perspective existing in the field of cognitive science and gave me a multifaceted way of thinking on research. I would like to thank Dr. Xiaokang Zhou, now an Associate Professor of Shiga University, for his help and guidance when I was writing my first paper, from how to investigate the literature and draw diagrams to correction and revision. I would like to express my gratitude to Dr. Kiichi Tago, now an Assistant Professor of Chiba Institute of Technology, for sharing the pressure of doctoral thesis, encouraging each other, and having constructive discussions every weekend. In addition, I would like to thank Mr. Takeshi Fujitani, with whom I have shared joys and sorrows since entering e-School as the same working students and always encouraged each other. I also extend my appreciation to all members of Networked Information Systems Laboratory. Finally, I would like to extend my heartfelt thanks to my family for their support and cooperation in my fulltime work and weekend research life.

Bibliography

[1] Vassilis Christophides and Themis Palpanas. Report on the first international workshop on personal data analytics in the internet of things (pda@ iot 2014). *ACM SIGMOD Record*, 44(1):52–55, 2015.

[2] World Health Organization et al. *Ecosystems and human well-being: health synthesis: a report of the Millennium Ecosystem Assessment.* Geneva: WorldHealth Organization, 2005.

[3] Seiji Kasuya, Xiaokang Zhou, Shoji Nishimura, and Qun Jin. Personal data analytics for well-being oriented life support: Experiment and feasibility study. In *Frontiers in Artificial Intelligence and Applications*, volume 282 of *Frontiers in Artificial Intelligence and Applications*, pages 172–179. IOS Press, 2016. 7th International Conference on Applications of Digital Information and Web Technologies, ICADIWT 2016 ; Conference date: 29-03-2016 Through 31-03-2016.

[4] Jim Gemmell, Gordon Bell, Roger Lueder, Steven Drucker, and Curtis Wong. Mylifebits: fulfilling the memex vision. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 235–238, 2002.

[5] Gang Luo, Chunqiang Tang, and Selena B Thomas. Intelligent personal health record: experience and open issues. *Journal of medical systems*, 36(4):2111–2128, 2012.

[6] Deborah Estrin. Small data, where n= me. *Communications of the ACM*, 57(4):32–34, 2014.

[7] Serguei Dobrinevski. Personal analytics as a factor of change in enterprise communication and collaboration patterns. In *2013 27th International Conference on Advanced Information Networking and Applications Workshops*, pages 1135–1140. IEEE, 2013.

[8] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on*

Pervasive and Ubiquitous Computing, pages 731–742, 2015.

[9] Teruhiko Teraoka. Organization and exploration of heterogeneous personal data collected in daily life. *Human-Centric Computing and Information Sciences*, 2(1):1–15, 2012.

[10] Min Mun, Shuai Hao, Nilesh Mishra, Katie Shilton, Jeff Burke, Deborah Estrin, Mark Hansen, and Ramesh Govindan. Personal data vaults: a locus of control for personal data streams. In *Proceedings of the 6th International COnference*, pages 1–12, 2010.

[11] Dominik Schuldhaus, Heike Leutheuser, and Bjoern M Eskofier. Classification of daily life activities by decision level fusion of inertial sensor data. In *Proceedings of the 8th International Conference on Body Area Networks*, pages 77–82, 2013.

[12] James Y Xu, Yuwen Sun, Zhao Wang, William J Kaiser, and Greg J Pottie. Context guided and personalized activity classification system. In *Proceedings of the 2nd Conference on Wireless Health*, pages 1–10, 2011.

[13] Weimin Li, Xunfeng Li, Mengke Yao, Jiulei Jiang, and Qun Jin. Personalized fitting recommendation based on support vector regression. *Human-centric computing and information sciences*, 5(1):1–11, 2015.

[14] Bryan Minor, Janardhan Rao Doppa, and Diane J Cook. Data-driven activity prediction: Algorithms, evaluation methodology, and applications. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 805–814, 2015.

[15] Saida Aissi, Mohamed Salah Gouider, Tarek Sboui, and Lamjed Ben Said. A spatial data warehouse recommendation approach: conceptual framework and experimental evaluation. *Human-centric Computing and Information Sciences*, 5(1):1–18, 2015.

[16] Soudip Roy Chowdhury, Florian Daniel, and Fabio Casati. Recommendation and weaving of reusable mashup model patterns for assisted development. *ACM Transactions on Internet Technology (TOIT)*, 14(2-3):1–23, 2014.

[17] Thomas G Morrell and Larry Kerschberg. Personal health explorer: A semantic health recommendation system. In *2012 IEEE 28th International Conference on Data Engineering Workshops*, pages 55–59. IEEE, 2012.

[18] Rachid Benlamri and Xiaoyun Zhang. Context-aware recommender for mobile learners.

Human-centric Computing and Information Sciences, 4(1):12, 2014.

[19] Sunny Consolvo, David W McDonald, and James A Landay. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 405–414, 2009.

[20] Laleh Jalali and Ramesh Jain. Building health persona from personal data streams. In *Proceedings of the 1st ACM international workshop on Personal data meets distributed multimedia*, pages 19–26, 2013.

[21] Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. Health mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(5):1–27, 2013.

[22] Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex Pentland. Daily stress recognition from mobile phone data, weather conditions and individual traits. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 477–486, 2014.

[23] Rischan Mafrur, I Gde Dharma Nugraha, and Deokjai Choi. Modeling and discovering human behavior from smartphone sensing life-log data for identification purpose. *Human-centric Computing and Information Sciences*, 5(1):31, 2015.

[24] Luis A Castro, Jesus Favela, Eduardo Quintana, and Moises Perez. Behavioral data gathering for assessing functional status and health in older adults using mobile phones. *Personal and Ubiquitous Computing*, 19(2):379–391, 2015.

[25] James McNaull, Juan Carlos Augusto, Maurice Mulvenna, and Paul McCullagh. Flexible context aware interface for ambient assisted living. *Human-Centric Computing and Information Sciences*, 4(1):1, 2014.

[26] Qun Jin, Bo Wu, Shoji Nishimura, and Atsushi Ogihara. Ubi-liven: a human-centric safe and secure framework of ubiquitous living environments for the elderly. In *2016 International Conference on Advanced Cloud and Big Data (CBD)*, pages 304–309. IEEE, 2016.

[27] Carol D Ryff and Corey Lee M Keyes. The structure of psychological well-being revisited. *Journal of personality and social psychology*, 69(4):719, 1995.

[28] K Schwab, A Marcus, JR Oyola, and W Hoffman. Personal data: The emergence of a new asset class. an initiative of the world economic forum (in collaboration with bain & company, inc.)(2011).

[29] Jawbone Fitness Tracker. Online. `https://jawbone.com/up`. Accessed: 24 Mar. 2017.

[30] Japan Meteorological Agency. Online. `https://www.data.jma.go.jp/obd/stats/etrn/index.php`.

[31] Twitter. Online. `https://twitter.com/home`.

[32] Erik Næsset. Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(2):49 – 56, 1997.

[33] B. L. Welch. The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29(3/4):350–362, 1938.

[34] Koji Yatani. Effect sizes and power analysis in hci. *In: Robertson J., Kaptein M. (eds) Modern Statistical Methods for HCI. Human–Computer Interaction Series. Springer*, pages 87–110, 2016.

[35] Nor Aishah Ahad and Sharipah Soaad Syed Yahaya. Sensitivity analysis of welch' st-test. In *AIP Conference proceedings*, volume 1605, pages 888–893. American Institute of Physics, 2014.

[36] Shlomo S. Sawilowsky. Fermat, schubert, einstein, and behrens-fisher: The probable difference between two means when $\sigma_1^2 \neq \sigma_2^2$. *Journal of Modern Applied Statistical Methods*, 1(2):461–472, 2002.

[37] N Matsubara, K Nawata, and N Nakai. Introduction to statistics. *Statistics Section, Department of Social Sciences, College of Arts and Sciences, The University of Tokyo, University of Tokyo Press*, pages 258–277, 1991.

[38] Seiji Kasuya, Xiaokang Zhou, Shoji Nishimura, and Qun Jin. A framework of personal data analytics for well-being oriented life support. In *Advanced Multimedia and Ubiquitous Engineering*, pages 443–449. Springer, 2016.

[39] S. Kasuya, X. Zhou, K. Tago, S. Nishimura, and Q. Jin. Cyber-enabled well-being oriented daily living support based on personal data analytics. *IEEE Transactions on Emerging Topics*

in Computing, pages 1–1, 2017.

[40] Riccardo Guidotti, Roberto Trasarti, and Mirco Nanni. Tosca: two-steps clustering algorithm for personal locations detection. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–10, 2015.

[41] Masanobu Abe, Daisuke Fujioka, and Hisashi Handa. A life log collecting system supported by smartphone to model higher-level human behaviors. In *2012 Sixth International Conference on Complex, Intelligent, and Software Intensive Systems*, pages 665–670. IEEE, 2012.

[42] Andrea Viel, Andrea Brunello, Angelo Montanari, and Federico Pittino. An original approach to positioning with cellular fingerprints based on decision tree ensembles. *Journal of Location Based Services*, 13(1):25–52, 2019.

[43] Attila Reiss and Didier Stricker. Personalized mobile physical activity recognition. In *Proceedings of the 2013 international symposium on wearable computers*, pages 25–28, 2013.

[44] Ao Guo and Jianhua Ma. A context-aware scheduling mechanism for smartphone-based personal data collection from multiple wearable devices. In *2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pages 528–533. IEEE, 2016.

[45] Shakiba Rahimiaghdam, Pinar Karagoz, and Alev Mutlu. Personalized time-aware outdoor activity recommendation system. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, pages 1121–1126, 2016.

[46] Ravi Karkar, Jasmine Zia, Roger Vilardaga, Sonali R Mishra, James Fogarty, Sean A Munson, and Julie A Kientz. A framework for self-experimentation in personalized health. *Journal of the American Medical Informatics Association*, 23(3):440–448, 2016.

[47] Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In *Proceedings of the 2014 conference on Designing interactive systems*, pages 667–676, 2014.

[48] Juha Pärkkä, Luc Cluitmans, and Miikka Ermes. Personalization algorithm for real-time activity recognition using pda, wireless motion bands, and binary decision tree. *IEEE Transactions*

on Information Technology in Biomedicine, 14(5):1211–1215, 2010.

[49] Francis Bond, Timothy Baldwin, Richard Fothergill, and Kiyotaka Uchimoto. Japanese semcor: A sense-tagged corpus of japanese. In *Proceedings of the 6th global WordNet conference (GWC 2012)*, pages 56–63. Citeseer, 2012.

[50] Moves - Actibity Daiary for iPhone and Android. Online. `https://www.moves-app.com`. Accessed: 25 Jun.2018.

[51] J Ross Quinlan. Improved use of continuous attributes in c4. 5. *Journal of artificial intelligence research*, 4:77–90, 1996.

[52] Piotr. Sekara Vedran. Lehmann Sune. Alessandretti, Laura. Sapiezynski and Andrea. Baronchelli. Evidence for a conserved quantity in human mobility. *Nature Human Behaviour*, 2:485–491, July 2018.

[53] Samiul Hasan and Satish V Ukkusuri. Urban activity pattern classification using topic models from online geo-location data. *Transportation Research Part C: Emerging Technologies*, 44:363–381, 2014.

[54] Changqing Zhou, Dan Frankowski, Pamela Ludford, Shashi Shekhar, and Loren Terveen. Discovering personally meaningful places: An interactive clustering approach. *ACM Trans. Inf. Syst.*, 25(3), July 2007.

[55] Ran Li and Daoqin Tong. Incorporating activity space and trip chaining into facility siting for accessibility maximization. *Socio-Economic Planning Sciences*, 60:1 – 14, 2017.

[56] Frank Bentley, Henriette Cramer, and Jörg Müller. Beyond the bar: the places where location-based services are used in the city. *Personal and Ubiquitous Computing*, 19(1):217–223, 2015.

[57] Krittika D'Silva, Anastasios Noulas, Mirco Musolesi, Cecilia Mascolo, and Max Sklar. Predicting the temporal activity patterns of new venues. *EPJ data science*, 7(1):13, 2018.

[58] Lauren Herckis, Jessica Cao, Jacqui Fashimpaur, Anna Henson, Rachel Rodgers, Thomas W. Corbett, and Jessica Hammer. Exploring hybrid virtual-physical homes. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, page 669–680, New York, NY, USA, 2020. Association for Computing Machinery.

[59] Steve Harrison and Paul Dourish. Re-place-ing space: the roles of place and space in collabora-

tive systems. *In: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pages 67–76, 1996.

[60] Paul Dourish. Re-space-ing place: "place" and "space" ten years on. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, CSCW '06, page 299–308, New York, NY, USA, 2006. Association for Computing Machinery.

[61] Chirath Kumarasiri and Cassim Farook. User centric mobile based decision-making system using natural language processing (nlp) and aspect based opinion mining (abom) techniques for restaurant selection. In *Intelligent Computing*, pages 43–56, Cham, 2019. Springer International Publishing.

[62] Victoria Hollis, Artie Konrad, Aaron Springer, Matthew Antoun, Christopher Antoun, Rob Martin, and Steve Whittaker. What does all this data mean for my future mood? actionable analytics and targeted reflection for emotional well-being. *Human–Computer Interaction*, 32(5-6):208–267, 2017.

[63] Norsidah Ujang and Khalilah Zakariya. The notion of place, place meaning and identity in urban regeneration. *Procedia-social and behavioral sciences*, 170:709–717, 2015.

[64] Norsidah Ujang and Khalilah Zakariya. Place attachment and the value of place in the life of the users. *Procedia - Social and Behavioral Sciences*, 168:373 – 380, 2015. Asia Pacific International Conference on Environment-Behaviour Studies (AicE-Bs 2014Berlin), Sirius Business Park Berlin-yard field, Berlin, Germany, 24-26.

[65] E. C. Dragut, H. Wang, P. Sistla, C. Yu, and W. Meng. Polarity consistency checking for domain independent sentiment dictionaries. *IEEE Transactions on Knowledge and Data Engineering*, 27(3):838–851, March 2015.

[66] Sergio Altares López and Juan J. Cuadrado-Gallego. Supervised learning methods application to sentiment analysis. In *Proceedings of the 23rd International Database Applications and Engineering Symposium*, IDEAS '19. Association for Computing Machinery, 2019.

[67] K. Jia and Z. Li. Chinese micro-blog sentiment classification based on emotion dictionary and semantic rules. In *2020 International Conference on Computer Information and Big Data Applications (CIBDA)*, pages 309–312, April 2020.

[68] Kiichi Tago, Kosuke Takagi, Seiji Kasuya, and Qun Jin. Analyzing influence of emotional tweets on user relationships using naive bayes and dependency parsing. *World Wide Web*, 22(3):1263–1278, May 2019.

[69] K. Tago and Q. Jin. Influence analysis of emotional behaviors and user relationships based on twitter data. *Tsinghua Science and Technology*, 23(1):104–113, Feb 2018.

[70] Lin Mu, Peiquan Jin, Lizhou Zheng, En-Hong Chen, and Lihua Yue. Lifecycle-based event detection from microblogs. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, page 283–290, Republic and Canton of Geneva, CHE, 2018. International World Wide Web Conferences Steering Committee.

[71] F. Chen, R. Ji, J. Su, D. Cao, and Y. Gao. Predicting microblog sentiments via weakly supervised multimodal deep learning. *IEEE Transactions on Multimedia*, 20(4):997–1007, April 2018.

[72] Yi-Fu Tuan. *Topophilia: A study of environmental perceptions, attitudes, and values.* Columbia University Press, 1990.

[73] Hakon Heimer. Topophilia and quality of life: defining the ultimate restorative environment, 2005.

[74] Hiroya Takamura, Takashi Inui, and Manabu Okumura. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 133–140, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.

[75] Kudo Taku, Yamamoto Kaoru, and Matsumoto Yuji. Applying conditional random fields to japanese morphological analysis. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004)*, pages 230–237, 2004.

[76] Henriette Cramer Frank Bentley and Jörg Müller. Beyond the bar: the places where location-based services are usedin the city. *Personal and Ubiquitous Computing*, 19:217–223, 2015.