IEEE TRANSACTIONS ON
EMERGING TOPICS
IN COMPUTING

Received 14 December 2016; revised 4 September 2017; accepted 5 September 2017.
Date of publication 17 October 2017; date of current version 9 June 2020.

Digital Object Identifier 10.1109/TETC.2017.2763966

Cyber-Enabled Well-Being Oriented Daily Living Support Based on Personal Data Analysis

SEIJI KASUYA, XIAOKANG ZHOU, KIICHI TAGO, SHOJI NISHIMURA, AND QUN JIN

S. Kasuya is with the Graduate School of Human Sciences, Waseda University, 2-579-15 Mikajima, Tokorozawa, Saitama 359-1192, Japan
X. Zhou is with the Faculty of Data Science, Shiga University, 1-1-1 Banba, Hikone, Shiga 522-8522, Japan
K. Tago is with the Graduate School of Human Sciences, Waseda University, 2-579-15 Mikajima, Tokorozawa, Saitama 359-1192, Japan
S. Nishimura is with the Faculty of Human Sciences, Waseda University, 2-579-15 Mikajima, Tokorozawa, Saitama 359-1192, Japan
Q. Jin is with the Faculty of Human Sciences, Waseda University, 2-579-15 Mikajima, Tokorozawa, Saitama 359-1192, Japan
CORRESPONDING AUTHOR: Q. JIN (jin@waseda.jp)

ABSTRACT We are living in a cyber-physical-social environment with a variety of lifestyles and values. Living support has become important in such a diverse society. Owing to the ability to collect a large amount of personal data or life logs in the cyber-physical-social environment, it is now possible for us to provide living support based on personal data analysis. Moreover, analyzing such data can facilitate a deep understanding of an individual. In this study, we focus on the provision of cyber-enabled well-being oriented daily living support for an individual based on personal data analysis. Three categories of personal data are identified from an individual's daily life data. In this paper, we discuss the basic concept, model, and framework for well-being oriented personal data analysis in order to offer suggestions and advices to improve the living quality of an individual. Finally, we report a feasibility study with an application scenario by using personal and environmental data.

INDEX TERMS Personal data analytics, well-being oriented living support, data-driven behavior analysis, cyber-enabled application

I. INTRODUCTION

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 the present 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 for an individual to others. In this paper, 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 paper, we demonstrate the feasibility of this framework through a study that used and analyzed a set of personal and environmental data.

The rest of this paper is organized as follows. We provide brief overview on related work in Section II. In Section III, after introducing the definition of well-being, we describe the basic concept and model of well-being oriented living support. In Section IV, we discuss the extraction and collection of heterogeneous data in the daily life cycle, and demonstrate

the method of conducting personal data analysis for well-being oriented living support. In Section V, a feasibility study with an actual data set is presented and discussed with an application scenario. Finally, we conclude this study and offer promising suggestions for future work in Section VI.

II. RELATED WORK

A. PERSONAL DATA ANALYSIS FOR DAILY LIFE SUPPORT

Gemmel *et al.* [5] 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. Luo *et al.* [6] 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 [7] discussed personalized data analysis, which can be used to infer health status and well-being from digital behaviors. Dobrinevski [8] 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.* [9] 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 [10] 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.* [11] 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.

B. DAILY ACTIVITY CLASSIFICATION AND PERSONALIZED RECOMMENDATION

Schuldhaus, *et al.* [12] 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.* [13] 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.* [14] 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.* [15] 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.* [16] proposed enhanced spatial data warehouse exploitation by recommending personalized MDX queries to the users while taking into account their preferences and needs. Chowdhury *et al.* [17] 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. Morrell and Kerschberg [18] 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 [19] 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.

C. BEHAVIOR ANALYSIS FOR DAILY LIFE MANAGEMENT

Consolvo *et al.* [20] 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 [21] 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.* [22] 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.* [23] 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.* [24] 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.* [25] 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.* [26] proposed a system which can provide feedback to highlight the important criteria for sleep quality during the night. Jin *et al.* [27] 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.

D. SUMMARY

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 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.

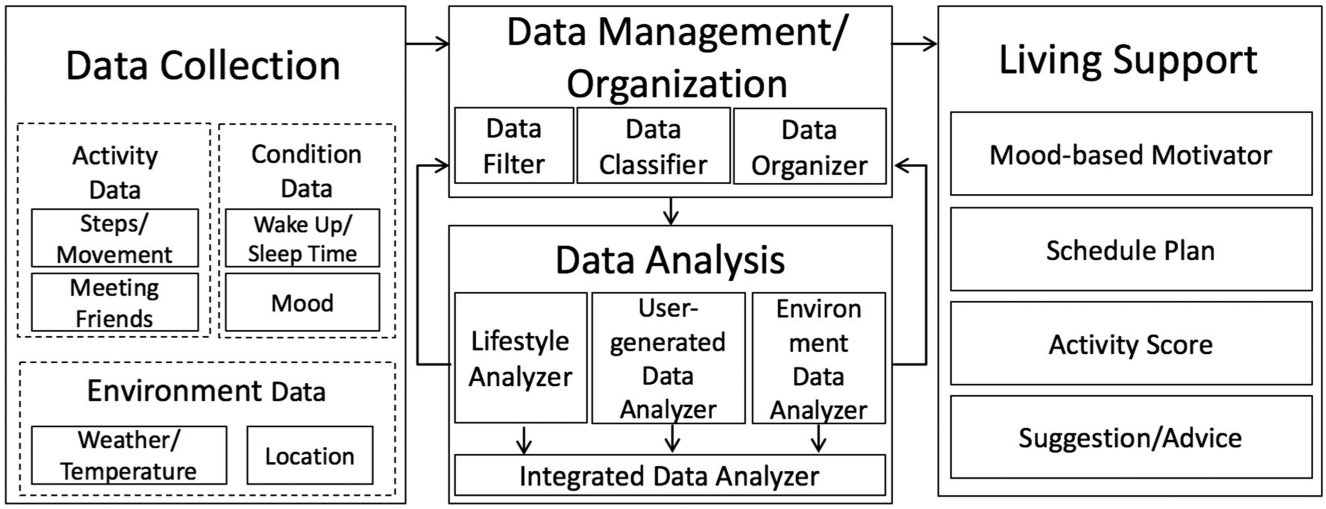


FIGURE 1. **Framework of well-being oriented living support.**

III. CONCEPT OF WELL-BEING AND LIVING SUPPORT

A. 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 [28]. 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.

B. FRAMEWORK OF WELL-BEING ORIENTED LIVING SUPPORT

Considering the above discussion, in the present paper, 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 Figure 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.

First, 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.

C. 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 (WD) or Weekend/Holiday (WH). WD or 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 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, WD, WH, CW, and LO. Subsequently, classify the data into three major categories: volunteered data, observed data, and inferred data [29];

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.

IV. DATA-DRIVEN AND CYBER-ENABLED LIVING SUPPORT

A. 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 [29]: 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 one's 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.

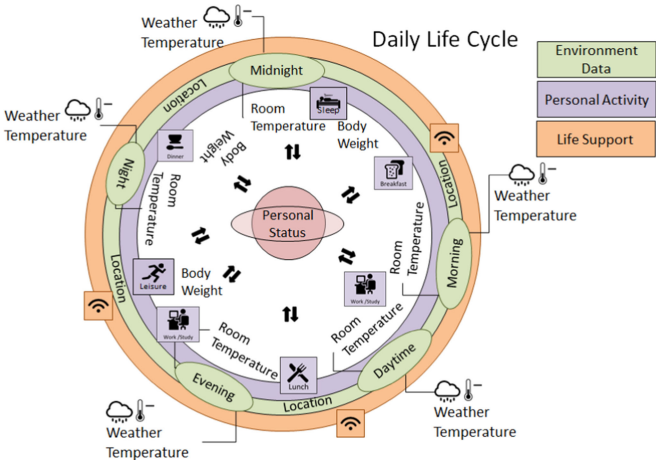


FIGURE 2. Conceptual image of the data collected in the daily life cycle.

Figure 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.

B. FRAMEWORK OF PERSONAL DATA ANALYSIS

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 3.

As shown in Figure 3, after data filtering and classifying, the three basic data, i.e., activity data, condition data, and

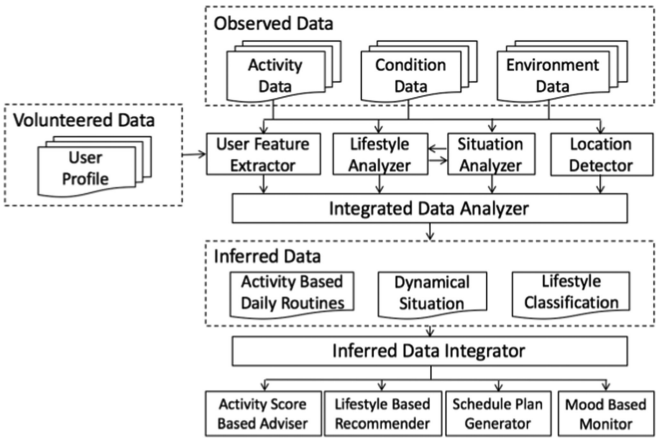


FIGURE 3. Personal data analysis for cyber-enabled living support.

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.

V. THE FEASIBILITY STUDY BASED ON THE STATISTICAL ANALYSIS

A. 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 [30]) includes the number of walk steps, sleep time, and deep sleep time. The environment-related data in one's 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 hours, and daytime and night weather were collected from the Japan Meteorological Agency (JMA) [31]. Besides, Twitter [32] data was utilized as a type of condition-related data.

Table 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).

TABLE 1. Data types, units, and sources of the collected personal data.

Data Type	*Unit*	*Data Source*
Sleep 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

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.

B. 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 [33] was calculated using Eq. (1).

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

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) [33] was obtained using Eq. (2).

$$SD = \sqrt{\frac{1}{n-1}\sum\nolimits_{i=1}^{n} (x_i - \bar{x})^2} \tag{2}$$

The Median $Q_{\frac{1}{2}}(x)$ was calculated using Eq. (3).

$$Q_{\frac{1}{2}}(x) = \begin{cases} x'_{\frac{n+1}{2}}, & \textit{if } n \textit{ is odd} \\ \frac{1}{2}\left(x'_{\frac{n}{2}} + x'_{\frac{n}{2}+1}\right), & \textit{if } n \textit{ is even} \end{cases} \tag{3}$$

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

Tables 2 and 3 present these descriptive statistics for the data on weekdays and weekends/holidays, respectively.

C. 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.

The experiment data were examined using the following four steps:

TABLE 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 (°C)	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

(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 [34], 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 [35]), 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 [36]. 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 [37] can be calculated by Eq. (4), and the Degree of Freedom (DF) [34] can be obtained by Eq. (5).

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{4}$$

$$DF = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \tag{5}$$

where $s_1 = \sqrt{\sum_1 / n_1(n_1 - 1)}$ and $s_2 = \sqrt{\sum_2 / n_2(n_2 - 1)}$, 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 Eq. (6), or more generally, by the so-called t-distribution table [38].

TABLE 3. Basic properties of personal data on weekends/holidays.

Weekend/Holiday ($N = 64$)	*Mean*	*SD*	*Median*	*Minimum*	*Maximum*
Precipitation (mm)	4.14	15.87	0	0	109.5
Average temperature (°C)	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	125	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

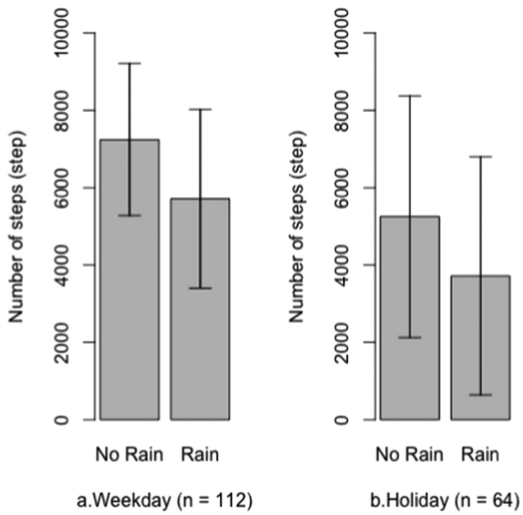


FIGURE 4. Average walk steps according to the weather on weekdays and weekends/holidays.

$$t_a(DF) : \int_{t_a}^{\infty} \frac{1}{\sqrt{DF}\ B\left(\frac{1}{2}, \frac{DF}{2}\right)\left(1 + \frac{t^2}{DF}\right)^{\frac{DF+1}{2}}} dt = \alpha \qquad (6)$$

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 4 to 7 and Table 4 as well. From Table 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

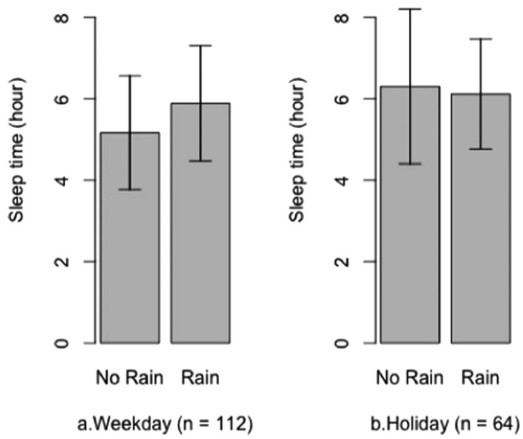


FIGURE 5. Average sleep time according to the weather on weekdays and weekends/holidays.

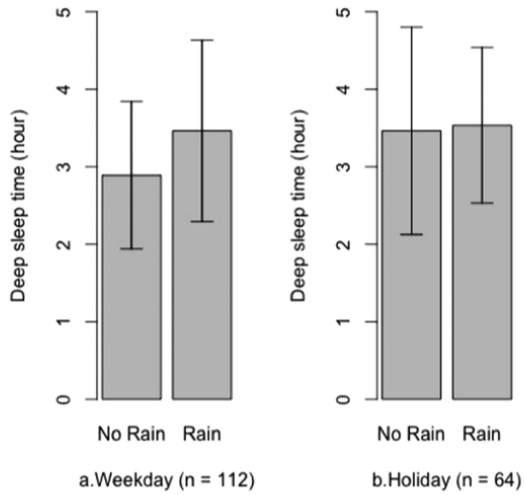


FIGURE 6. Average deep sleep time according to the weather on weekdays and weekends/holidays.

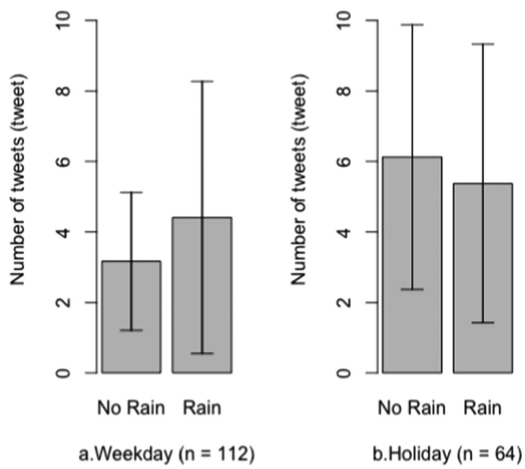


FIGURE 7. Average number of tweets according to the weather on weekdays and weekends/holiday.

TABLE 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			

Note: N=176. SD = standard deviation. DF = degree of freedom. *<0.05, **<0.01, n.s.: not significant

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 of the data were missed for a day, the 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 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.

D. 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 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 for an individual to others. If fact, it provides an individual with well-being oriented living support based on evidence obtained specially for that specific individual.

E. 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 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

TABLE 5. Comparison of this work with related works on living support.

Research	*Individual Behavior*	*Life Event*	*Personal Trait*	*Local Environment*
Gemmell *et al.* [5], Cosolvo *et al.* [20]	✓			
Jalali and Jain [21]		✓		
Mun *et al.* [11], Xu *et al.* [13]	✓		✓	✓
Luo *et al.* [6]		✓	✓	✓
Teraoka [10], Schuldhaus *et al.* [12]	✓	✓	✓	
This work	✓	✓	✓	✓

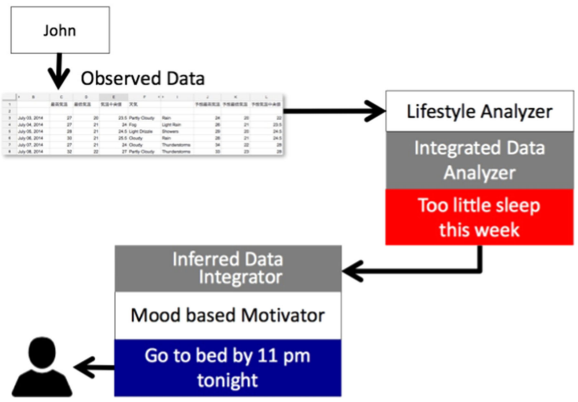


FIGURE 8. Application scenario for the mood-based motivator.

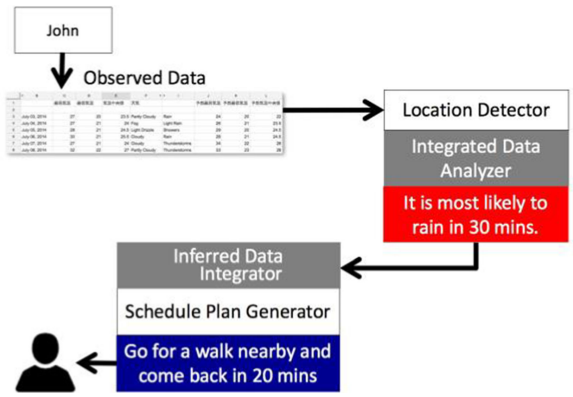


FIGURE 9. Application scenario for the schedule plan generator.

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 maintain good health. As shown in Figure 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.

VI. CONCLUSION

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 paper, 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.

ACKNOWLEDGMENTS

The work was supported in part by 2016 and 2017 Waseda University Grants for Special Research Projects No. 2016B-233 and No. 2017B-302, and 2016-2018 Masaru Ibuka Foundation Research Project on Oriental Medicine.

REFERENCES

[1] V. Christophides and T. Palpanas, "Report on the first international workshop on personal data analytics in the internet of things (PDA@IOT 2014)," *Newspaper, ACM SIGMOD Rec.*, vol. 44, no. 1, pp. 52–55, 2015.

[2] World Health Organization: Ecosystems and Human Well-Being: Health Synthesis, (2005). [Online]. Available: http://apps.who.int/iris/bitstream/10665/43354/1/9241563095.pdf, Accessed on: Mar. 24, 2017.

[3] S. Kasuya, *et al.*, "A framework of personal data analytics for well-being oriented life support," *Adv. Multimedia Ubiquitous Eng.*, vol. 354, pp. 443–449, 2016.

[4] S. Kasuya, *et al.*, "Personal data analytics for well-being oriented life support: Experiment and feasibility study," *Frontiers Artif. Intell. Appl.*, vol. 282, pp. 172–179, 2016.

[5] J. Gemmell, *et al.*, "MyLifeBits: Fulfilling the memex vision," in *Proc. 10th ACM Int. Conf. Multimedia*, 2002, pp. 235–238.

[6] G. Luo, *et al.*, "Intelligent personal health record: Experience and open issues," in *Proc . 1st ACM Int. Health Informat. Symp.*, 2010, pp. 326–335.

[7] D. Estrin, "Small data, where n = me," *Commun. ACM*, vol. 57, no. 4, pp. 32–34, 2014.

[8] S. Dobrinevski, "Personal analytics as a factor of change in enterprise communication and collaboration patterns," in *Proc . 27th Int. Conf . Adv. Inf. Netw . Appl. Workshops*, 2013, pp. 1135–1140.

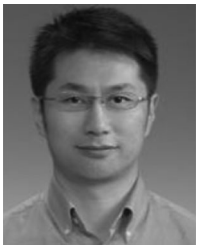
[9] D. A. Epstein, *et al.*, "A lived informatics model of personal informatics," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous*, 2015, pp. 731–742.

[10] T. Teraoka, "Organization and exploration of heterogeneous personal data collected in daily life," *Human-Centric Comput. Inf. Sci.*, vol. 2, 2012, Art. no. 1.

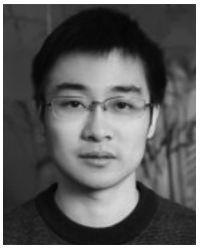
[11] M. Mun, *et al.*, "Personal data vaults: A locus of control for personal data streams," in *Proc. 6th Int. Conf.*, 2010,, Art. no. 17.

[12] D. Schuldhaus, *et al.*, "Classification of daily life activities by decision level fusion of inertial sensor data," in *Proc. 8th Int. Conf. Body Area Netw.*, 2013, pp. 77–82.

[13] J. Y. Xu, *et al.*, "Context guided and personalized activity classification system", in *Proc. 2nd Conf. Wireless Health*, 2011, Art. no. 12.
[14] W. Li, *et al.*, "Personalized fitting recommendation based on support vector regression," *Human-Centric Comput. Inf. Sci.*, vol. 5, 2015, Art. no. 21.
[15] B. Minor, J. R. Doppa, and D. J. Cook, "Data-driven activity prediction: Algorithms, evaluation methodology, and applications," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 805–814.
[16] S. Aissi, *et al.*, "A spatial data warehouse recommendation approach: Conceptual framework and experimental evaluation," *Human-Centric Comput. Inf. Sci.*, vol. 5, 2015, Art. no. 30.
[17] S. R. Chowdhury, *et al.*, "Recommendation and weaving of reusable mashup model patterns for assisted development," *ACM Trans. Internet Technol.*, vol. 14, 2014, Art. no. 2.
[18] T. G. Morrell and L. Kerschberg, "Personal health explorer: A semantic health recommendation system," in *Proc. IEEE 28th Int. Conf. Data Eng. Workshop*, 2012, pp. 55–59.
[19] R. Benlamri and X. Zhang, "Context-aware recommender for mobile learners," *Human-Centric Comput. Inf. Sci.*, vol. 4, 2014, Art. no. 12.
[20] S. Consolvo, *et al.*, "Theory-driven design strategies for technologies that support behavior change in everyday life," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2009, pp. 405–414.
[21] L. Jalali and R. Jain, "Building health persona from personal data streams," in *Proc. 1st ACM Int. Workshop Personal Data Meets Distrib. Multimedia*, 2013, pp. 19–26.
[22] F. Bentley, *et al.*, "Health mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change," *ACM Trans. Comput.-Human Interaction*, vol. 20, 2013, Art. no. 5.
[23] A. Bogomolov, *et al.*, "Daily stress recognition from mobile phone data, weather conditions and individual traits," in *Proc. ACM Int. Conf. Multimedia*, 2014, pp. 477–486.
[24] R. Mafrur, *et al.*, "Modeling and discovering human behavior from smartphone sensing life-log data for identification purpose," *Human-Centric Comput. Inf. Sci.*, vol. 5, 2015, Art. no. 31.
[25] L. A. Castro, *et al.*, "Behavioral data gathering for assessing functional status and health in older adults using mobile phones," *Int. J. Personal Ubiquitous Comput.*, vol. 19, no. 2, pp. 379–391, 2015.
[26] J. McNaull, *et al.*, "Flexible context aware interface for ambient assisted living," *Human-Centric Comput. Inf. Sci.*, vol. 4, no. 1, 2014, pp. 1–41.
[27] Q. Jin, B. Wu, S. Nishimura, and A. Ogihara, "Ubi-liven: A human-centric safe and secure framework of ubiquitous living environments for the elderly," in *Proc. 4th Int. Conf. Adv. Cloud Big Data*, 2016, pp. 304–309.
[28] C. D. Ryff and C. L. M. Keyes, "The structure of psychological well-being revisited," *J. Personality Social Psychol.*, vol. 69, no. 4, pp. 719–727, 1995.
[29] World Economic Forum, Personal Data: The Emergence of a New Asset Class, (2011). [Online]. Available: http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf, Accessed on: Mar. 24, 2017.
[30] Jawbone Fitness Tracker. [Online]. Available: https://jawbone.com/up Accessed on: Mar. 24, 2017.
[31] Japan Meteorological Agency. [Online]. Available: http://www.jma.go.jp/jma/index.html, Accessed on: Mar. 24, 2017.
[32] Twitter. [Online]. Available: https://twitter.com/, Accessed on: Mar. 24, 2017.
[33] E. Nñsset, "Determination of mean tree height of forest stands using airborne laser scanner data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 52, no. 2, pp. 49–56, 1997.
[34] B. L. Welch, "The significance of the difference between two means when the population variances are unequal," *Biometrika*, vol. 29, no. 3, pp. 350–361, 1938.
[35] K. Yatani, "Effect sizes and power analysis in HCI," *Modern Statist. Methods HCI*, pp. 87–110, 2016.
[36] N. A. Aishad and S. S. S. Yahaya, "Sensitivity analysis of Welch's t-test," *Proc. AIP*, pp. 888–893, 2014.
[37] S. S. Sawilowsky, "Fermat, Schubert, Einstein, and Behrens-Fisher: The probable difference between two means when," *J. Modern Appl. Statist. Methods*, vol. 1, no. 2, pp. 461–472, 2002.
[38] M. Nozomu, *et al.*, *Introduction to Statistics*. Tokyo, Japan: University of Tokyo Press, 1991, Art. no. 281. (in Japanese).



SEIJI KASUYA (M'17) received the MS degree from the Graduate School of Human Sciences, Waseda University, Japan, in 2012. He is currently working toward the PhD degree in the Graduate School of Human Sciences, Waseda University, Japan. His research interests include well-being oriented living support, personal data analytics, healthcare, and big data. He is a member of the IEEE, the IEEE CS, USA, and IPSJ and JSAI, Japan.



XIAOKANG ZHOU (M'12) is currently a lecturer of the faculty of Data Science, Shiga University, Japan. He also works as a visiting researcher in the RIKEN Center for Advanced Intelligence Project (AIP), RIKEN, Japan. He has been engaged in interdisciplinary research work in the fields of computer science and engineering, information systems, and human informatics. He has published more than 40 refereed papers in high-quality academic journals and international conference proceedings in the related research areas. His research interests include ubiquitous and social computing, data mining and analytics, behavior and cognitive informatics, user modeling, human-computer interaction, and information seeking and recommendation. He is a member of the IEEE, the IEEE CS, and the ACM, USA, and IPSJ and JSAI, Japan.

KIICHI TAGO received the bachelor degree of arts in 2013 from Waseda University, Japan. He is currently working toward the MS degree in the Graduate School of Human Sciences, Waseda University. His research interests include big data, social networking service, natural language processing, cognitive psychology, and behavioral analysis.

SHOJI NISHIMURA received the bachelor's degree in mathematics from Waseda University, the MSc degree in applied physics from Waseda University, and the PhD degree in human sciences from Osaka University. He is a professor with Waseda University, Japan. His current research interests include educational technology, especially education and the Internet. In 1991, he joined the Advance Research Center, INES Corporation as a senior researcher. Then, he joined the School of Human Sciences, Waseda University as an assistant professor in 1997. Currently, he is a professor in the Faculty of Human Sciences, Waseda University. He is a member of Japan Society for Educational Technology, Japanese Society for Information and Systems in Education, and Information Processing Society of Japan.

QUN JIN (M'95, SM'17) is a professor in the Department of Human Informatics and Cognitive Sciences, Faculty of Human Sciences, Waseda University, Japan. He has been engaged extensively in research works in the fields of computer science, information systems, and social and human informatics. He seeks to exploit the rich interdependence between theory and practice in his work with interdisciplinary and integrated approaches. He has published more than 200 refereed papers in world-renowned academic journals and international conference proceedings in related research areas. His current research interests cover human-centric ubiquitous computing, behavior and cognitive informatics, big data, personal analytics and individual modeling, cyber security, cyber-enabled applications in healthcare, and computing for well-being. He is a senior member of the IEEE and the IEEE CS, USA, and IPSJ, Japan, and a member of the ACM, USA, IEICE and JSAI, Japan, and the CCF, China.