Journal of Information Processing Vol.24 No.2 237–246 (Mar. 2016)

[DOI: 10.2197/ipsjjip.24.237]

Regular Paper

Estimating Physical Characteristics with Body-worn Accelerometers Based on Activity Similarities

Akira Masuda[1,a)] Takuya Maekawa[1,b)]

Received: May 4, 2015, Accepted: November 5, 2015

Abstract: This paper describes our experimental investigation of the end user physical characteristics (e.g., gender, height, weight, dominant hand, and skill at sport) that can be successfully estimated solely from sensor data obtained during daily activities (e.g., walking and dish washing) from body-worn accelerometers. For this purpose we use the huge quantities of data that we have collected, which include 14,880 labeled activities obtained from 61 subjects. Our proposed method tries to estimate various kinds of characteristics based on our simple idea 'When the activity sensor data of two users are similar, the physical characteristics of the two users may also be similar.' We consider that estimating the end user's physical characteristics will enable us to realize new kinds of applications that automatically recommend information/services to an end user according to her estimated physical characteristics such as gender and weight.

Keywords: activity sensing, accelerometers, physical characteristics

1. Introduction

Advances in sensing and wireless communication technologies have led to the low cost production of small wireless body-worn and environment-embedded sensor devices that are used to recognize human daily activities. In the near future, end users will wear small accelerometers and/or accelerometer-embedded devices such as wristwatches, cellphones, and shoes [16], and their daily lives will be continuously recorded by these sensors.

In this paper, we try to estimate information about an end user's physical characteristics such as height, weight, gender, dominant hand, and age solely by employing acceleration data obtained from accelerometers attached to several parts of her body and/or accelerometers embedded in daily objects. We believe that the automatic estimation of physical characteristics with acceleration sensors will prove very useful in many application domains. We show some examples. (1) We can construct a personalized recommender system that provides user-specific advertisement/information according to the user's estimated characteristics. For example, if an accelerometer is attached to a shopping basket and physical characteristics of a shopper who carry around the basket are estimated, a public display or humanoid robot in the shop can recommend items for him/her according to the estimated characteristics such as gender and age. Such information is also useful for analyzing customer shopping behaviors according to their gender and age. Furthermore, recent amusement parks have introduced RFID sensor bracelets worn by customers for easy payment. If accelerometers are embedded in such sensor bracelets and physical characteristics of the customers are estimated, physical characteristics aware attractions (e.g., according to gender and age) and recommender systems at souvenir shops and restaurants will be achieved. (2) If an end user's estimated physical characteristics are useful information for the end user herself, we can simply provide her with the information. For example, if the end user is taking tennis lessons, she will definitely find it useful to receive her monthly progress reports. For example, monthly progress in the estimated tennis skillfulness of the end user is definitely useful. (3) While the evaluation described below mainly focuses on known (apparent) physical characteristics such as height, dominant hand, and sports experience, we consider our method applicable to the automatic estimation of hidden physical characteristics such as diseases, health indicators, and sporting ability. In particular, the automatic diagnosis of diseases that affect body movements by using always-on accelerometers could be a significant application of wearable sensor systems. In fact, several physiology studies have determined the difference between normal subjects' gait acceleration signals and those of subjects with a disease, e.g., between healthy subjects and subjects with diabetes [17]. (4) If such physical characteristics as age, gender, and height of a person are estimated, these characteristics can be used as soft biometric features to identify the person. Assume that accelerometers are embedded into such daily objects as house shoes in a home environment, and basic information of house residents such as name, age, gender, and height are known. By using the estimated physical characteristics and the known physical characteristics, we can identify a house resident and provide personalized services to the resident.

We estimate an end user's physical characteristics by employing machine learning approaches. That is, we prepare labeled time-series acceleration data obtained from many other users

1 Graduate School of Information Science and Technology, Osaka University, Suita, Osaka 565–0871, Japan

a) masuda.akira@ist.osaka-u.ac.jp

b) maekawa@ist.osaka-u.ac.jp

(*training users*) and information about their physical characteristics, and construct models that estimate physical characteristics in advance. (A label attached to acceleration data includes information about the class label of its related activity and activity start and end times.) We then estimate the end user's physical characteristics by using the models and the end user's acceleration sensor data. We construct a model that estimates each kind of physical characteristic, for example height or gender. The model is based on our idea that when the activity sensor data of two users are similar, the physical characteristics of the two users may also be similar. That is, by computing similarities between activities of training users, we learn the relationship between their activity similarities and the similarities of their physical characteristics. For example, 'write' activity sensor data obtained from two left-handed users may be similar. When we know in advance that one of the two users is left-handed, we can estimate the dominant hand of the other user by using similarities between the sensor data from their 'write' activities.

In the rest of this paper, we first introduce work related to acceleration data analysis. Then we explain our proposed method and evaluate it with acceleration sensor data obtained from 61 subjects. The contribution of this paper is that we experimentally investigate various kinds of physical characteristics that can be successfully estimated from acceleration data obtained from body-worn sensors by using 147.5 hours of labeled sensor data obtained from 61 subjects. To the best of our knowledge, our method is the first method that can estimate various physical characteristics from various activity sensor data based on activity similarity.

2. Related Work

We introduce several studies that estimate the properties of an accelerometer (or its wearer) by analyzing its acceleration data. Reference [11] estimates the on-body position of an accelerometer by using its sensor data. Reference [11] prepares sensor data obtained from accelerometers worn on various parts of the body in advance, and learns the characteristics of the acceleration data that are peculiar to each on-body position by using discriminative classifiers such as the Support Vector Machine (SVM) and the C4.5 decision tree. Many studies estimate the calories expended in such physical activities as walking and running by using body-worn accelerometers (and other sensors) [2], [15]. Several studies employ 'walk' acceleration data to achieve biometric gait authentication [5], [14]. Also, Ref. [8] tried to detect drunk walking by using an accelerometer based on an assumption that the effects of alcohol intake on gait data are similar for each user. The studies that come closest to ours involve acceleration-based evaluation systems related to sport training and health care that evaluate skillfulness in sports and several health indicators. For example, Ref. [6] evaluates a golf swing by using the estimated angle of wrist rotation with accelerometers and gyroscopes attached to the body and a golf club. Reference [21] distinguishes experienced and inexperienced runners with body-worn accelerometers. Reference [18] attempts to estimate activity levels of subjects from the length of 'walk' activities in subjects' daily lives. By contrast, we investigate various kinds of physical characteristics such as height, dominant hand, and sport experience in a simple and unified framework by using large amounts of data. Also, Ref. [23] tries to estimate basic physical characteristics (gender, weight, and height) by using an accelerometer on a smartphone. They directly estimate physical characteristics from only walk sensor data features. On the other hand, our approach can estimate physical characteristics from various activity data. (This method will be compared with our method in the evaluation section.)

In the medical research field, many surveillance studies have been undertaken concerning the relationship between physical activity levels obtained from accelerometers attached to subjects and the physical characteristics of the subjects [7], [22]. For example, Ref. [22] investigates age and gender differences in relation to physical activity levels in the United States. Reference [7] examines gender, day, and time of day differences in the physical activity levels of adolescents. Also, as mentioned above, Ref. [17] investigates the difference between the gait acceleration signals of subjects with and without diabetes. In contrast, we try to estimate various kinds of end user physical characteristics by using similarities between activity data.

Our proposed method is similar to collaborative filtering methods in the information retrieval research field [20] because the collaborative filtering methods recommend an item to a user based on the computed similarity between users. Since the collaborative filtering methods compute the similarity based on sets of items that the users have purchased, the similarity is computed by using the correlation between the purchased items. In contrast, this study focuses on acceleration sensor data, i.e., continuous values, and we compute the user similarity based on distributions of sensor data.

3. Estimation Method

3.1 Assumed Environment

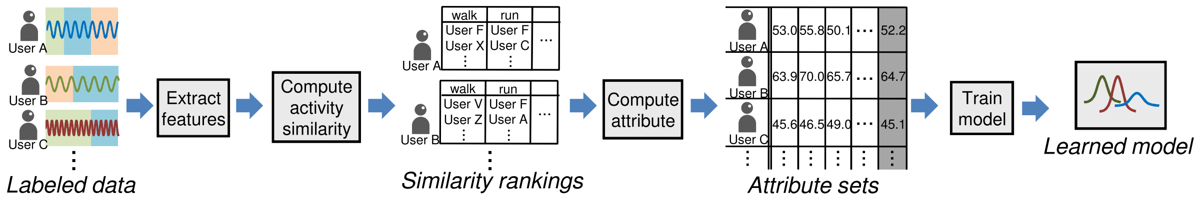
We assume that an end user wears several accelerometers (or has devices with accelerometers). Sensor data obtained from the accelerometers are analyzed by an activity recognition system. That is, the recognition system labels the sensor data. A label includes information about the class label of its related activity and activity start and end times. We estimate the end user's physical characteristics by using the labeled sensor data.

3.2 Outline of Method

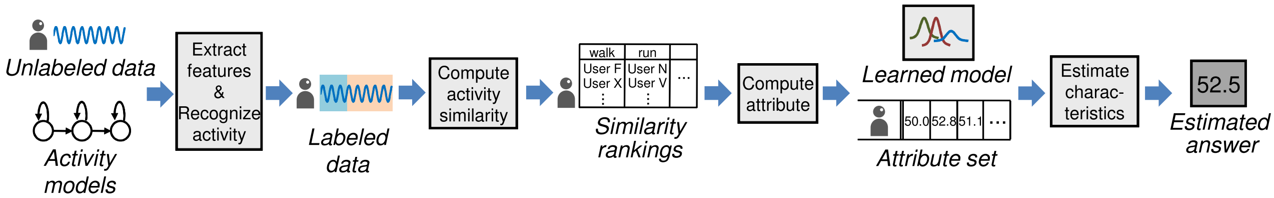
Our method consists of main two procedures; *Training* and *Physical characteristic estimation* as summarized in **Fig. 1**. In the *Training* procedure, we construct a model for each of the physical characteristics we want to learn/estimate in advance. A model estimates a given physical characteristic of an end user. In the *Physical characteristic estimation* procedure, we estimate a given physical characteristic of an end user by using her acceleration data labeled by the activity recognition system. We detail the two procedures below.

3.3 Training

We obtain labeled acceleration data including various kinds of activities from many training users in advance. Figure 1 (a) shows the outline of this procedure. In this procedure, (1) we extract fea-



(a) Training: Learn a model for estimating physical characteristics.



(b) Physical characteristic estimation: Estimate physical characteristics from unlabeled acceleration data.

Fig. 1 Overview of our method. (a) Training and (b) Physical characteristic estimation.

tures from the sensor data, (2) we compute similarities between the activities of training users according to each activity class by using the extracted features and construct rankings of each training user that include other training users in ascending order of similarity to the training user, (3) we compute attribute sets from the rankings to construct a model that estimates a certain kind of physical characteristic, e.g., height or age, and (4) we then train the model by using the computed attributes and answers. An answer corresponds to the training user's physical characteristic value that we want to estimate, e.g., height or age. We compute the attributes from the physical characteristics of other similar training users based on our idea that when the activity sensor data of two users are similar, the physical characteristics of the two users may also be similar. When we want to estimate a height, for example, we use the heights of similar training users to compute the attributes. We detail the four sub-procedures.

3.3.1 Feature Extraction

Before computing activity similarities, we extract features from the training users' sensor data based on existing activity recognition studies. Because we assume time-series acceleration data, we compute a feature vector for each sliding time window. We extract features based on the FFT components of 64 sample time windows. As features, we use the mean, energy, and dominant frequency according to the existing activity recognition studies [1], [24]. The mean is the DC component of the FFT coefficients, and can characterize the posture of parts of the body. For example, a mean corresponding to a hand posture during tooth brushing may have particular characteristics. The energy feature is calculated by summing the magnitudes of the squared discrete FFT components. Note that the DC component of the FFT coefficients is excluded from this summation. The energy can be used to distinguish low intensity activities such as standing from high intensity activities such as walking [1], [24]. The dominant frequency is the frequency that has the largest FFT component, and it allows us to distinguish between repetitive motions with similar energy values [13]. We construct a feature vector that concatenates the above features extracted from the body-worn accelerometers for each time window (time slice).

3.3.2 Computing Similarities

Because the training users' sensor data are labeled, we can obtain one or more feature vector sequences for each activity class such as 'walk' and 'run' for each training user. That is, we can compute an activity similarity according to a certain activity class exhibited by two training users by using their feature vector sequences.

We first construct a feature vector that concatenates the features extracted from sensor data obtained from a training user's accelerometers for each time slice. Then, we compute the activity similarity of each pair of training users by using their feature vector sequences that correspond to the activity class. We compute activity similarities between training users simply by using a Gaussian mixture model (GMM). Assume that we wish to compute an activity similarity between the 'walk' activities of training users A and B. We regard user A as a base user and user B as an object user, and compute the similarity between the object user's 'walk' activity feature vectors and the base user's 'walk' activity model by using

$$p(\boldsymbol{f_o}|\boldsymbol{\lambda_b}) = \sum_{i=0}^{M} \pi_i N(\boldsymbol{f_o}|\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}), \tag{1}$$

where $\boldsymbol{f_o}$ is a feature vector of the object user, $\boldsymbol{\lambda_b}$ collectively shows all the parameters of the base user's GMM, M is a number of mixtures, which is 32 in our implementation. π_i, $\boldsymbol{\mu_i}$ and $\boldsymbol{\Sigma_i}$ are the mixture weight, mean vector and covariance matrix of the ith multivariate Gaussian distribution of the GMM, respectively. We employ the EM algorithm to estimate the GMM parameters [3]. We use a GMM to compute the activity similarity because a GMM is usually used to model a complex data distribution. For example, 'wash dishes' sensor data include data corresponding to 'rub with a sponge' and 'flush with water.' Therefore, we use a GMM to capture such a complex distribution of activity sensor data. The similarity D of the 'walk' activity between the base user and the object user is simply computed as

$$D = \frac{1}{2}\left(\frac{1}{J_o}\sum_{j=0}^{J_o} p(\boldsymbol{f_{oj}}|\boldsymbol{\lambda_b}) + \frac{1}{J_b}\sum_{j=0}^{J_b} p(\boldsymbol{f_{bj}}|\boldsymbol{\lambda_o})\right), \tag{2}$$

where J_o and J_b are the number of the object and base user's 'walk' feature vectors, respectively. We compute the similarity for each pair of training users. By doing so, we can obtain the activity similarity between the 'walk' activities of a training user

	walk									···	bicycle									height
	hand (R)				···	all four				···	hand (R)				···	all four				
	top-1	top-3	···	top-9	···	top-1	top-3	···	top-9	···	top-1	top-3	···	top-9	···	top-1	top-3	···	top-9	
User A	150.1	154.7	···	152.0	···	150.1	162.2	···	162.3	···	172.0	170.9	···	165.1	···	175.5	160.7	···	163.8	160.8
User B	180.0	178.5	···	175.5	···	170.1	171.9	···	176.7	···	172.0	169.8	···	168.5	···	160.8	162.0	···	162.1	177.9
⋮					⋮										⋮					⋮
User Z	153.0	151.9	···	154.0	···	153.0	153.7	···	151.1	···	172.0	171.4	···	165.1	···	175.5	172.7	···	170.1	155.0
	attributes																			answer

Fig. 3 Example attribute sets and answers for estimating height (cm). A pair consisting of an attribute set and an answer is computed for each training user.

	walk				run				···
	hand (R)	hand (L)	···	all four	hand (R)	hand (L)	···	all four	···
User A	User D User N User Z User L ⋮	User B User X User L User K ⋮	···	User D User L User J User Y ⋮	User C User K User L User H ⋮	User B User N User D User Q ⋮	···	User D User L User Z User B ⋮	···

Fig. 2 Example activity similarity rankings of user A. Each ranking includes training users in ascending order of similarity to user A. A ranking is computed for each activity class and each sensor variation.

(base user) A and those of each other training user. That is, we can rank training users from the computed similarities. This ranking reflects the similarities of 'walk' activities to those of user A. In our implementation, we define the following six sensor variations; (1) the right hand, (2) the left hand, (3) the waist, (4) the thigh, (5) both hands, and (6) both hands, waist and thigh, and compute the rankings for each sensor variation. (In the experiment described below, subjects wear four three-axis accelerometers; one on each wrist, one on the right thigh, and one at the waist.) With the fifth variation, for example, we construct a feature vector that concatenates features extracted from accelerometers on the right and left hands, and compute the rankings with GMMs. The leftmost column in **Fig. 2** shows an example of the ranking of 'walk' activities when we focus on the right hand. In the example, users D, N, Z, and L are similar to user A in this order with D being the most similar. We can compute an activity similarity ranking for each sensor variation and for each activity class as shown in Fig. 2. As regard to training user A, when we use six sensor variations and assume fourteen activity classes, we generate 84 rankings (6×14). We compute an attribute set by using the rankings to learn a model that estimates a certain kind of physical characteristic.

3.3.3 Computing Attributes and Learning Model

We construct a model that estimates a certain kind of physical characteristic by using the above rankings and the physical characteristics of training users. We assume two types of physical characteristic information; numerical and nominal information. Numerical information includes a user's height, weight, and age. Nominal information includes a user's gender, dominant hand, and sport experience. For example, gender information has 'male' and 'female' values, and sport experience information has 'yes,' 'somewhat,' and 'no' values.

Here, assume that we construct a model that estimates an end user's height. We focus on a training user A and compute a pair consisting of an attribute set and an answer as shown in the first row (User A's row) in **Fig. 3**. The answer corresponds to the height of user A. We compute the attribute set by using the heights of similar users in the activity similarity rankings shown in Fig. 2. This is based on our idea that when the activity sensor data of two users are similar, the physical characteristics of the two users may also be similar. In each ranking, we compute attribute values by using the physical characteristics of the top-n similar training users. When we deal with numerical physical characteristics such as height and age, we employ the average value for the top-n similar users as an attribute. For example, the first attribute value of user A corresponds to the height of the most similar user as regard to the right hand in the 'walk' activity class, i.e., user D as shown in the leftmost column of Fig. 2. The second attribute value of user A corresponds to the average height for the top-3 similar users, i.e., users D, N, and Z as listed in the leftmost column of Fig. 2. In our implementation, each attribute value are computed from its corresponding activity similarity described as

$$A_{cn} = \frac{1}{n} \sum_{i=1}^{n} a_c(u^i) \tag{3}$$

where $n = 1, 3, 5, 7, 9$, u^i shows the ith ranked user in the activity similarity ranking, and $a_c(u)$ shows the value of physical characteristics c (e.g., height).

When we deal with nominal physical characteristics such as gender and dominant hand, we employ a simple majority voting protocol. Assume that we focus on gender. When n is 3 and the top-3 similar users consist of two males and one female, for example, we use 'male' as an attribute value. We use five n values (1, 3, 5, 7 and 9) same as numerical characteristics, and compute an attribute for each n value. As above, we can compute attributes from any kinds of physical characteristics information by using activity similarity rankings.

Here, it is unclear which kind of attribute is useful for estimating a certain kind of physical characteristic, e.g., height and gender. Therefore, we learn a model that estimates the physical characteristic by using large numbers of pairs consisting of an attribute set and an answer with machine learning approaches. When we want to estimate nominal physical characteristics such as gender and dominant hand, we use such classifiers as the Naive Bayes classifier and the SVM. When we want to estimate numerical physical characteristics such as height and age, we use such models as regression models and neural networks.

3.4 Physical Characteristic Estimation

Unlabeled acceleration data obtained from an end user are given. Figure 1 (b) shows the outline of this procedure. In this procedure, (1) we extract features from the sensor data, (2) an activity recognition system labels the feature vectors, (3) we compute activity similarities between the end user and each training user according to each activity class and each sensor variation by using the labeled data, and construct rankings of the end user that include training users in ascending order of similarity to the end user, (4) we compute an attribute set from the rankings, and (5) we then estimate a certain kind of physical characteristic by using its corresponding model learned in the *Training* procedure and the attribute set. The first, third, and fourth sub-procedures are identical to those in the *Training* procedure. In the fifth sub-procedure, by using the attribute set, the model learned in the *Training* procedure estimates the value of the end user's physical characteristics. We explain the second sub-procedure below.

3.4.1 Activity Recognition

We classify an extracted feature vector at each time slice in an appropriate activity class by employing supervised machine learning techniques. That is, we first model each activity by using labeled training data, and then recognize test data with the learned models. We learn an activity class with a left-to-right hidden Markov model (HMM) where the values of its observed variables correspond to extracted feature vectors, and we represent its output distributions by using Gaussian mixture densities. We learn an activity model for each activity class that we want to recognize.

4. Experiment

4.1 Dataset

We collected sensor data with our developed wireless sensor nodes equipped with three-axis acceleration sensors and sampling rates of 30 Hz. Each subject wore the sensor nodes on the wrists of both hands, the waist, and the right thigh. Here, the most natural data would be acquired from the normal daily lives of the subjects. However, obtaining sufficient samples of such data from many subjects is very costly. Therefore, we collected sensor data by using a semi-naturalistic collection protocol [1] that permits greater variability in subject behavior than laboratory data. In the protocol, the subjects perform a random sequence of activities following instructions on a worksheet. The subjects are relatively free about how they perform each activity because the instructions on the worksheet are not very strict, e.g., "brush your teeth at the sink" and "vacuum the room with a hand-held vacuum cleaner." During the experimental period, the subjects completed data collection sessions (10 minutes average) that included a random sequence of the activities listed in **Table 1**. Most of the activities were basically selected from those reported in existing activity recognition studies [1], [12], [13], [19]. Also, we collected 'draw on whiteboard' and 'write in notebook' data because we wanted to try to estimate subjects' writing dominant hand. We collected 'play pingpong' data because we try to estimate a user's sport experience. Each subject completed ten sessions in total in our experimental environment.

To annotate the collected data, a companion recorded the subjects with a web camera during the experiment. The web camera was connected to a mobile computer carried by the companion. The sensor data from the four sensor nodes attached to the subject were also sent to the mobile computer. We describe how several of the activities in Table 1 were performed in detail. In activity J, we instructed the subjects to enter several sentences on the computer keyboard. In activities K and L, we instructed the subjects to write some sentences in a notebook and on a whiteboard, respectively. In activity M, each subject played pingpong with a worker in our laboratory. We collected a total of 14,880 labeled activities from the 61 subjects. Note that in our experiment, several activities were performed at different places, e.g., 'wash dishes' at kitchen, 'ascend stairs' at stairs, and 'play pingpong' at gym. Therefore, when a participant walked between the places, 'walk' activities were collected, and thus multiple 'walk' labels were included in a session.

Each subject also filled out a questionnaire that asked for information about the physical characteristics listed in **Table 2**. We selected various kinds of physical characteristics related to the subjects ranging from basic information such as weight and gender to information about the activities listed in Table 1. The activity-related information included the dominant hand used in several activities. **Figure 4** shows the distribution of the physical characteristics of our subjects. As shown in Fig. 4, the subjects were

Table 1 Activities performed in our experiment.

A	stand	F	descend stairs	K	draw on whiteboard
B	walk	G	bicycle	L	write in notebook
C	run	H	brush teeth	M	play pingpong
D	sit	I	wash dishes	N	vacuum
E	ascend stairs	J	use PC		

Table 2 Physical characteristic information used in our experiment.

name	value
gender	{male, female}
age	numerical
height	numerical
weight	numerical
dominant hand (writing)	{right, left}
dominant hand (brushing)	{right, left, both}
dominant hand (pingpong)	{right, left}
pingpong experience	{yes(2), somewhat(1), no(0)}

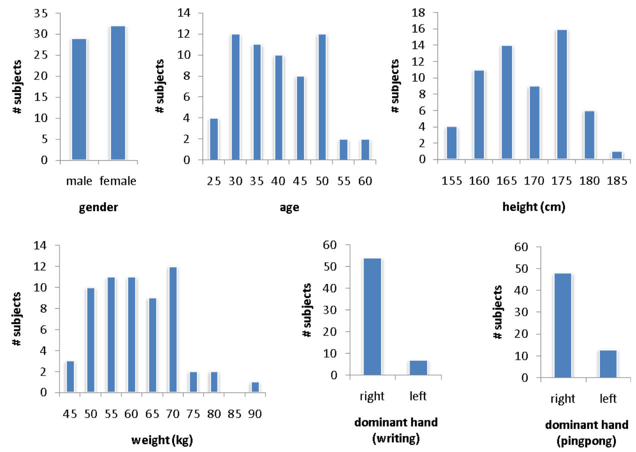


Fig. 4 Distribution of physical characteristics of our subjects (selected).

well-balanced in terms of gender. The subjects ranged from 22 to 58 years old. Also, the height and weight of the subjects covered a wide range. There were few left-handed subjects because there are few left-handed people.

4.2 Evaluation Methodology

We evaluated our method using 'leave-one-subject-out' cross validation. That is, we regarded one subject as an end user and the remaining subjects as training users. We iterated the procedure so that each subject became an end user once, and we computed the estimation accuracies of the physical characteristics listed in Table 2. Note that when we trained activity models that labeled the unlabeled acceleration data of an end user, we used the labeled acceleration data of all the training users as the training data. That is, we can recognize the activities and estimate the physical characteristics of an end user without needing her labeled acceleration data.

As mentioned above, when we estimate nominal physical characteristics such as 'gender' and 'dominant hand,' we use discriminative classifiers. In the evaluation, we tested the C4.5 decision tree, SVM, Naive Bayes classifier (NB), and multinominal logistic regression (Logi) implemented in Ref. [25]. When we estimated such numerical physical characteristics as height and age, we tested the sequential minimal optimization algorithm for regression (SMOreg), linear regression (Linear), and least median squared linear regression (LMS). We use parameters and hyperparameters of the machine learning methods that yielded good performance in our preliminary test using small data. Also, as for SVM and SMOreg, we use a polynomial kernel with the SMO algorithm.

Note that 'pingpong experience' has ordered values. That is, the level of experience decreases in the order; 'yes,' 'somewhat,' and 'No.' We also regard the information as numerical information. We consider that each nominal values of the information shown in Table 2 correspond to numerical values associated with the nominal values in the table.

In addition, we prepare naive methods for estimating physical characteristics and then compare them with our methods. A naive method for estimating nominal characteristics simply outputs the major nominal value among all training users. For example, when we want to estimate 'gender,' and the respective numbers of male and female training users are 10 and 20, the method outputs 'female.' A naive method for estimating a numerical characteristic simply outputs the average value for all the training users.

Furthermore, we prepare baseline methods that are designed based on Ref. [23]. Here we briefly explain a method proposed in Ref. [23]. The method also employs training users' sensor data and their physical characteristics information. The method first extracts feature vectors obtained from an end user's acceleration sensor data in the same way as our method. Then the method finds a feature vector from a training user that is most similar to each feature vector from the end user (i.e., 1-NN search). That is, a similar training user is retrieved for each feature vector of the end user. Finally, the method computes the average physical characteristics value for the retrieved similar training users (e.g., average height). The average value will be the estimation of the method. (When we estimate nominal physical characteristics, a simple majority voting protocol is employed.) We call the baseline method *base1*. We also prepare an activity-aware baseline method based on Ref. [23]. The method first identifies an activity class that each feature vector belongs to. Then, the method performs an 1-NN search for each activity class. Note that the 1-NN searcher is prepared for each activity class. That is, the method can retrieve a similar user for each feature vector by taking account of its activity class. However, the method cannot learn which kind of activity class is useful for estimating a certain kind of physical characteristic because the method outputs the average physical characteristics value for the retrieved similar training users. We call the baseline method *base2*. On the other hand, our method can learn which kind of activity class is useful for estimating a certain kind of physical characteristics because we train a classifier by using features prepared for each activity class. (The classifier learns the importance of a feature by using training data.)

Table 3 Accuracies (percentages) of nominal physical characteristics.

	gender	dominant hand (writing)	dominant hand (pingpong)	dominant hand (brushing)	pingpong expe-rience
naive	52.5	88.5	78.7	73.8	**44.3**
base1	62.3	88.5	72.8	75.4	39.3
base2	61.6	89.3	78.1	71.2	39.1
C4.5	59.0	**98.4**	93.4	**91.8**	26.2
SVM	80.3	**98.4**	**96.7**	90.2	39.3
NB	**86.9**	**98.4**	**96.7**	90.2	39.3
Logi	77.0	95.1	93.4	83.6	34.4
NB(100)	91.8	96.7	95.1	88.5	49.2

4.3 Results

4.3.1 Estimation Accuracy for Nominal Information

Table 3 tabulates the classification accuracies of nominal physical characteristics for each method. The accuracy means the percentage of correctly classified instances (end users).

[Gender] We first focus on 'gender' results. We had considered it very difficult to estimate a subject's gender using only accelerometers. Contrary to our expectation, the accuracies of SVM, NB, and Logi methods were good and much higher than the accuracies of the naive and baseline methods. The accuracy of NB was approximately 87%. **Figure 5** shows the accuracies for each physical characteristic obtained with the NB method, which provided the best results in three of the five categories. As shown in Fig. 5 (a), the accuracy for female subjects was better than that for male subjects. This may be because there were fewer male subjects. Here, we investigate which extracted feature contributed to the estimation of the subject's gender. We evaluated the worth (contributions) of attributes by using the ReliefF evaluator, which can deal with both numerical and nominal values. ReliefF evaluates an attribute by using a sampled attribute set and its k-nearest neighbors from both the same and different classes. For more detail, see Ref. [10]. With the ReliefF evaluator, we found that the features computed from the 'walk,' 'run' 'vacuum,' and 'play pingpong' activity similarities contributed greatly to estimating the subject's gender. With 'walk,' the activity similarities computed from all of the four sensors were important contributors. This may be caused by gender differences in a physique. With

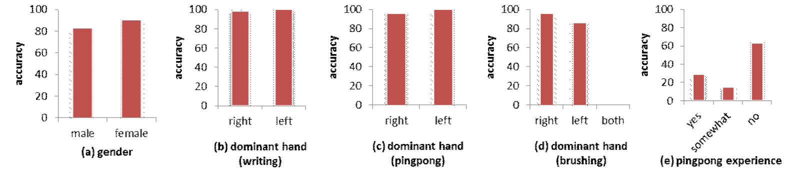


Fig. 5 Accuracies for subjects for each physical characteristic value (NB method).

Table 4 Mean absolute errors (MAEs) of numerical physical characteristics.

	age	height (cm)	weight (kg)	pingpong experience
naive	7.95	6.56	8.01	0.71
base1	7.95	6.00	8.00	0.71
base2	7.95	6.27	7.64	0.73
SMOreg	7.44	**4.30**	**4.50**	**0.61**
Linear	**7.13**	4.62	4.81	**0.61**
LMS	7.41	4.63	4.73	0.63
SMOreg(100)	7.28	4.01	4.34	0.61

the 'run' activity, we also found a gender difference regarding the thigh acceleration data. The average dominant frequencies of the male and female subjects in the x-axis data were 2.54 and 3.04 Hz, respectively. We confirmed a significant difference between the two average frequencies with a two-tail t-test ($p < .05$). With 'vacuum,' we found the gender differences in using a vacuum cleaner. Many female subjects vacuumed with a stoop and held the nozzle of the canister vacuum with both hands (76.2%). On the other hand, many male subjects vacuumed standing upright and held the nozzle with one hand (84.0%). As regard to 'play pingpong,' the ratio of female subjects with no pingpong experience (53.1%) was higher than that for male subjects (34.5%). Men are reported to participate in sport more frequently than women [9].

[Dominant hand] With 'dominant hand (writing),' the accuracies of the C4.5, SVM, and NB methods were high, and better than those of the naive and baseline methods as shown in Table 3. As shown in Fig. 5 (b), the results with the NB method were better for both right- and left-handed subjects. With ReliefF, we confirmed that the features computed from the 'draw on whiteboard' and 'write on notebook' activity similarities contributed greatly to estimating a subject's dominant hand when writing. As regard to 'dominant hand (pingpong),' the accuracies of the SVM and NB methods were good and much higher than the accuracy of the naive method. As shown in Fig. 5 (c), the accuracies for both right- and left-handed subjects were high. The features computed from the 'play pingpong' and 'brush teeth' activity similarities contributed greatly to estimating a subject's dominant hand in pingpong. With 90.2% of subjects, the dominant hand for pingpong and brushing teeth were the same. With 'dominant hand (brushing),' the accuracy of the C4.5 method was good as shown in Table 3. However, the accuracy for both-handed (ambidextrous) subjects was zero in Fig. 5 (d). This is because there were only two ambidextrous subjects. Note that an ambidextrous subject means a subject who switches the toothbrush from one hand to the other while brushing her teeth.

[Sports experience] With 'pingpong experience,' the accuracy of the naive method outperformed our methods. With ReliefF, the features computed from the 'vacuum' and 'run' activity similarities were listed as high contributors. This means that our approach could not successfully capture characteristics of body movements related to 'pingpong experience.' However, when we used training data with manual activity labels, the accuracy of NB increased to 54.1%. That is, we consider that the activity recognition performance decreased the accuracy. (The F-measure of activity recognition was 0.855, and that of 'play pingpong' was 0.743.)

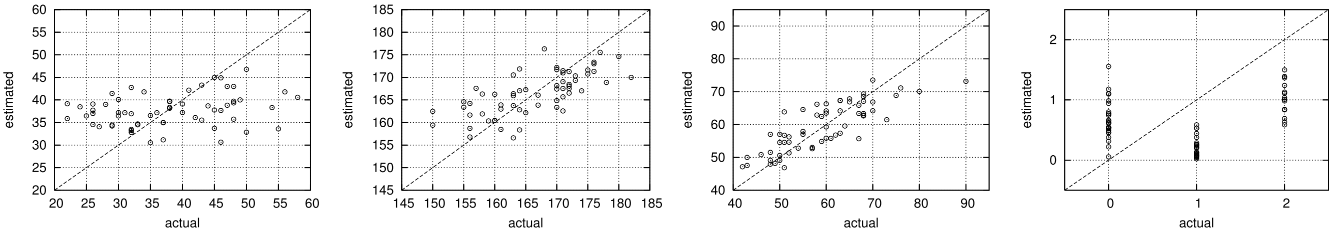
In the above classification results, NB outperformed the other methods as shown in Table 3. This may be because NB deals well with highly dimensional data and does not require huge amounts of training data.

4.3.2 Estimation Accuracy for Numerical Information

Table 4 shows mean absolute errors (MAEs) for numerical physical characteristics for each method. A smaller MAE indicates a higher estimation accuracy.

[Basic physical characteristics (age, height, and weight)] We first focus on 'age' results. Linear achieved the lowest MAE, namely an error of about seven years. However, we found no significant difference between the MAE of SMOreg and that of the naive method with a two-tail t-test ($p > .05$). The scatter chart in **Fig. 6** (a) shows the relationship between an actual physical characteristic value and the value estimated for each subject. The x-axis shows an actual physical characteristic value and the y-axis shows an estimated value. That is, points closer to the dashed line in the chart have better corresponding estimated results. As shown in the chart, the estimated results for subjects around 25 years old were poor. We consider it to be difficult to estimate a subject's age with high accuracy using only her activity similarities. Note that our 'age' data do not have a decimal part, and thus this limitation can have a negative effect on the estimation results. However, because the estimation error is larger than seven, we believe that the effect of the limitation of our data is small.

With 'height,' the MAEs of SMOreg, Linear, and LMS were very small (approximately 4.5 cm), and much smaller than the MAEs of the naive and baseline methods as shown in Table 4. In fact, we could find significant differences between the MAEs of these methods and the MAEs of the naive and baseline methods using a two-tail t-test ($p < .05$). It is surprising that we could estimate a subject's height with an average error of approximately 4.5 cm by using an acceleration-based approach. Figure 6 (b) is a scatter chart showing the SMOreg estimation results. While the results for 'height' seem to be better than those for 'age,' the results for subjects around 155 cm tall were poor. By using the ReliefF evaluator, we found that the features computed from the 'draw on whiteboard' and 'walk' activity similarities contributed considerably to estimating a subject's height. The relationship between the 'walk' activity and the subject's height is convincing because the lengths of the arms and legs are relatively proportional to the height, and so the hand acceleration data obtained from subjects with similar arm lengths may also be similar. In fact, the features computed from accelerometers on the hand and



(a) age – Linear (b) height (cm) – SMOreg (c) weight (kg) – SMOreg (d) pingpong exp. – SMOreg

Fig. 6 Scatter charts showing relationships between actual numerical physical characteristic values and estimated values.

thigh were important contributors. However, it is hard to understand why the 'draw on whiteboard' activity contributed to estimating the subjects' heights. In particular, the features computed from accelerometers on the hands contributed significantly. We consider that the tall subjects wrote on a whiteboard with a stoop, and so the postures of their hands may be different from those of the short subjects.

As regard to 'weight,' the MAE of SMOreg was the lowest and much smaller than that of the naive method as shown in Table 4. We found that there was significant difference between the MAE of SMOreg and the MAEs of the naive and baseline methods using a two-tail t-test ($p < .05$). Figure 6 (c) is a scatter chart showing the estimation results obtained with SMOreg. These results appear to be good. The features computed from the 'vacuum' and 'walk' activity similarities were important contributors. In particular, the features computed from accelerometers on the hand were important, which were also important in estimating 'height.' This may be because 'weight' and 'height' have a strong correlation. In the 'run' activity, the features computed from accelerometers on the waist made a particularly important contribution. We consider that the lower halves of relatively heavy subjects' bodies are stable and so the subjects move with less waist movement. This is reflected in the waist acceleration data. In fact, during the 'run' activities, the average x-axis energy value of the lighter subjects (lighter than 50 kg) was about 1.33 times larger than that of the heavier subjects (heavier than 70 kg).

[Sports experience] With 'pingpong experience,' the MAEs of SMOreg, Linear, and LMS were better than that of the naive method as shown in Table 4. However, we could not find a significant difference between the MAE of SMOreg and that of the naive method using a one-tail t-test ($p > .05$). Figure 6 (d) shows a scatter chart giving the SMOreg estimation results. As shown in the chart, it was difficult to distinguish the subjects whose pingpong experience was 'no (0)' from the subjects whose pingpong experience was 'somewhat (1).' However, the estimated values of the subjects whose pingpong experience was 'yes (2)' were relatively larger than those of other subjects in the chart. This may be because it was not very difficult to find subjects who were very good at playing pingpong by using their acceleration data. Note that these results are derived from only 'play pingpong' sensor data. Investigating sensor data of other sports and the effectiveness of our method on the data is one of our important future work. When we deal with sensor data of sports with multiple players such as football, we should consider the difference of sensor data among different roles (positions). For example, sensor data of a goalkeeper and those of an offensive player are completely different. Therefore, we should distinguish different roles.

In the above classification results, SMOreg outperformed Linear and LMS as shown in Table 4. This may be because SMOreg, which employs a polynomial kernel, is a non-linear regression method. Meanwhile, Linear and LMS are linear regression methods. We believe that non-linear methods are suitable for this problem.

4.3.3 Estimation with Few Activity Data and Small Numbers of Sensors

The above results show the estimation performance when we used sensor data of all 14 activities. Also, we used all four sensors. Here we show the estimation performance when we used sensor data of fewer activities from fewer sensors in **Table 6**. To compare estimation performance with the above results, we used Naive Bayes classifier (NB) and linear regression (Linear) for estimation.

We first show results obtained when using fewer activities. When we used only 'walk' sensor data from all four sensors, the classification accuracy for 'gender' was 83.6%. It was not very different from that of when we used sensor data of all activities from all four sensors. Also, estimating the dominant hand of a person by using only 'walk' sensor data was difficult. When we used only 'walk' sensor data from all of the four sensors, the MAEs related to 'age,' 'height,' and 'weight' were 7.49, 4.63 cm, and 5.29 kg, respectively. These results were somewhat poorer than those of when we used sensor data of all activities from all four sensors. When we used only 'run' sensor data from all of the four sensors, the estimation accuracies were somewhat poorer than those when using 'walk' sensor data. The 'walk' activity is very common in our daily life and we found that it is useful for estimating basic physical characteristics such as gender, height, and weight.

We then show results obtained when using fewer sensors. When we used only sensor data from both hands, while the 'gender' and 'dominant hand' estimation accuracies were fine, the estimation accuracies for the other physical characteristics were poor. Meanwhile, when we used only sensor data from the waist and thigh, while the 'gender' estimation accuracy was poor, the accuracies for 'age,' 'height,' and 'weight' were somewhat fine. We believe that the wrist sensor data are useful for estimating 'gender' and the body sensor data (waist and thigh) are useful for estimating body-related physical characteristics (height and weight).

Table 6 also shows estimation results obtained when we used

Table 5 Activity recognition accuracies.

	precision	recall	F-measure
A: stand	0.892	0.917	0.904
B: walk	0.953	0.950	0.951
C: run	0.951	0.950	0.950
D: sit	0.877	0.919	0.897
E: ascend stairs	0.746	0.770	0.758
F: descend stairs	0.794	0.809	0.801
G: bicycle	0.948	0.950	0.949
H: brush teeth	0.714	0.735	0.724
I: wash dishes	0.837	0.828	0.833
J: use PC	0.833	0.852	0.842
K: draw on whiteboard	0.835	0.836	0.836
L: write in notebook	0.850	0.787	0.818
M: play pingpong	0.744	0.742	0.743
N: vacuum	0.742	0.716	0.728

Table 6 Estimation performances by NB and Linear using fewer sensors and activities.

	gender	dominant hand (writing)	pingpong experience (nominal)	age	height	weight
only **walk**	83.6	88.5	49.2	7.49	4.53	5.29
only **run**	68.9	86.9	36.1	7.12	5.51	6.58
only wrists	83.6	98.4	36.1	7.91	4.89	6.99
only waist and thigh	77.0	88.5	47.5	6.83	4.84	5.31
only non-dominant	70.5	100.0	44.3	7.83	5.45	7.57
only **walk** and non-dominant	65.6	88.5	49.2	7.81	6.04	7.06
only **run** and non-dominant	57.3	88.5	39.3	7.56	5.76	7.98
only **walk** and wrists	77.0	88.5	45.9	7.95	5.53	7.00
only **walk** and waist	59.0	86.9	42.6	7.95	5.30	7.14
only **walk** and thigh	65.6	88.5	44.3	7.10	5.23	5.86
only **run** and wrists	73.8	88.5	31.1	7.27	5.58	8.16
only **run** and waist	63.9	86.9	42.6	7.66	5.94	7.64
only **run** and thigh	73.8	86.9	47.5	7.71	6.15	7.35

sensor data from the non-dominant hand. So, we assume that we use an accelerometer embedded in a smart watch. However, the classification accuracy for 'gender' and the MAEs related to 'age,' 'height,' and 'weight' were very poor. This may be because our participants mainly used their dominant hands to perform the activities.

We show results obtained when using fewer sensors and fewer activities. When we used only 'walk' sensor data from both hands, the classification accuracy for 'gender' was 77.0%. It was about 10% lower than that of when we used sensor data of all activities from all four sensors (86.9%). When we used only 'walk' sensor data from the waist, the classification accuracy for 'gender' was 59.0%. Also, when we used only 'walk' sensor data from the thigh, the classification accuracy for 'gender' was 65.6%. That is, walk sensor data from the hands contributed to estimating 'gender.' However, when we used only 'walk' sensor data from the non-dominant hand, the classification accuracy for 'gender' was only 65.6%. In contrast, when we used only 'walk' sensor data from the dominant hand, the classification accuracy for 'gender' was 83.9%. From these results, we believe that 'walk' sensor data from the dominant hand are useful for estimating 'gender.' Thoroughly investigating the reason why sensor data from the dominant hand are useful is our important future work.

When we used only 'walk' sensor data from both hands, the MAEs related to 'height' and 'weight' were 5.53 cm and 7.00 kg, respectively. Also, when we used only 'walk' sensor data from the waist, the MAEs related to 'height' and 'weight' were 5.30 cm and 7.14 kg, respectively. In addition, when we used only 'walk' sensor data from the thigh, the MAEs related to 'height' and 'weight' were 5.23 cm and 5.86 kg, respectively. That is, 'walk' sensor data from the thigh contributed to estimating 'height' and 'weight.'

Meanwhile, when we used only 'run' sensor data from few sensors, the estimation accuracies were somewhat poorer than those of 'walk' sensor data as shown in Table 6. We believe that 'walk' sensor data are more useful than 'run' sensor data for estimating physical characteristics.

4.3.4 Accuracy of Activity Recognition

Here we discuss the effect of the accuracy of activity recognition on the physical characteristics estimation. **Table 5** shows the accuracy of activity recognition. As shown in the results, we can recognize the activities with high accuracies, and this result is comparable to previous activity recognition studies [1], [13], [19]. Also, the row of 'NB(100)' in Table 3 shows the accuracies of nominal physical characteristics when we assume that the activity recognition accuracy is 100%. In addition, the row of 'SMOreg(100)' in Table 4 shows the accuracies of numerical physical characteristics when we assume that the activity recognition accuracy is 100%. As shown in the results, even when the activity recognition accuracy is 100%, the estimation results are not very different from the other methods. Therefore, we believe that a further improvement of the activity recognition cannot improve the physical characteristics estimation accuracy so much. Note that the classification accuracy of NB(100) for 'pingpong experience' is much higher than the accuracies of the other methods. This may be because the activity recognition accuracy of 'play pingpong' is poor as shown in Table 5.

5. Conclusion

In this paper, we experimentally investigated the kinds of physical characteristic that can be successfully estimated from activity acceleration data by using vast amounts of sensor data obtained from 61 subjects. As a result, we were able to estimate such physical characteristics as the dominant hand in an activity that are easy to detect using accelerometers with very high accuracies. In addition, we were able to estimate such characteristics as gender, height, and weight that are not directly apparent to accelerometers with good accuracies by employing our activity similarity-based approach. Estimating physical characteristics permits us to realize new applications that automatically change the services provided to an end user according to her estimated physical characteristics (e.g., characteristics-based recommender applications).

As a part of our future work, we plan to develop a shopping basket with an accelerometer to estimate physical characteristics of shoppers. Note that the ways of carrying a basket can depend on shoppers, e.g., carrying a shopping basket in the hand or in the crook of the arm, and thus the collected sensor data also depend on the ways of carrying. To cope with this problem, we should prepare an estimation model for each way of carrying, and switch the model depending on the way of carrying. In order to distinguish the ways of carrying a basket, we can utilize previous studies detecting the sensor (smart phone) position (e.g., pants pocket vs. breast pocket) [4].

References

[1] Bao, L. and Intille, S.S.: Activity recognition from user-annotated acceleration data, *Pervasive 2004*, pp.1–17 (2004).

[2] Choi, J., Lee, J., Hwang, H., Kim, J., Park, J. and Shin, K.: Estimation of activity energy expenditure: Accelerometer approach, *International Conference of the Engineering in Medicine and Biology Society*, pp.3830–3833 (2006).

[3] Dempster, A., Laird, N., Rubin, D., et al.: Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B* (*Methodological*), Vol.39, No.1, pp.1–38 (1977).

[4] Fujinami, K. and Kouchi, S.: Recognizing a Mobile Phone's Storing Position as a Context of a Device and a User, *MobiQuitous 2013*, pp.76–88 (2013).

[5] Gafurov, D., Helkala, K. and Sondrol, T.: Biometric gait authentication using accelerometer sensor, *Journal of Computers*, Vol.1, No.7, pp.51–59 (2006).

[6] Ghasemzadeh, H., Loseu, V., Guenterberg, E. and Jafari, R.: Sport training using body sensor networks: A statistical approach to measure wrist rotation for golf swing, *International Conference on Body Area Networks*, pp.1–8 (2009).

[7] Jago, R., Anderson, C., Baranowski, T. and Watson, K.: Adolescent Patterns of Physical Activity: Differences by Gender, Day, and Time of Day, *American Journal of Preventive Medicine*, Vol.28, No.5, pp.447–452 (2005).

[8] Kao, H.-L.C., Ho, B.-J., Lin, A.C. and Chu, H.-H.: Phone-based gait analysis to detect alcohol usage, *UbiComp 2012*, pp.661–662 (2012).

[9] Klesges, R., Klesges, L., Haddock, C. and Eck, L.: A longitudinal analysis of the impact of dietary intake and physical activity on weight change in adults, *The American Journal of Clinical Nutrition*, Vol.55, No.4, pp.818–822 (1992).

[10] Kononenko, I.: Estimating Attributes: Analysis and Extensions of RELIEF, *European Conference on Machine Learning*, pp.171–182 (1994).

[11] Kunze, K., Lukowicz, P., Junker, H. and Tröster, G.: Where am I: Recognizing On-Body Positions of Wearable Sensors, *LOCA 2005*, pp.264–275 (2005).

[12] Logan, B., Healey, J., Philipose, M., Tapia, E. and Intille, S.: A long-term evaluation of sensing modalities for activity recognition, *Ubicomp 2007*, pp.483–500 (2007).

[13] Maekawa, T., Yanagisawa, Y., Kishino, Y., Ishiguro, K., Kamei, K., Sakurai, Y. and Okadome, T.: Object-based activity recognition with heterogeneous sensors on wrist, *Pervasive 2010*, pp.246–264 (2010).

[14] Mantyjarvi, J., Lindholm, M., Vildjiounaite, E., Makela, S. and Ailisto, H.: Identifying users of portable devices from gait pattern with accelerometers, *IEEE International Conference on Acoustics, Speech, and Signal Processing* (*ICASSP'05*), Vol.2 (2005).

[15] Nachman, L., Baxi, A., Bhattacharya, S., Darera, V., Deshpande, P., Kodalapura, N., Mageshkumar, V., Rath, S., Shahabdeen, J. and Acharya, R.: Jog Falls: A Pervasive Healthcare Platform for Diabetes Management, *Pervasive 2010*, pp.94–111 (2010).

[16] NIKE Inc.: NIKE+, available from ⟨http://nikeplus.nike.com/plus/⟩.

[17] Petrofsky, J., Lee, S. and Bweir, S.: Gait characteristics in people with type 2 diabetes mellitus, *European Journal of Applied Physiology*, Vol.93, No.5, pp.640–647 (2005).

[18] Rabbi, M., Ali, S., Choudhury, T. and Berke, E.: Passive and In-Situ assessment of mental and physical well-being using mobile sensors, *UbiComp 2011*, pp.385–394 (2011).

[19] Ravi, N., Dandekar, N., Mysore, P. and Littman, M.: Activity recognition from accelerometer data, *IAAI 2005*, Vol.20, No.3, pp.1541–1546 (2005).

[20] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J.: Item-based collaborative filtering recommendation algorithms, *WWW 2001*, pp.285–295 (2001).

[21] Strohrmann, C., Harms, H. and Tröster, G.: What do sensors know about your running performance?, *Int'l Symp. Wearable Computers*, pp.101–104 (2011).

[22] Troiano, R., Berrigan, D., Dodd, K., Masse, L., Tilert, T. and McDowell, M.: Physical activity in the United States measured by accelerometer, *Medicine & Science in Sports & Exercise*, Vol.40, No.1, pp.181–188 (2008).

[23] Weiss, G.M. and Lockhart, J.W.: Identifying user traits by mining smart phone accelerometer data, *5th International Workshop on Knowledge Discovery from Sensor Data*, pp.61–69 (2011).

[24] Welk, G. and Differding, J.: The utility of the Digi-Walker step counter to assess daily physical activity patterns, *Medicine & Science in Sports & Exercise*, Vol.32, No.9, pp.S481–S488 (2000).

[25] Witten, I. and Frank, E.: *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann (2004).

Akira Masuda is a student at Graduate School of Information Science and Technology, Osaka University, Japan. His research interests include sensor data mining and activity recognition.

Takuya Maekawa is an associate professor at Osaka University, Japan. His research interests include ubiquitous and mobile sensing. He received his Ph.D. in information science and technology from Osaka University in 2006.