

## Research Article

# Applying Data Mining Techniques to Identify Suitable Activities

Yu-Fang Yeh<sup>1</sup> and Ching-Pao Chang<sup>2</sup>

<sup>1</sup>*Sports Information and Communication, Aletheia University, New Taipei City 25103, Taiwan*

<sup>2</sup>*Information Engineering, Kun Shan University, Tainan 71003, Taiwan*

Correspondence should be addressed to Yu-Fang Yeh; [yufang.yeh@gmail.com](mailto:yufang.yeh@gmail.com)

Received 13 August 2015; Revised 6 October 2015; Accepted 8 October 2015

Academic Editor: Meng Du

Copyright © 2015 Y.-F. Yeh and C.-P. Chang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Identifying suitable physical activities is crucial for personal health management. However, a big challenge in identifying suitable physical activities is the influencing factors are extremely complex. Therefore, this study aims to propose an approach to facilitate the construction of suitable physical activity models. In the approach, association rule mining and clustering technique are applied to analyze personal activity-physiological information. To demonstrate how the proposed approach can be used for constructing the activity models, an experiment using mobile devices to collect personal activity-physiological information was designed. The revealed models can be used to not only understand personal health conditions but also provide useful information about proper and improper physical activities.

## 1. Introduction

Providing high quality healthcare with limited medical resources is a crucial concern for medical institutes [1–3]. Many medical institutes integrate user records and medical information to provide health management decision support [4, 5]. Many researches use wireless and sensor network technology for remote health monitoring [6, 7]. The collected physiological data are submitted to a medical center for medical diagnosis [8, 9]. Bauer et al. used a multilayer network to collect and analyze physiological data to provide useful information [10]. Although the collected data can be evaluated by medical personnel, it still requires significant efforts to provide useful information individually [11]. Tremblay et al. proposed the architecture of a health agent that analyzes the collected physiological data to facilitate personal health management [12]. Sim et al. proposed an evidence-based approach in which expert judgment is applied to data collected from user action and medical evidence to facilitate decision-making pertaining to personal activities [13]. The challenge in the evidence-based approach is that the data used to construct personal models are collected by users, and the characteristics of the same data may be different when the sources (users) differ. To overcome this problem, Li et al.

proposed a personal activity analysis process, which includes an interviewing and investigation stage to collect and verify data, an information integration stage to analyze the data, and a reflection stage to take decisions on personal health management [14]. However, the analysis process requires considerable efforts to support the data collection, validation, and analysis.

This study proposes an approach to the construction of activity models; in the approach, a data mining technique is applied to personal physiological data for the construction of personal activity models (PAMs). The proposed approach is based on modeling techniques to measure suitability of activities and construct activity models [15, 16]. Many modeling techniques can be used to measure the attributes of activities. For example, the regression analysis is a simple technique that can be used to obtain the relationship (linear or exponential) among attributes. The constructed models can be used to predict subsequent values of certain attributes [17]. Multivariate analysis is another technique used to analyze the relationship among multiple variables; for example, the analysis of variance (ANOVA) can be used to test the variances of means of variables [18].

Data mining techniques are commonly used to construct prediction models; for example, the classification technique

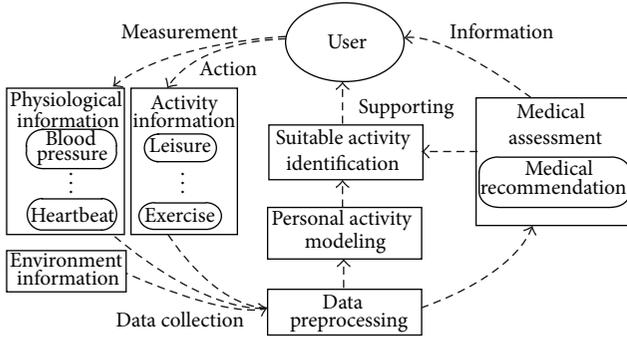


FIGURE 1: Architecture of the PAM approach.

is used to construct a decision tree that can be used to predict subsequent data. Before the modeling process, the collected data are preprocessed according to the mining technique used. The modeling process attempts to identify the association rules between these items [19, 20]. Association rule mining is another mining technique that can be used to construct prediction models by using the relational dataset. By appropriately setting the confidence and support, association rule mining can be used to identify many useful rules [21]. Clustering is another mining technique that groups collected data into several clusters according to the data characteristics. These clusters can be used to predict subsequent data [22, 23].

The main advantage of the proposed approach is that the obtained PAMs can be used not only to identify suitable activities but also to facilitate medical evaluation by medical personnel. The PAM can be constructed using real-time data collected from mobile devices. To demonstrate the performance of the proposed approach, the modeling techniques are applied in an experiment pertaining to health management. The results show that the constructed models can be used to select suitable activities. The remaining part of this paper is ordered as follows: the architecture of the proposed approach is introduced in Section 2, and the results and discussion are provided; finally, suggestions for further studies and conclusions pertaining to the proposed approach are presented.

## 2. Methods

This study proposed an approach to construct activity models; in the approach, modeling techniques are applied on the collected physiological data. The architecture of the proposed approach is depicted in Figure 1; personal physiological data and activity data are collected using wearable devices based on the predefined schema. The collected data are preprocessed and analyzed to construct the PAMs. The obtained models provide information that can be used to identify the proper activity or avoid improper activities. The proposed approach mainly involves the data collection (physiological data and activity data) and activity modeling. Physiological data, such as blood pressure, heart rate, and blood glucose, can be treated as retroactive information, and they can be collected automatically using mobile devices. The activity

TABLE 1: Example of integrating collected data.

| Time     | Environment information | Physiological information Attr. | Activity information   |
|----------|-------------------------|---------------------------------|------------------------|
|          | $a_1^e \cdots a_h^e$    | $a_1^p \cdots a_k^p$            | $a_1^a \cdots a_n^a$   |
| $t_1$    | $s_{11} \cdots s_{1h}$  | $v_{11} \cdots v_{1k}$          | $u_{11} \cdots u_{1n}$ |
| $t_2$    | $s_{21} \cdots s_{2h}$  | $v_{21} \cdots v_{2k}$          | $u_{21} \cdots u_{2n}$ |
| $t_3$    | $s_{31} \cdots s_{3h}$  | $v_{31} \cdots v_{3k}$          | $u_{31} \cdots u_{3n}$ |
| $\vdots$ | $\vdots$                | $\vdots$                        | $\vdots$               |
| $t_r$    | $s_{r1} \cdots s_{rh}$  | $v_{r1} \cdots v_{rk}$          | $u_{r1} \cdots u_{rn}$ |

data can be treated as proactive information and represent the activities of the user. In addition to physiological data and activity data, environmental information, such as the environmental temperature and humidity, can also be collected for the modeling process. The collected data are preprocessed according to the predefined schema and are used to construct the PAMs. The obtained models provide information for medical advice and recommendations for medical consultants. The medical recommendations are expressed as rules. The suitable activity identification (SAI) component identifies suitable activity according to the rules of medical advice and the obtained PAMs. The information can be used for selecting suitable activities.

Details of the implementation of these components, including data collection, data preprocessing, activity modeling, and SAI, are depicted in the following subsections.

**2.1. Modeling Techniques.** To facilitate data collection, a set of attributes should be defined to describe the physiological data and activities. The physiological data may be collected by different types of devices at different timestamps. The collected data are integrated and preprocessed according to the analysis engine. Table 1 shows integrated data collected at different timestamps;  $T = \{t_1, \dots, t_r\}$  denotes the timestamps,  $E = \{a_1^e, \dots, a_h^e\}$  denotes the attributes of environmental data, and  $P = \{a_1^p, \dots, a_k^p\}$  denotes the attributes of physiological data. The variables  $e_{ij}$  and  $v_{ij}$  denote environmental data and physiological data that can be collected automatically in a specified time interval by using wearable mobile devices. The environmental data also contain information on the device location (obtained from Global Positioning System), which indicates the location where the user performs activities.

The activity data contain information on the activities performed, while the physiological data contains information on the user's physiological status, which can be treated as an outcome of the performed activities. The performed activities cannot be recorded using a single term. The characteristics of an activity may be different for different persons. For example, the characteristics of the activity "jogging" are different for two different persons, and characteristics such as speed, duration, and direction (uphill or downhill) may affect the outcomes of the activities. For activity modeling, this study used the attributes of movement, such as the speed, direction, and acceleration, to represent the activities performed by

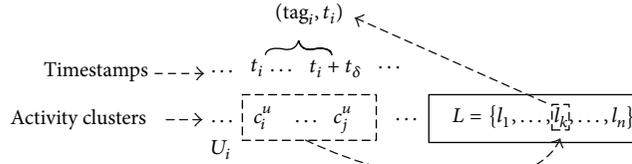


FIGURE 2: Linking the tags to the obtained activity patterns.

users. The information can be collected using the sensors of wearable mobile devices, such as the G-sensor. In Table 1,  $A = \{a_1^a, \dots, a_n^a\}$  denotes the attributes of activity, and  $u_{ij}$  denotes the collected activity data.

The data, including the environmental, physiological, and activity data, are collected by wearable devices at different timestamps. An activity can be represented as a sequence of activity data. For example, let  $U = \{u_1, \dots, u_r\}$  denote the collected activity data and let  $M = \{m_1, \dots, m_k\}$  denote the set of activities performed by the user, where  $m_i = \{u_j, \dots, u_k\} \subset U$  denotes an activity contained in  $U$ . An activity can be represented as a sequence of  $\{u_i\}$  (for some  $i$ ). The algorithm used to identify the possible activities is shown as follows.

- (1) Assume that the set of clusters  $C^u = \{c_1^u, \dots, c_r^u\}$  is obtained by applying the clustering technique to the collected activity data  $U$ .
- (2) The elements of  $U$  are considered to represent clusters of  $C^u$ ; that is,  $U_c = \{c_i^u \mid c_i^u \in C^u\}$ .
- (3) Use  $U_c$  to identify the frequent itemsets  $L = \{l_1, \dots, l_n\}$ , where  $l_i \subset U_c$ , such that the number of occurrences (*support*) of the patterns is greater than a preselected threshold  $\delta_l$ .

The modeling process identifies the relationship between  $L$  and the physiological information. Each activity is assigned a tag by the user for easy identification. For example, an activity can be expressed as  $(\text{tag}_i, t_i)$ , where  $\text{tag}_i$  denotes the activity tag (such as jogging or climbing) and  $t_i$  denotes the timestamp. The steps used to link the tags to the obtained activity patterns are shown in Figure 2;  $t_\delta$  denotes the preselected duration and is less than the duration of possible activities. For a tag  $(\text{tag}_i, t_i)$ , a sequence of activity clusters  $U_i = \{c_i^u, \dots, c_j^u\}$  within the time interval  $[t_i, t_i + t_\delta]$  is selected. For the pattern  $l_k$  that is satisfied,  $U_i$  is linked to  $\text{tag}_i$ . A pattern  $l_k$  is said to satisfy an activity cluster  $U_i$  if all elements of  $l_k$  are contained in  $U_i$  in order.

The main purpose of data preprocessing is to apply the cluster technique to the collected data to identify possible activities that can be linked to the physiological data. The identified activities can be used to construct PAMs.

**2.2. Activity Modeling.** The main purpose of personal activity modeling is to find the relationship between activity and physiological data. The collected physiological data are grouped according to the activity transactions, such as the set of activity clusters within a time interval  $t_\delta$ . For example, in Figure 3, the activity clusters  $\{c_i^u, \dots, c_{i+1}^u\}$  form an activity

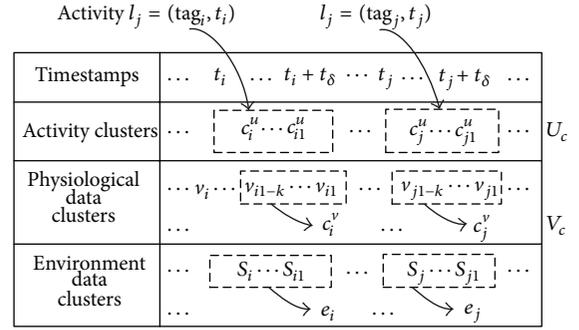


FIGURE 3: Clustering the collected physiological data based on timestamps.

transaction;  $\{v_i, \dots, v_{i+1}\}$  denote the set of physiological data based on the activity transaction and can be treated as representing the personal physiological state in the time interval  $[t_i, t_i + t_\delta]$  (the effects of the performed activities). For personal activity analysis, the last  $k$  physiological data  $\{v_{i-k}, \dots, v_{i+1}\}$  are selected to represent the effects of the activity transaction and are denoted by a centroid. The personal activity modeling process can be divided into the following steps.

In the first step, the clustering technique is applied to the physiological data of all activity transactions to obtain *physiological clusters*  $C^v = \{c_i^v \mid c_i^v \in C^v\}$ , which indicate the effects of certain activities. Figure 3 shows the physiological clusters  $V_c = \{c_i^v \mid c_i^v \in C^v\}$  obtained from the collected physiological data. The purpose of this step is to distribute the collected physiological data among several clusters, with all the data contained in a cluster having similar characteristics (attributes).

In the second step, the clustering technique is applied to the environmental data to obtain *environment clusters*. Since the effects of an activity may be influenced by environmental parameters, such as temperature or humidity, the environment clusters can be used to group the activity data. For an activity transaction, a sequence of environmental data, such as  $\{s_i, \dots, s_{i+1}\}$ , can be selected and represented by a center  $o_i$ . The clustering technique can be applied to these centers  $\{o_1, \dots, o_n\}$  to obtain environment clusters  $E = \{e_i, \dots, e_n\}$ . Therefore, an activity transaction can be represented as  $\text{tr}_i = (\{c_i^u, \dots, c_{i+1}^u\}, e_i; c_i^v, [b_i])$ , where a suitability tag  $b_i \in \{1, 0, -1\}$  can be attached to  $c_i^v$ . The identified transactions of activities are denoted as  $\text{TR} = \{\text{tr}_1, \dots, \text{tr}_n\}$ .

In the third step, the association rule mining technique is applied to the obtained transactions  $\text{TR}$  to construct PAMs. The PAMs can be expressed as a set of rules, and each rule contains antecedent and subsequent items. The following algorithm is used to generate the antecedent items.

- (1) Calculate the support (the number) for each activity cluster  $c_i^u \in U_c$  using the obtained transactions  $\text{TR}$  and select the items with a support greater than  $\text{supp}_a$  (preselected) to form the first-level large itemset  $L_1$ .
- (2) Generate the second-level itemset  $L_1 \times L_1$ , and select the items with a support greater than  $\text{supp}_a$  to form the second-level large itemset  $L_2$ .

- (3) The process is repeated until no more large itemsets can be formed. The last  $n$ -level large itemset is denoted as  $L_n$ .

The obtained large itemset  $L_n$  contains the items, such as  $(c_i^u, \dots, c_j^u)$ , that appear in the collected dataset frequently. The PAMs can represent the relationship between activities and physiological clusters. The physiological clusters  $V_c$  denote the subsequent items, and the algorithm used to construct PAMs is as follows.

- (1) Calculate the support for each physiological cluster  $c_i^v \in V_c$ , and select the items with support greater than  $\text{supp}_a$  to form the large itemset  $L_R$ .
- (2) Generate the candidate rules  $L_n \times L_R$ , and, for each item  $r_i \in L_n \times L_R$ , count the number of appearances in TR. The total number of appearances shown in the antecedent part is denoted as  $n_a$ , and the number of appearances shown in both antecedent and subsequent parts is denoted as  $n_b$ . The confidence of the rule  $r_i$  is calculated as  $n_b/n_a$ . A rule is selected if its *confidence* is greater than the preselected *threshold confidence*.

An obtained rule  $r_i$  can be expressed as  $\{c_i^u, \dots, c_j^u\} \rightarrow \{c_i^v\}$ , indicating that an activity  $l_i$  with the pattern  $\{c_i^u, \dots, c_j^u\}$  causes physiological effect  $c_i^v$ . The environmental information can be used to indicate the environment where the obtained rules can be applied and where the confidences of these rules with the environment cluster  $e_i$  are recalculated. The confidence of the suitability of the rules can be evaluated using the attached suitability tag, and the confidences of the rules with the suitability tag  $b_i \in B = \{1, 0, -1\}$  are recalculated. The rule  $r_i$  with an environment cluster and a suitability tag can be expressed as  $\{c_i^u, \dots, c_j^u, [e_i]\} \rightarrow \{c_i^v, [b_i]\}$ , and the generated activity rule set is denoted as  $R$ .

**2.3. Activity Identification.** The obtained PAMs provide information that can be used to evaluate of suitability of subsequent activities. The evaluation process can be done before or during an activity, and it can be described as follows. First, when applying the obtained rules to select suitable activities, the environmental data collected at the current timestamp are clustered and used to select rules. For example, in Figure 4, let  $\{s_{i-k}, \dots, s_i\}$  denote the data collected from the current environment and let it be represented as  $e_i^u \in E$ ; the rules with the same environmental data are selected such that  $e_i^u = e_i$  (the same cluster).  $L_s = \{l_1, \dots, l_k\}$  denotes the possible activity patterns.

Second, the obtained rules can be used to evaluate the suitability of the activity. Since the activity data are collected during activities, the collected data can be used to evaluate the suitability of the activity. As shown in Figure 4, let  $U_i^u = \{c_{i-k}^u, \dots, c_i^u\}$  denote the activity cluster of the data collected in the time interval  $[t_{i-\delta}, t_i]$  when an activity is being performed. Since the activity is being performed by the user, the new data  $U_i^u$  are collected continuously. The collected data may contain only the first several activity clusters of certain rules ( $r_i \in R$ , for some  $i$ ). Therefore, these rules can be selected

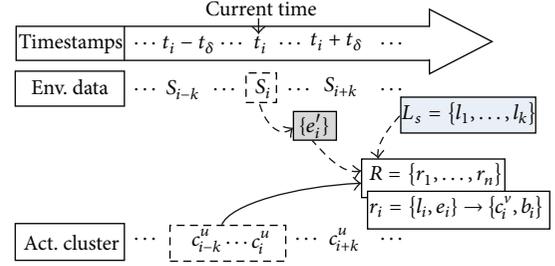


FIGURE 4: Identification of suitable activities and evaluation of the current activity.

TABLE 2: The selected attributes of the experiment.

| Number | Data sources | Attr.  | Description                     |
|--------|--------------|--------|---------------------------------|
| 1      | Timestamps   | MTime  | The timestamp of records        |
| 2      | Env. data    | ETemp  | The environmental temperature   |
| 3      | Env. data    | Humid  | The environmental humidity      |
| 4      | Phy. data    | SBP    | The systolic blood pressure     |
| 5      | Phy. data    | DBP    | The diastolic blood pressure    |
| 6      | Phy. data    | HB     | The heart beat                  |
| 7      | Phy. data    | BTemp  | The body temperature            |
| 8      | Act. data    | ActX   | The x-axis value of accelerator |
| 9      | Act. data    | ActY   | The y-axis value of accelerator |
| 10     | Act. data    | ActZ   | The z-axis value of accelerator |
| 11     | Act. data    | ActTag | The activity tags               |

as candidate rules of  $U_i^u$ , and the evaluation results can be derived by considering the selected rules. For example, let the rule  $r_i = \{l_i, e_i\} \rightarrow \{b_i\}$  be selected as a candidate rule. If  $c_i^u \in l_i$  and  $c_i^u \in U_i^u$ , then any  $c_j^u \in l_i$  ( $j < i$ ) is also contained in  $U_i^u$  ( $U_i^u$  contains the first several activity clusters of the rule  $r_i$ ). Let  $L^1$  denote the candidate rules that are suitable (for  $b_i = 1$ ), and let  $L^{-1}$  denote the candidate rules that are unsuitable. Then,  $L^1$  and  $L^{-1}$  can be used to evaluate the suitability of the activity.

### 3. Results

To demonstrate the construction of the PAMs using the proposed approach, an experiment was designed. Table 2 shows the set of attributes used to collect data, and *data sources* column denotes the sources from where data are collected. The environmental attributes consist of the environmental temperature (ETemp) and humidity (Humid). The physiological attributes include the blood pressure (SBP and DBP), heart rate (HB), and body temperature (BTemp). The activity attributes include the x-, y-, and z-axis (ActX, ActY, and ActZ) values of the accelerator, and ActTag denotes the activity tag.

The activity data are collected (every second) using a mobile device. An activity tag is assigned by the user before performing the activity; for example, go uphill (U), downhill (D), or on a smooth road (F). The physiological data are collected (every 5 s) using an electric sphygmomanometer,

TABLE 3: Example of collected data.

| Number | MTime                 | ETemp | Humi | ActX  | ActY  | ActZ  | ActTag | BTemp | SBP | DBP | HB |
|--------|-----------------------|-------|------|-------|-------|-------|--------|-------|-----|-----|----|
| 0      | 2014/4/21 11:23:17 PM | 24.2  | 63.2 | -0.96 | 0.88  | 7.73  | U      | 36.1  | 115 | 62  | 76 |
| 1      | 2014/4/21 11:23:18 PM | 24.2  | 63.2 | -0.23 | 1.62  | 8.69  | U      | 36.1  | 115 | 62  | 76 |
| 2      | 2014/4/21 11:23:19 PM | 24.2  | 63.2 | -1.69 | 1.88  | 7.73  | U      | 36.1  | 115 | 62  | 76 |
| 3      | 2014/4/21 11:23:20 PM | 24.2  | 63.2 | 0.04  | 1.62  | 9.19  | U      | 36.1  | 115 | 62  | 76 |
| 4      | 2014/4/21 11:23:21 PM | 24.2  | 63.2 | 0.81  | 0.69  | 6.81  | U      | 36.1  | 117 | 62  | 76 |
| 5      | 2014/4/21 11:23:22 PM | 24.2  | 63.1 | 0.69  | -1.00 | 14.38 | U      | 36.1  | 117 | 74  | 78 |
| 6      | 2014/4/21 11:23:23 PM | 24.2  | 63.1 | 0.35  | 2.12  | 10.81 | U      | 36.1  | 117 | 74  | 78 |
| 7      | 2014/4/21 11:23:24 PM | 24.2  | 63.1 | -0.54 | 1.23  | 6.31  | U      | 36.0  | 117 | 74  | 78 |
| 8      | 2014/4/21 11:23:25 PM | 24.2  | 63.2 | -1.27 | 0.12  | 14.08 | U      | 36.0  | 117 | 74  | 78 |
| 9      | 2014/4/21 11:23:26 PM | 24.2  | 63.2 | 0.62  | 1.73  | 5.92  | U      | 36.0  | 117 | 74  | 78 |
| 10     | 2014/4/21 11:23:27 PM | 24.2  | 63.2 | 1.15  | 1.12  | 9.31  | U      | 36.0  | 125 | 73  | 86 |
| 11     | 2014/4/21 11:23:28 PM | 24.2  | 63.2 | -0.73 | 1.85  | 8.50  | U      | 36.0  | 125 | 73  | 86 |
| 12     | 2014/4/21 11:23:29 PM | 24.2  | 63.1 | -0.54 | 0.46  | 8.88  | U      | 36.0  | 125 | 73  | 86 |
| 13     | 2014/4/21 11:23:30 PM | 24.2  | 63.1 | -0.81 | -0.12 | 11.73 | U      | 35.9  | 125 | 73  | 86 |
| 14     | 2014/4/21 11:23:31 PM | 24.2  | 63.2 | 0.42  | -0.62 | 9.11  | U      | 35.9  | 125 | 73  | 86 |
| 15     | 2014/4/21 11:23:32 PM | 24.2  | 63.2 | 0.38  | 1.62  | 8.27  | U      | 35.9  | 119 | 78  | 85 |
| 16     | 2014/4/21 11:23:33 PM | 24.2  | 63.2 | 0.31  | 3.42  | 6.77  | D      | 35.9  | 119 | 78  | 85 |
| 17     | 2014/4/21 11:23:34 PM | 24.2  | 63.2 | 0.19  | 1.23  | 6.31  | D      | 35.9  | 119 | 78  | 85 |
| 18     | 2014/4/21 11:23:35 PM | 24.2  | 63.2 | 0.31  | 0.15  | 5.04  | D      | 35.9  | 119 | 78  | 85 |
| 19     | 2014/4/21 11:23:36 PM | 24.2  | 63.2 | 2.58  | 2.88  | 14.42 | D      | 35.9  | 119 | 78  | 85 |

TABLE 4: The activity clusters obtained from collected data.

| Data number | MTime                 | ActX  | ActY  | ActZ  | ActTag | Cluster |
|-------------|-----------------------|-------|-------|-------|--------|---------|
| 0           | 2014/4/21 11:23:17 PM | -0.96 | 0.88  | 7.73  | U      | 0       |
| 1           | 2014/4/21 11:23:18 PM | -0.23 | 1.62  | 8.69  | U      | 2       |
| 2           | 2014/4/21 11:23:19 PM | -1.69 | 1.88  | 7.73  | U      | 0       |
| 3           | 2014/4/21 11:23:20 PM | 0.04  | 1.62  | 9.19  | U      | 2       |
| 4           | 2014/4/21 11:23:21 PM | 0.81  | 0.69  | 6.81  | U      | 4       |
| 5           | 2014/4/21 11:23:22 PM | 0.69  | -1.00 | 14.38 | U      | 3       |
| 6           | 2014/4/21 11:23:23 PM | 0.35  | 2.12  | 10.81 | U      | 2       |

and they are transmitted to a mobile device over a Bluetooth interface. The environmental data, including temperature and humidity, are also collected, and they are transmitted to the mobile device using a Bluetooth interface. The collected data consist of 143 data items collected over five different days. Table 3 shows an example of the collected data based on the attributes shown in Table 2. The collected data are preprocessed and used to construct PAMs.

## 4. Discussion

*4.1. Suitable Activity Identification.* The activity data, including ActX, ActY, and ActZ, are used to cluster the collected data. Table 4 shows the activity clusters obtained from the collected data shown in Table 2; the number of clusters is selected as 5 (denoted by the numbers 0 to 4). ActTag is not used as the attribute for clustering and is intended only for helping the user recognize the activity.

The clustered activities are used to form activity transactions for a transaction length of 6. The items of a new transaction are selected from the previous transaction by shifting the items to the right by one item. For example, the first transaction is selected from item 0 to item 5 (0, 2, 0, 2, 4, 3), while the next transaction is selected from 1 to 6 (2, 0, 2, 4, 3, 2). An activity tag, denoted as TranTag, is attached to each transaction for recognizing the transaction. TranTag can be determined by the number of activity tags appearing in the transaction. The obtained transactions can be used to name the identified activities. In addition to the activity tags, a physiological cluster is associated with each transaction for indicating the effects of the activity transaction. The physiological clusters of the transactions are selected as follows. First, for each transaction, the last three physiological data items are selected and used for calculating a centroid. The centroid can be used to represent the effects of the transaction. Second, the obtained centroids are clustered. Table 5 shows an example of physiological clusters obtained

TABLE 5: The physiological clusters.

| Tran. number | Physiological centroid          | Phy. cluster |
|--------------|---------------------------------|--------------|
| 0            | 36.100, 115.000, 62.000, 76.000 | 0            |
| 1            | 36.100, 116.000, 68.000, 77.000 | 1            |
| 2            | 36.050, 117.000, 74.000, 78.000 | 1            |
| 7            | 36.000, 121.000, 73.500, 82.000 | 3            |
| 8            | 35.950, 125.000, 73.000, 86.000 | 3            |
| 9            | 35.900, 125.000, 73.000, 86.000 | 3            |
| 17           | 35.900, 119.500, 78.000, 79.000 | 4            |
| 18           | 35.900, 119.500, 78.000, 79.000 | 4            |
| 71           | 36.200, 123.000, 72.500, 79.500 | 2            |
| 72           | 36.150, 126.000, 73.000, 79.000 | 2            |
| 73           | 36.100, 126.000, 73.000, 79.000 | 2            |

TABLE 6: The environment clusters.

| Tran. number | Environment centroid | Env. cluster |
|--------------|----------------------|--------------|
| 0            | 24.200, 63.183       | 1            |
| 1            | 24.200, 63.167       | 1            |
| 2            | 24.200, 63.150       | 1            |
| 43           | 28.100, 83.200       | 0            |
| 44           | 28.100, 83.200       | 0            |
| 45           | 28.100, 83.200       | 0            |

TABLE 7: The activity transactions.

| Tran. number | Activity transaction | TranTag | Phy. cluster | Env. cluster |
|--------------|----------------------|---------|--------------|--------------|
| 0            | 0, 2, 0, 2, 4, 3     | U       | 0            | 1            |
| 1            | 2, 0, 2, 4, 3, 2     | U       | 1            | 1            |
| 2            | 0, 2, 4, 3, 2, 4     | U       | 1            | 1            |
| 7            | 4, 3, 4, 2, 2, 1     | U       | 3            | 1            |
| 17           | 4, 4, 3, 4, 3, 3     | D       | 4            | 1            |

by applying  $k$ -means (the number of clusters is 5) to the physiological centroids of the transactions.

The environment clusters are obtained using the collected environmental data. The centroids of these environmental data are computed and clustered. Examples of obtained environment centroids and clusters are shown in Table 6; the environmental parameters, including temperature and humidity, do not change rapidly and are clustered into two clusters (denoted as 0 and 1). The transaction tags, physiological clusters, and environment clusters are added to the transactions  $(\{c_i^u, \dots, c_j^u\}, e_i; c_i^v, [b_i])$ . The suitability tags  $b_i$  are assigned by medical personnel according to the physiological cluster of the transaction. Table 7 shows an example of obtained transactions, and the total number of obtained transactions is 137.

**4.2. Personal Activity Modeling.** The activity transaction with physiological clusters  $\{c_i^u, \dots, c_j^u\} \rightarrow \{c_i^v\}$  can be used to construct PAMs. Table 8 shows the PAMs obtained using association mining; the minimum support is 0.04 and the minimum confidence is 80%. *Activity items* denote the antecedent items,

TABLE 8: Activity models with physiological clusters.

| Rule number | Activity items    | Phy. item              | Support       | Confidence    |
|-------------|-------------------|------------------------|---------------|---------------|
| 0           | 0, 0, 0           | $\rightarrow$ 0        | 0.1314        | 1.0000        |
| 1           | 0, 1, 3           | $\rightarrow$ 2        | 0.0438        | 1.0000        |
| 2           | <b>2, 0, 0</b>    | $\rightarrow$ <b>0</b> | <b>0.0730</b> | <b>0.9000</b> |
| 3           | 2, 0, 1           | $\rightarrow$ 2        | 0.0511        | 0.8571        |
| 4           | 4, 2, 0           | $\rightarrow$ 1        | 0.0511        | 1.0000        |
| 5           | 0, 0, 0, 0        | $\rightarrow$ 0        | 0.0584        | 1.0000        |
| 6           | 0, 4, 1, 3        | $\rightarrow$ 2        | 0.0438        | 1.0000        |
| 7           | <b>2, 0, 0, 0</b> | $\rightarrow$ <b>0</b> | <b>0.0511</b> | <b>1.0000</b> |
| 8           | 2, 0, 2, 2        | $\rightarrow$ 0        | 0.0511        | 0.8571        |
| 9           | 2, 2, 0, 0        | $\rightarrow$ 0        | 0.0511        | 1.0000        |
| 10          | 2, 2, 2, 0        | $\rightarrow$ 0        | 0.0657        | 0.8889        |
| 11          | 2, 2, 2, 2        | $\rightarrow$ 0        | 0.0803        | 1.0000        |
| 12          | 1, 1, 1, 1        | $\rightarrow$ 0        | 0.1022        | 0.9286        |
| 13          | 2, 2, 2, 2, 2     | $\rightarrow$ 0        | 0.0438        | 1.0000        |
| 14          | 1, 1, 1, 1, 1     | $\rightarrow$ 0        | 0.0584        | 1.0000        |

while the *phy. item* represents the subsequent item of the rule. *Support* indicates the frequency of appearance of the patterns, while *confidence* denotes the accuracy of the rule. For example, rule 2 indicates that an activity with the pattern  $\{2, 0, 0\}$  may yield physiological cluster 0 with 7% support and 90% confidence. This pattern can further be extended to  $\{2, 0, 0, 0\}$ , which also yields cluster 0 with support 5% and confidence 100%.

The environment clusters  $e_i$  of the transaction indicate the situation in which the activities are performed. Different environment clusters may be associated with different effects. Therefore, the environment clusters should be considered while applying the rule on prediction of suitable activity. The confidences of the obtained rules with environment clusters are shown in Table 9; no minimum confidence is used to filter the rules. For example, according to rule 2, an activity with the pattern  $p_1 = \{2, 0, 0\}$  may yield physiological cluster 0 with confidence 90%, with all transactions with pattern  $p_1$  (100%) contained in environment cluster 0. However, when rule 3 is applied, an activity with the pattern  $p_2 = \{2, 0, 1\}$  may yield physiological cluster 2 with confidence 86%, with only 71% of transactions with pattern  $p_2$  being contained in environment cluster 0.

For determining the suitability of activities, the physiological clusters are evaluated by medical personnel and suitability tags are attached to the transactions. In this study, the suitability tags are evaluated according to the obtained physiological centroids and clusters (as shown in Table 10). The suitability tag is 0 for physiological clusters 0 and 1, 1 for physiological clusters 2 and 3, and  $-1$  for physiological cluster 4. An example of activity transactions with suitability tags (without minimum confidence) is shown in Table 11, in which the confidences for suitability tags  $-1, 0,$  and  $1$  are shown in columns 7, 8, and 9, respectively.

The confidences for suitability tags indicate the probability of the corresponding rule yielding suitable physiological

TABLE 9: Confidence of activity models with environment clusters.

| Rule number | Activity items | Phy. item | Support  | Confidence    | Env. confidence |               |               |
|-------------|----------------|-----------|----------|---------------|-----------------|---------------|---------------|
|             |                |           |          |               | Cluster 0       | Cluster 1     |               |
| 0           | 0, 0, 0        | →         | 0        | 0.1314        | 1.0000          | 1.0000        | 0.0000        |
| 1           | 0, 1, 3        | →         | 2        | 0.0438        | 1.0000          | 0.5000        | 0.5000        |
| 2           | <b>2, 0, 0</b> | →         | <b>0</b> | <b>0.0730</b> | <b>0.9000</b>   | <b>1.0000</b> | <b>0.0000</b> |
| 3           | <b>2, 0, 1</b> | →         | <b>2</b> | <b>0.0511</b> | <b>0.8571</b>   | <b>0.7143</b> | <b>0.2857</b> |
| 4           | 4, 2, 0        | →         | 1        | 0.0511        | 1.0000          | 1.0000        | 0.0000        |
| 5           | 0, 0, 0, 0     | →         | 0        | 0.0584        | 1.0000          | 1.0000        | 0.0000        |
| 6           | 0, 4, 1, 3     | →         | 2        | 0.0438        | 1.0000          | 0.5000        | 0.5000        |
| 7           | 2, 0, 0, 0     | →         | 0        | 0.0511        | 1.0000          | 1.0000        | 0.0000        |
| 8           | 2, 0, 2, 2     | →         | 0        | 0.0511        | 0.8571          | 0.8571        | 0.1429        |
| 9           | 2, 2, 0, 0     | →         | 0        | 0.0511        | 1.0000          | 1.0000        | 0.0000        |
| 10          | 2, 2, 2, 0     | →         | 0        | 0.0657        | 0.8889          | 0.8889        | 0.1111        |
| 11          | 2, 2, 2, 2     | →         | 0        | 0.0803        | 1.0000          | 1.0000        | 0.0000        |
| 12          | 1, 1, 1, 1     | →         | 0        | 0.1022        | 0.9286          | 1.0000        | 0.0000        |
| 13          | 2, 2, 2, 2, 2  | →         | 0        | 0.0438        | 1.0000          | 1.0000        | 0.0000        |
| 14          | 1, 1, 1, 1, 1  | →         | 0        | 0.0584        | 1.0000          | 1.0000        | 0.0000        |

TABLE 10: The activity transactions with suitability tags.

| Tran. number | Activity transaction | TranTag | Phy. cluster | Env. cluster | Suitability tag |
|--------------|----------------------|---------|--------------|--------------|-----------------|
| 0            | 0, 2, 0, 2, 4, 3     | U       | 0            | 0            | 0               |
| 1            | 2, 0, 2, 4, 3, 2     | U       | 1            | 0            | 0               |
| 2            | 0, 2, 4, 3, 2, 4     | U       | 1            | 0            | 0               |
| 7            | 4, 3, 4, 2, 2, 1     | U       | 3            | 0            | 1               |
| 17           | 4, 4, 3, 4, 3, 3     | D       | 4            | 0            | -1              |
| 30           | 2, 2, 1, 0, 4, 1     | D       | 2            | 0            | 1               |

clusters. For example, the *confidence* of the suitability tag 0 of rule 2 is 90% which means the probability to cause the physiological cluster 2 is approximately 90%. An activity satisfying a rule with higher suitability can be treated as suitable activities for users. Therefore, rules 1, 3, and 6 shown in Table II can be used to identify suitable activities. An activity satisfying the patterns of suitable rules can be identified as suitable activity.

The TranTag associated with each transaction can be used to name the identified activities. Table 12 shows the confidences for different activity tags, including D (downhill), U (uphill), and F (smooth road). The activity tags with high confidences can be used for naming the identified activities, such as F for rules 0 and 5 and D for rule 4. The D and U activities may contain pattern  $\{0, 1, 3\}$ .

**4.3. Suitable Activity Identification.** The obtained PAMs can be used to identify suitable activities. In this subsection, the application of the obtained PAMs to another test dataset is discussed for demonstrating how the proposed approach can be applied to identify suitable activities. The attributes of the test dataset are displayed in Table 1. The test dataset contains 70 data items collected from different activities, including 28 U activities, 28 D activities, and 14 F activities (the dataset is available at <http://mgiga.com.tw/PAM/>). The test dataset are divided into 10 transactions, and each transaction (containing 7 data items) denotes an activity.

Each activity can be treated as a sequence of activity items. The performed activity items in the sequence are used to evaluate the *activity* (which is currently being performed), while the physiological data of the last three activities are used to validate the evaluations (the selection of the physiological data of a transaction is identical to the process presented in Section 4.1). Table 13 shows an example of a test dataset. The suitable activity identification process can be described as follows.

First, the collected activity items are clustered based on the centroids of the activity clusters of the obtained PAMs. Table 14 shows the centroids of the activity clusters obtained in Section 4.2. The *norm* denotes the module of the centroid of each cluster that indicates the strength of the activities. The rule  $r_6 = \{0, 4, 1, 3\}$  shown in Table II can be expressed as the dash line shown in Figure 5, while the solid line depicts the rule  $r_8 = \{0, 2, 1, 3\}$ . Second, the PAMs (shown in Table 12) are applied to the current activity clusters, and they provide information (suitability) on the current activity. For example, when the first activity item  $\{0.19, -1.04, 13.73\}$  is collected and classified into cluster 3, there is no rule shown in Table II which can be selected. The second activity item  $\{0.88, -0.58, 7.35\}$  is then collected and classified as 0. The activity transaction is  $\{3, 0\}$ , and rules 0, 1, 5, and 6 are selected; among these rules, no rule is unsuitable, while rules 1 and 6 are suitable. Therefore, the performing activity may be

TABLE II: Confidence of activity models with suitability tags.

| Rule number | Activity items |   | Phy. item | Support       | Confidence    | Confidences for suitability tags |               |               |
|-------------|----------------|---|-----------|---------------|---------------|----------------------------------|---------------|---------------|
|             |                |   |           |               |               | -1                               | 0             | 1             |
| 0           | 0, 0, 0        | → | 0         | 0.1314        | 1.0000        | 0.0000                           | 1.0000        | 0.0000        |
| 1           | 0, 1, 3        | → | 2         | 0.0438        | 1.0000        | 0.0000                           | 0.0000        | <b>1.0000</b> |
| 2           | <b>2, 0, 0</b> | → | <b>0</b>  | <b>0.0730</b> | <b>0.9000</b> | 0.0000                           | <b>0.9000</b> | 0.1000        |
| 3           | <b>2, 0, 1</b> | → | <b>2</b>  | <b>0.0511</b> | <b>0.8571</b> | 0.0000                           | 0.1429        | <b>0.8571</b> |
| 4           | 4, 2, 0        | → | 1         | 0.0511        | 1.0000        | 0.0000                           | 1.0000        | 0.0000        |
| 5           | 0, 0, 0, 0     | → | 0         | 0.0584        | 1.0000        | 0.0000                           | 1.0000        | 0.0000        |
| 6           | 0, 4, 1, 3     | → | 2         | 0.0438        | 1.0000        | 0.0000                           | 0.0000        | <b>1.0000</b> |
| 7           | 2, 0, 0, 0     | → | 0         | 0.0511        | 1.0000        | 0.0000                           | 1.0000        | 0.0000        |
| 8           | 2, 0, 2, 2     | → | 0         | 0.0511        | 0.8571        | 0.0000                           | 1.0000        | 0.0000        |
| 9           | 2, 2, 0, 0     | → | 0         | 0.0511        | 1.0000        | 0.0000                           | 1.0000        | 0.0000        |
| 10          | 2, 2, 2, 0     | → | 0         | 0.0657        | 0.8889        | 0.0000                           | 1.0000        | 0.0000        |
| 11          | 2, 2, 2, 2     | → | 0         | 0.0803        | 1.0000        | 0.0000                           | 1.0000        | 0.0000        |
| 12          | 1, 1, 1, 1     | → | 0         | 0.1022        | 0.9286        | 0.0000                           | 1.0000        | 0.0000        |
| 13          | 2, 2, 2, 2, 2  | → | 0         | 0.0438        | 1.0000        | 0.0000                           | 1.0000        | 0.0000        |
| 14          | 1, 1, 1, 1, 1  | → | 0         | 0.0584        | 1.0000        | 0.0000                           | 1.0000        | 0.0000        |

TABLE 12: The activity models with activity tags.

| Rule number | Activity items |   | Phy. item | Support | Confidence | Conf. for activity tags |        |               |
|-------------|----------------|---|-----------|---------|------------|-------------------------|--------|---------------|
|             |                |   |           |         |            | D                       | U      | F             |
| 0           | 0, 0, 0        | → | 0         | 0.1314  | 1.0000     | 0.0000                  | 0.0000 | <b>1.0000</b> |
| 1           | 0, 1, 3        | → | 2         | 0.0438  | 1.0000     | 0.5000                  | 0.5000 | 0.0000        |
| 2           | 2, 0, 0        | → | 0         | 0.0730  | 0.9000     | 0.0000                  | 0.1000 | 0.9000        |
| 3           | 2, 0, 1        | → | 2         | 0.0511  | 0.8571     | 0.2857                  | 0.5714 | 0.1429        |
| 4           | 4, 2, 0        | → | 1         | 0.0511  | 1.0000     | <b>1.0000</b>           | 0.0000 | 0.0000        |
| 5           | 0, 0, 0, 0     | → | 0         | 0.0584  | 1.0000     | 0.0000                  | 0.0000 | <b>1.0000</b> |
| 6           | 0, 4, 1, 3     | → | 2         | 0.0438  | 1.0000     | 0.5000                  | 0.5000 | 0.0000        |
| 7           | 2, 0, 0, 0     | → | 0         | 0.0511  | 1.0000     | 0.0000                  | 0.0000 | 1.0000        |
| 8           | 2, 0, 2, 2     | → | 0         | 0.0511  | 0.8571     | 0.0000                  | 0.1429 | 0.8571        |
| 9           | 2, 2, 0, 0     | → | 0         | 0.0511  | 1.0000     | 0.0000                  | 0.0000 | 1.0000        |
| 10          | 2, 2, 2, 0     | → | 0         | 0.0657  | 0.8889     | 0.1111                  | 0.0000 | 0.8889        |
| 11          | 2, 2, 2, 2     | → | 0         | 0.0803  | 1.0000     | 0.0000                  | 0.0000 | 1.0000        |
| 12          | 1, 1, 1, 1     | → | 0         | 0.1022  | 0.9286     | 0.0714                  | 0.0714 | 0.8571        |
| 13          | 2, 2, 2, 2, 2  | → | 0         | 0.0438  | 1.0000     | 0.0000                  | 0.0000 | 1.0000        |
| 14          | 1, 1, 1, 1, 1  | → | 0         | 0.0584  | 1.0000     | 0.0000                  | 0.0000 | 1.0000        |

TABLE 13: Example of collected testing data.

| Number | MTime       | ETemp | Humi | ActX  | ActY  | ActZ  | ActTag | BTemp | SBP | DBP | HB |
|--------|-------------|-------|------|-------|-------|-------|--------|-------|-----|-----|----|
| 7      | 10:46:11 PM | 25.0  | 67.4 | 0.19  | -1.04 | 13.73 | U      | 35.9  | 120 | 76  | 78 |
| 8      | 10:46:12 PM | 25.0  | 67.4 | 0.88  | -0.58 | 7.35  | U      | 35.9  | 120 | 76  | 78 |
| 9      | 10:46:13 PM | 25.0  | 67.4 | -0.58 | 0.08  | 9.15  | U      | 35.9  | 120 | 76  | 78 |
| 10     | 10:46:14 PM | 25.0  | 67.5 | -1.73 | -0.04 | 6.73  | U      | 35.9  | 124 | 72  | 80 |
| 11     | 10:46:15 PM | 25.1  | 67.5 | -0.35 | 0.96  | 14.58 | U      | 35.9  | 124 | 72  | 80 |
| 12     | 10:46:16 PM | 25.1  | 67.5 | 0.04  | 1.65  | 12.84 | U      | 35.9  | 124 | 72  | 80 |
| 13     | 10:46:17 PM | 25.1  | 67.5 | 0.96  | 0.65  | 6.08  | U      | 35.9  | 124 | 72  | 80 |

TABLE 14: Centroids of the activity clusters of activity models.

| Cluster | ActX   | ActY   | ActZ   | Norm  |
|---------|--------|--------|--------|-------|
| 0       | -0.493 | 0.918  | 7.632  | 7.70  |
| 1       | -0.002 | -0.134 | 9.590  | 9.59  |
| 2       | -0.302 | 1.652  | 9.702  | 9.85  |
| 3       | 0.358  | 0.792  | 13.053 | 13.08 |
| 4       | 0.343  | 1.200  | 6.407  | 6.53  |

TABLE 15: Test transactions.

| Tran. number | Transaction         | Suitability prediction | Applied rule | Suitability confidence |
|--------------|---------------------|------------------------|--------------|------------------------|
| 0            | 1, 4, 1, 2, 1, 0, 3 | 0                      | 4            | 100%                   |
| 1            | 3, 0, 1, 0, 3, 3, 4 | 1                      | 1            | 100%                   |
| 2            | 2, 4, 0, 3, 0, 2, 3 | 0                      | 2            | 90%                    |
| 3            | 4, 3, 0, 4, 2, 2, 4 | —                      | —            | —                      |
| 4            | 0, 1, 0, 4, 1, 2, 3 | 1                      | 6            | 100%                   |
| 5            | 0, 2, 4, 0, 1, 2, 4 | 1                      | 3            | 86%                    |
| 6            | 2, 2, 4, 4, 0, 3, 4 | —                      | —            | —                      |
| 7            | 0, 4, 3, 2, 4, 0, 3 | 0                      | 4            | 100%                   |
| 8            | 4, 3, 3, 3, 2, 1, 0 | 0                      | 4            | 100%                   |
| 9            | 3, 1, 4, 4, 4, 2, 0 | 0                      | 4            | 100%                   |

TABLE 16: Centroids of the physiological clusters of activity models.

| Cluster | BTemp | SBP | DBP | HB |
|---------|-------|-----|-----|----|
| 0       | 35.8  | 112 | 63  | 77 |
| 1       | 36.1  | 119 | 72  | 77 |
| 2       | 36.0  | 124 | 71  | 79 |
| 3       | 35.9  | 123 | 75  | 84 |
| 4       | 36.0  | 120 | 78  | 74 |

a suitable activity, and the information provided for the user is “suitable.” Data items 3, 4, and 5 are collected and classified into clusters 1, 0, and 3, respectively. The activity transaction is {3, 0, 1, 0, 3}, and only suitable rule 1 is selected. The activity can be evaluated as “suitable.” The last two activity items are classified into clusters 3 and 4, and no further rule is selected for this activity. The test transactions are depicted in Table 15; the third column denotes the prediction results obtained using the collected activity data. The fourth and fifth columns denote the applied rule and suitability confidence, respectively. Transactions 3 and 6 cannot be predicted because no activity pattern can be found in the obtained rules.

To validate the prediction results, certain physiological data items were selected and the centroids were classified into physiological clusters of PAMs; the centroids of these clusters are shown in Table 16. The physiological clusters of the test transactions are shown in Table 17, which also shows the environmental data classified into cluster 0. The suitability tag (column 6) for each transaction was based on the physiological cluster (column 4). The prediction results (column 3) were compared with the suitability tag (column 6), and it was observed that 80% of the activities could be predicted correctly and that all suitable activities could be

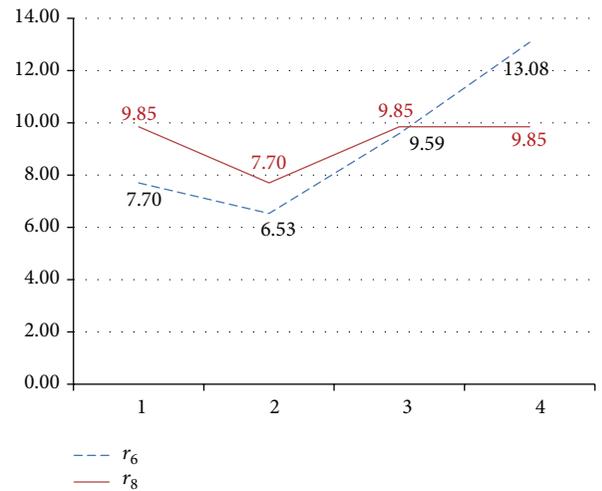


FIGURE 5: The visualization of rules with different activity clusters.

identified, such as transactions 1, 4, and 5. The remaining 20% activity transactions, that is, transactions 3 and 6, could not be predicted by using the obtained rules, because no obtained rules could be applied on transactions 3 and 6. However, new transactions can be added into the activity records to update the original models so that the updated models can be used to predict subsequent activity records. The obtained PAMs are used to evaluate the activities performed, and they provide evaluation results to users. The constructed PAMs can be updated using subsequent activity data.

Table 18 presents the results of test transactions which were obtained by the proposed approach on the dataset with 400 activity records to construct PAMs, and then the constructed PAMs were used to evaluate 138 activity records. All suitable activities were predicted correctly by PAMs.

TABLE 17: The predicting results obtained by using 10 test transactions.

| Tran. number | Transactions        | Predictions | Phy. cluster | Env. cluster | Suitability tag |
|--------------|---------------------|-------------|--------------|--------------|-----------------|
| 0            | 1, 4, 1, 2, 1, 0, 3 | <b>0</b>    | 1            | 0            | <b>0</b>        |
| 1            | 3, 0, 1, 0, 3, 3, 4 | <b>1</b>    | 2            | 0            | <b>1</b>        |
| 2            | 2, 4, 0, 3, 0, 2, 3 | <b>0</b>    | 1            | 0            | <b>0</b>        |
| 3            | 4, 3, 0, 4, 2, 2, 4 | —           | 1            | 0            | <b>0</b>        |
| 4            | 0, 1, 0, 4, 1, 2, 3 | <b>1</b>    | 2            | 0            | <b>1</b>        |
| 5            | 0, 2, 4, 0, 1, 2, 4 | <b>1</b>    | 2            | 0            | <b>1</b>        |
| 6            | 2, 2, 4, 4, 0, 3, 4 | —           | 1            | 0            | <b>0</b>        |
| 7            | 0, 4, 3, 2, 4, 0, 3 | <b>0</b>    | 1            | 0            | <b>0</b>        |
| 8            | 4, 3, 3, 3, 2, 1, 0 | <b>0</b>    | 1            | 0            | <b>0</b>        |
| 9            | 3, 1, 4, 4, 4, 2, 0 | <b>0</b>    | 1            | 0            | <b>0</b>        |

TABLE 18: The predicting results obtained by using 20 test transactions.

| Tran. number | Transactions        | Pred.    | Suit. tag |
|--------------|---------------------|----------|-----------|
| 1            | 0, 0, 4, 3, 3, 2, 0 | <b>1</b> | <b>1</b>  |
| 2            | 0, 0, 4, 2, 4, 1, 0 | <b>1</b> | <b>1</b>  |
| 3            | 2, 2, 3, 2, 0, 0, 2 | <b>1</b> | <b>1</b>  |
| 4            | 2, 3, 1, 2, 0, 0, 2 | <b>0</b> | <b>0</b>  |
| 5            | 3, 2, 0, 0, 0, 2, 3 | <b>0</b> | <b>0</b>  |
| 6            | 4, 1, 0, 0, 2, 3, 1 | <b>0</b> | <b>0</b>  |
| 7            | 0, 0, 0, 2, 4, 4, 3 | <b>0</b> | <b>0</b>  |
| 8            | 1, 1, 0, 0, 4, 4, 3 | <b>0</b> | <b>0</b>  |
| 9            | 1, 1, 0, 0, 0, 4, 3 | <b>0</b> | <b>0</b>  |
| 10           | 4, 1, 0, 0, 2, 3, 2 | <b>0</b> | <b>0</b>  |
| 11           | 0, 0, 0, 4, 2, 2, 2 | <b>0</b> | <b>0</b>  |
| 12           | 0, 0, 4, 4, 2, 1, 1 | <b>0</b> | <b>0</b>  |
| 13           | 0, 0, 2, 3, 1, 0, 0 | <b>0</b> | <b>0</b>  |
| 14           | 4, 2, 1, 1, 0, 0, 0 | <b>0</b> | <b>0</b>  |
| 15           | 3, 3, 1, 0, 0, 0, 1 | <b>0</b> | <b>0</b>  |
| 16           | 3, 2, 2, 0, 0, 0, 0 | <b>0</b> | <b>0</b>  |
| 17           | 4, 2, 3, 4, 0, 0, 0 | <b>0</b> | <b>0</b>  |
| 18           | 4, 2, 4, 4, 0, 4, 2 | <b>0</b> | <b>0</b>  |
| 19           | 4, 2, 3, 4, 0, 0, 0 | <b>0</b> | <b>0</b>  |
| 20           | 4, 3, 3, 2, 0       | <b>1</b> | <b>1</b>  |

## 5. Conclusion

This study proposes an analyzing approach for planning suitable physical activities. The approach uses techniques of association rule mining and clustering on a dataset of activity and physiological information to construct physical activity models (PAMs), which can be used to predict suitable physical activities for personal health and leisure management.

A potential limitation of this study is the prediction models are constructed using historic activity data and thus subsequent activities with unknown patterns cannot be predicted. Although collecting large amount of activity data may address this problem, new patterns still possibly occur in subsequent activities. Another limitation of this study is that the obtained physiological clusters require medical personnel to determinate the suitability tag to construct

prediction models. Further research should apply the knowledge transfer techniques on obtained rule sets to infer new rules for subsequent activities with new activity patterns. Furthermore, the similarity measure techniques should be applied to measure the similarity of obtained and new-build physiological clusters to reduce the efforts of tagging process.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

- [1] J. Beecham, T. Snell, M. Perkins, and M. Knapp, "Health and social care costs for young adults with epilepsy in the UK," *Health and Social Care in the Community*, vol. 18, no. 5, pp. 465–473, 2010.
- [2] J.-L. Fernández, T. Snell, and M. Knapp, "Social care services in England: policy evolution, current debates and market structure," *Cuadernos Aragoneses de Economía*, vol. 19, no. 2, pp. 265–282, 2010.
- [3] B. R. Schatz and R. B. Berlin Jr., *Healthcare Infrastructure: Health Systems for Individuals and Populations*, Health Informatics, Springer, London, UK, 2011.
- [4] J. Burrington-Brown, D. Claybrook, M. Dolan et al., "Defining the personal health information management role," *Journal of the American Health Information Management Association*, vol. 79, no. 6, pp. 59–63, 2008.
- [5] M. J. Ball, C. Smith, and R. S. Bakalar, "Personal health records: empowering consumers," *Journal of Healthcare Information Management*, vol. 21, no. 1, pp. 76–86, 2007.
- [6] U. M. Dholakia, R. P. Bagozzi, and L. K. Pearo, "A social influence model of consumer participation in network- and small-group-based virtual communities," *International Journal of Research in Marketing*, vol. 21, no. 3, pp. 241–263, 2004.
- [7] U. Varshney, "Pervasive healthcare and wireless health monitoring," *Mobile Networks and Applications*, vol. 12, no. 2-3, pp. 113–127, 2007.
- [8] S. Boudjit, N. Chelghoum, S. Allal, and M. Otsmani, "Multi-sensors' data gathering management system for a wireless health monitoring platform," in *Proceedings of the 1st ACM MobiHoc Workshop on Pervasive Wireless Healthcare (MobileHealth '11)*, Paris, France, 2011.

- [9] I. Qudah, P. Leijdekkers, and V. Gay, "Using mobile phones to improve medication compliance and awareness for cardiac patients," in *Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '10)*, ACM, Samos, Greece, June 2010.
- [10] P. Bauer, M. Sichitiu, R. Istepanian, and K. Premaratne, "The mobile patient: wireless distributed sensor networks for patient monitoring and care," in *Proceedings of the IEEE EMBS International Conference on Information Technology Applications in Biomedicine*, pp. 17–21, IEEE, Arlington, Va, USA, November 2000.
- [11] V. Stanford, "Using pervasive computing to deliver elder care," *IEEE Pervasive Computing*, vol. 1, no. 1, pp. 10–13, 2002.
- [12] M. C. Tremblay, A. R. Hevner, and D. J. Berndt, "Design of an information volatility measure for health care decision making," *Decision Support Systems*, vol. 52, no. 2, pp. 331–341, 2012.
- [13] I. Sim, G. D. Sanders, and K. M. MacDonald, "Evidence-based practice for mere mortals: the role of informatics and health services research," *Journal of General Internal Medicine*, vol. 17, no. 4, pp. 302–308, 2002.
- [14] I. Li, A. Dey, and J. Forlizzi, "A stage-based model of personal informatics systems," in *Proceedings of the 28th ACM Conference on Human Factors in Computing Systems (CHI '10)*, pp. 557–566, ACM, Atlanta, Ga, USA, April 2010.
- [15] M. Zerkouk, P. Cavalcante, A. Mhamed, J. Boudy, and B. Messabih, "Behavior and capability based access control model for personalized telehealthcare assistance," *Mobile Networks and Applications*, vol. 19, no. 3, pp. 392–403, 2014.
- [16] M. A. Saleem, Y.-K. Lee, and S. Lee, "Trajectory patterns mining towards lifecare provisioning," *Wireless Personal Communications*, vol. 76, no. 4, pp. 747–762, 2014.
- [17] M. R. Scott and H. Smith, *Applied Logistic Regression Analysis*, Sage, London, UK, 2nd edition, 2002.
- [18] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability and Statistics for Engineers and Scientists*, Prentice-Hall, Upper Saddle River, NJ, USA, 2002.
- [19] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, 3rd edition, 2011.
- [20] C.-P. Chang and C.-P. Chu, "Defect prevention in software processes: an action-based approach," *Journal of Systems and Software*, vol. 80, no. 4, pp. 559–570, 2007.
- [21] C.-P. Chang, C.-P. Chu, and Y.-F. Yeh, "Integrating in-process software defect prediction with association mining to discover defect pattern," *Information and Software Technology*, vol. 51, no. 2, pp. 375–384, 2009.
- [22] S. Zhong, T. M. Khoshgoftaar, and N. Seliya, "Analyzing software measurement data with clustering techniques," *IEEE Intelligent Systems*, vol. 19, no. 2, pp. 20–27, 2004.
- [23] A. Ahmad and L. Dey, "A  $k$ -mean clustering algorithm for mixed numeric and categorical data," *Data and Knowledge Engineering*, vol. 63, no. 2, pp. 503–527, 2007.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

