Pattern Anal Applic
DOI 10.1007/s10044-012-0300-z

ORIGINAL ARTICLE

Recognizing multi-modal sensor signals using evolutionary learning of dynamic Bayesian networks

Young-Seol Lee · Sung-Bae Cho

Received: 29 February 2012 / Accepted: 12 September 2012

Abstract Multi-modal context-aware systems can provide user-adaptive services, but it requires complicated recognition models with larger resources. The limitations to build optimal models and infer the context efficiently make it difficult to develop practical context-aware systems. We developed a multi-modal context-aware system with various wearable sensors including accelerometers, gyroscopes, physiological sensors, and data gloves. The system used probabilistic models to handle the uncertain and noisy time-series sensor data. In order to construct the efficient probabilistic models, this paper uses an evolutionary algorithm to model structure and EM algorithm to determine parameters. The trained models are selectively inferred based on a semantic network which describes the semantic relations of the contexts and sensors. Experiments with the real data collected show the usefulness of the proposed method.

Keywords Multi-modal sensor recognition · Dynamic Bayesian networks · Evolutionary algorithm · Selective inference

1 Introduction

Recent advances in the development of the mobile sensor technologies enable us to collect a large amount of data for activity recognition [1]. Activity-aware systems provide new applications in many fields such as health care support, smart environments, surveillance, emergency response, military missions [2], industrial work-flow optimization, and novel HCI paradigms [3]. Moreover, activity recognition is also important to analyze human behavior, routines, rituals, and social relationships. Wearable sensors can provide effective data collection about the user's behaviors. There were several attempts to recognize user contexts and to provide situation-appropriate services using the wearable sensors [4]. These systems used the fusion of various sensors to perceive the high-level contexts more accurately [5].

Activity recognition model for practical applications has to detect diverse activities that are performed under many different environment conditions. The model must have the ability to deal with the real-world's noisy data and complexities. Furthermore, it is difficult to train the inference model because a large number of sensors are used for the systems and there are highly complicated causal relationships and constraints among sensing information. On the other hand, the energy consumption and computational complexity must be considered [6, 7]. Murao et al. [8] tried to reduce energy consumption and maintain the accuracy of activity recognition in wearable computing environment by dynamic sensor selection. They developed a wearable sensor system to group correlated sensors with contexts and shut off unnecessary sensors for power saving. Other researchers have studied modularization methods which include grouping correlated sensors with contexts [9] and decomposing the complex problem into sub-problems [10] to work out the problems.

However, the optimal model training and the efficient context recognition are still important issues to construct the systems effectively. In order to recognize context efficiently, it is important to select necessary models and perform inference with the selected models. Although the models are successfully modularized, all the models have

Y.-S. Lee · S.-B. Cho (✉)
Department of Computer Science, Yonsei University,
50 Yonsei-ro, Sudaemoon-gu, Seoul 120-749, Korea
e-mail: sbcho@cs.yonsei.ac.kr; scho621@gmail.com

Published online: 26 September 2012

to be inferred because the systems do not know which contexts have to be recognized and which contexts do not need to be identified. In this case, the essential problems of the large-scale context-awareness still remain.

This paper presents a novel context-aware system using multi-modal wearable sensors such as accelerometers, a physiological sensor, and flexure sensors. In order to reduce the cost of inference for complex contexts under a computationally restricted environment, the necessary model is selected by the semantic network (SN) which represents the relations among the sensor values and contexts, and the selected model is used to recognize suitable contexts. The system recognizes user's activities using probabilistic models which are built by evolutionary algorithm. The optimal probabilistic models, one context versus-all (OVA) dynamic Bayesian networks (DBNs), deal with the uncertain, incomplete, and temporal observations from the sensors.

2 Related works

Context-aware computing has been an active area of research in recent years due to its potential and usefulness. Current research in activity recognition using wearable sensors covers various fields such as user-adaptive service, service recommendation, and health management. Accelerometer is one of the most popular sensors to detect human behaviors because of its usefulness and efficiency. Bao and Intille [11] attempted to recognize human behaviors using five 2-axis accelerometers. They used C4.5 classifier to recognize behaviors which include 20 different daily activities like 'walking', 'running', 'watching TV', and so on. Brezmes et al. [12] studied k-nearest neighbor algorithm-based activity recognition using a Nokia 95 mobile phone. They considered a user's way to hold the mobile phone (a chest pocket, a front pocket, a rear trouser pocket, and an inner jacket pocket, etc.). Huỳnh et al. [13] modeled low-level activities and high-level activities. They recognized two types of activities from a 2-axis accelerometer and nine tilt switches for sensing a user's motion and orientation. Jatoba et al. [14] used a 3-axis accelerometer to estimate daily life activities such as walking, jogging, climbing upstairs, etc. They compared the performance of six different methods which include ID3 decision tree, k-nearest neighbor, naïve Bayes classifier, nearest neighbor, classification and recognition tree, and adaptive neuro-fuzzy inference system. Godfrey et al. [15] investigated human movement detection using acceleration. They analyzed and summarized various methods to determine a user's activities using accelerometers. Murao et al. [8] developed CLAD prototype wearable sensor platform, and tried to recognize a user's activities with the sensor platform which includes five 3-axis accelerometers.

Several researchers have investigated context awareness using multi-modal sensors as well as accelerometers. Blum et al. [16] investigated to recognize diverse contexts such as eating, typing, shaking hands, clapping hands, driving, brushing teeth, and washing the dishes. Maurer et al. [17] developed eWatch which is a wearable sensing platform to gather light, acceleration, audio, and temperature. They recognized a user's current location and activities using some classifiers such as C4.5 decision tree, k-nearest neighbor, naïve Bayes, and Bayesian network from the sensor data. Oliver et al. [18] proposed the use of layered probabilistic representations for modeling office activities such as phone conversation and presentation. They collected real-time streams of evidence from video, audio, and computer interactions.

Harrison et al. [19] presented their experiences to build a wearable system for automatic activity recognition. They developed mobile sensing platform (MSP) and UbiFit Garden system which infers several types of activities (e.g., walking, running, cycling, etc.). Ermes et al. [20] recognized daily activities and sports with wearable sensors of 3D accelerometers and a GPS receiver, in which the activities included walking, running, cycling with an exercise bike, playing football, etc. Oliver and Horvitz [21] developed S-SEER system which added environmental sensors on personal computer systems to recognize human activities. They used layered hidden Markov models (LHMMs) to diagnose a user's activity from video, audio, and computer (keyboard and mouse) interactions. Schmidt et al. [22] presented a layered architecture for context-aware adaptation based on collections of multimodal sensors (including a photodiode, two accelerometers, a passive IR, a temperature/pressure sensor, and a CO gas sensor) in the TEA project. Li et al. recognized a user's physical activities from electrocardiogram (ECG) and accelerometer signals [23]. They extracted two types of features (temporal and cepstral features) and used support vector machine (SVM) and Gaussian mixture model (GMM) to recognize physical activities. Wang et al. [6] proposed a hierarchical sensor management scheme based on the state transition rules which turn on and off the sensors. Krause et al. [7] suggested a collection of selective sampling strategies to reduce the number of required sensor readings and the computation duty cycles even further.

However, most studies have typically performed all models constructed from training data. They tried to recognize all types of context in any time, and it requires additional cost to infer unnecessary context. In this paper, we propose a context-aware system using selective inference to deal with various contexts efficiently. The proposed system performs inference process using only the selected model.

3 Multi-modal context-aware system

Figure 1 shows the overview of the proposed system to recognize a user's behaviors from various wearable sensors. In this paper, four types of sensors are used to collect data as shown in Fig. 1. The input data from accelerometers, flexure sensors, and a physiological sensor are pre-processed through smoothing, normalization, and clustering. The user input and web data are used to select DBN models by the semantic network in which the activation values (degrees of occurrence) are propagated from low-level sensor nodes to high-level context nodes. The corresponding DBN module is chosen by the activated contexts, and then the probabilities of the contexts are calculated by the chosen OVA DBNs which are trained by using GA. Figure 2, where x axis denotes time and y axis means ten types of attributes, illustrates raw data samples.

3.1 Sensor data collection and feature extraction

A sensor platform used in this paper integrates two data gloves, the armband on a right arm, five sets of accelerometers, and gyroscopes on a head and both arms and wrists as shown in Fig. 3. The platform includes various sensors as shown in Table 1. User input and web data are not handled in this process because they are only used to select DBN models as mentioned in previous section.

Data gloves (5DT Inc.: http://www.5dt.com) sense the degree of bending fingers from five channels of the left hand and 14 channels of the right hand. Bodymedia's armband (http://sensewear.bodymedia.com) captures the user's physiological signals, related to emotion and stress, including a skin temperature, heat flux, and galvanic skin response. In order to track human motion, XSens XBus Kit Mtx (http://http://www.xsens.com) is used which includes a 3-axis accelerometer and a 3-axis gyroscope for each sensor node. Table 2 summarizes the collected data from the sensors.

Since raw data include continuous values collected from diverse types of sensors, it is required to integrate them in a unified format. The continuous values are smoothed to remove sparking noise using sliding window and normalized to calculate the distance among data. The system discretizes the data into three states by k-means clustering as shown in Fig. 4. Here, k value, the number of clusters, is empirically determined because the cluster validation measures such as Dunn's index were not effective in this case. Determining it automatically should be one of the issues that we have to explore for the future work. The preprocessing procedure calculates the average of current and previous data within the window as follows:

$$D_{\text{smoothed}} = \frac{\sum_{i=1}^{N} D_i}{N}, \quad D_i = \{x_1, \ldots, x_m\} \tag{1}$$

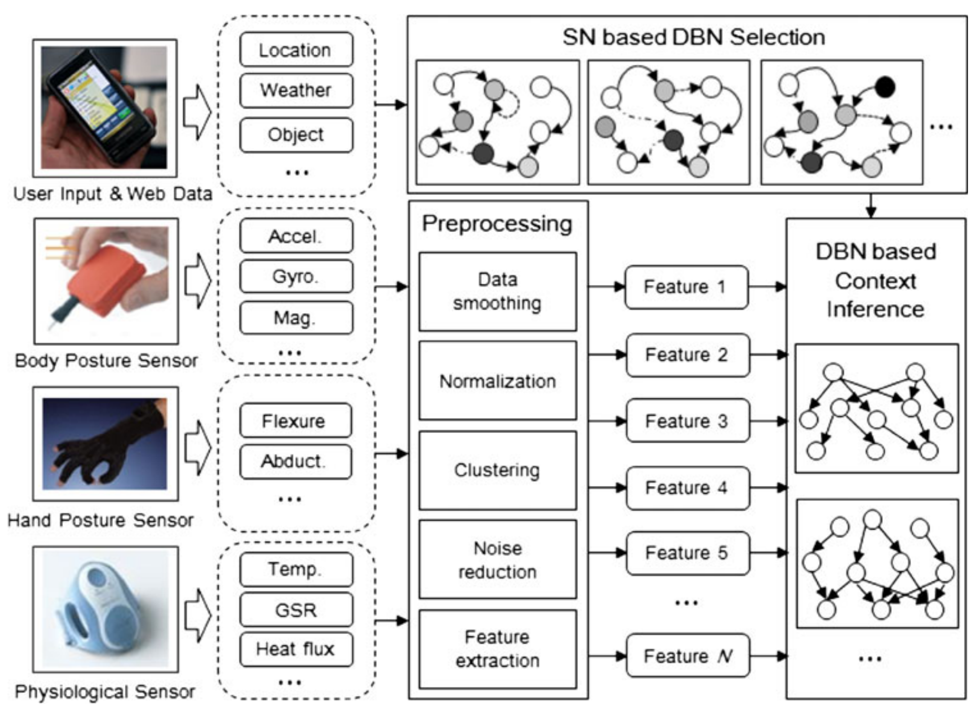

Fig. 1 The proposed system overview

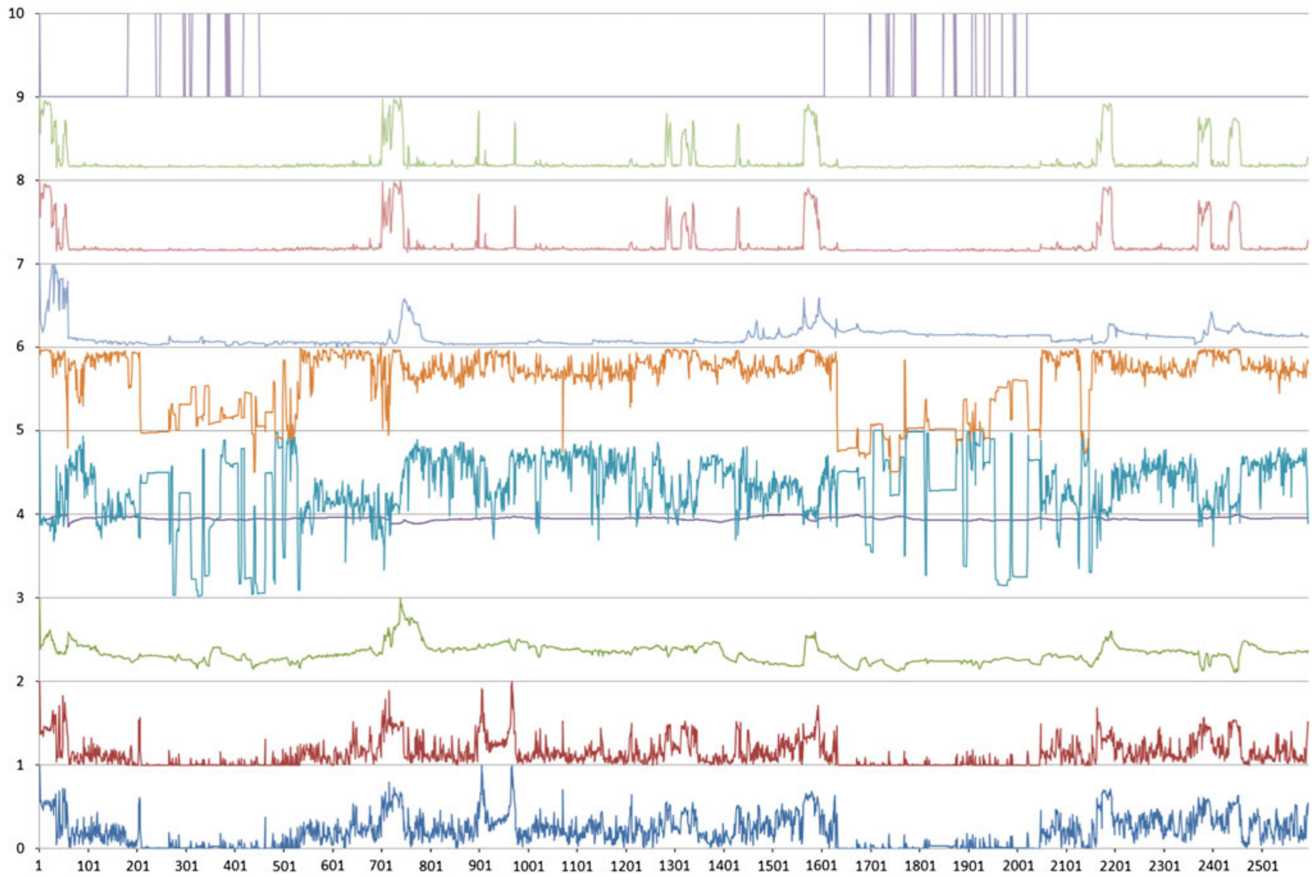

Fig. 2 Multi-modal sensor data (*1 x* acc., *2 y* acc., *3* heat flux, *4* skin temp., *5* 2nd *x* acc., *6* 2nd *y* acc., *7* GSR, *8* energy expenditure, *9* METS, *10* sleep activity)

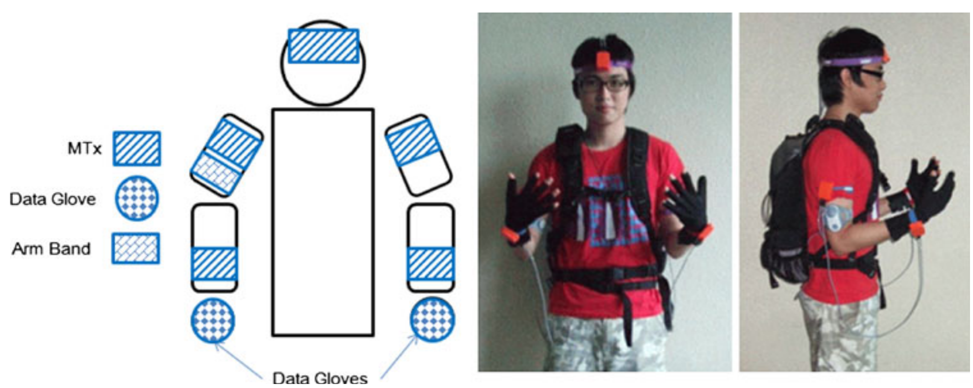

Fig. 3 Sensor positions on a user's body

where D_i is a vector which represents an input sensor data item, and N is a window size for smoothing. The raw data often contain some noises due to the sensitivity of sensors, and this rectangular smoothing reduces the noises and captures significant patterns in the data.

After smoothing, we normalize the continuous values to calculate the distance among the data. Each sensor produces the values with its own scale, and it can be compared after normalization like the following equation:

$$x_{i\_\mathrm{normalized}} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{2}$$

where x_i is a real value, which lies between $\min(x)$ and $\max(x)$. Some sensor values such as 3-axis acceleration and 3-axis angular velocity have more than one attribute and the features v_{scalar} and v_{delta} are extracted as follows:

Table 1 Multi-modal sensors for context awareness

Module	Sensor Position	Specification
Bodymedia armband	Right upper arm	2-Axis accelerometer, heat flux sensor, skin temperature sensor, galvanic skin response (GSR) sensor
5DT data glove (right hand)	2 flexure sensors per finger, 1 sensor for knuckle and 1 for first joint	14 flexure sensors
5DT data glove (left hand)	1 flexure sensor per finger	5 flexure sensors
XSens XBus Kit (Mtx)	Head, right upper arm, right lower arm, left upper arm, left lower arm	3-Axis accelerometer, 3-axis gyroscopes

Table 2 Collected raw data from sensors

Module	Data
Armband	Time, longitudinal acceleration, transverse acceleration, galvanic skin response, skin temperature, near-body temperature, heat flux, average metabolic equivalent unit (Kcal)
Data Glove (Right hand)	Thumb angle 1, thumb angle 2, abduction between thumb and index, index angle 1, index angle 2, abduction between index and middle, middle angle 1, middle angle 2, abduction between middle and ring, ring angle 1, ring angle 2, abduction between ring and little, little angle 1, little angle 2
Data Glove (Left hand)	Thumb angle, index angle, middle angle, ring angle, little angle
XSens XBus Kit (Mtx)	x axis Euler angle, y axis Euler angle, z axis Euler angle, x axis acceleration (m/s^2), y axis acceleration (m/s^2), z axis acceleration (m/s^2), x axis angular velocity (rad/s), y axis angular velocity (rad/s), z axis angular velocity (rad/s), x axis magnetic field, y axis magnetic field, z axis magnetic field, temperature

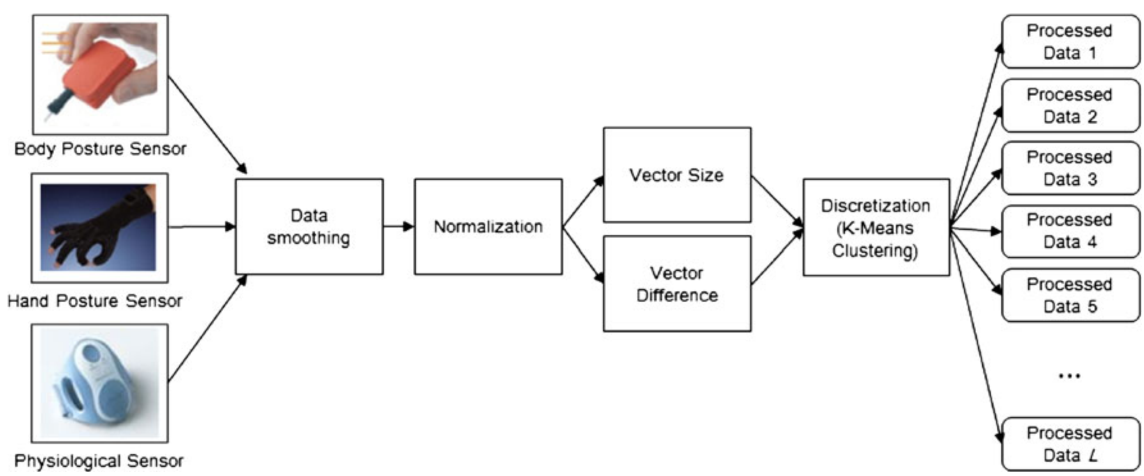

Fig. 4 Preprocessing process for sensor data

$$v_{\text{scalar}} = |v| = \sqrt{v_x^2 + v_y^2 + v_z^2} \tag{3}$$

where v_x, v_y, and v_z represent x axis, y axis, and z axis at a 3-dimensional vector v, respectively.

$$\begin{aligned} v_{\text{delta}} &= |v_{\text{current}} - v_{\text{previous}}| \\ &= \sqrt{(x_{\text{current}} - x_{\text{previous}})^2 + (y_{\text{current}} - y_{\text{previous}})^2 + (z_{\text{current}} - z_{\text{previous}})^2} \end{aligned} \tag{4}$$

where v_{previous}, and v_{current} mean previous sensory input and current sensory input of acceleration or angular velocity.

As the final step in the preprocessing, we discretize the continuous value into discrete state by using k-means clustering to generate inputs suitable to discrete DBNs. The algorithm assigns the sensor data into different k sets to map the values to k discrete states.

3.2 Activity recognition using DBNs selected by semantic network

A Bayesian network is a powerful tool for representing a large number of joint probability among the variables, and it effectively handles the uncertainty as well [10]. DBN is a Bayesian network designed suitable for sequential data.

Each DBN is used to estimate the probability of the context, $P(C_c = \text{yes}|C_p, O_c, O_p)$, which is calculated under Bayes theorem with Markov chain by:

$$P(C_c = \text{yes}|C_p, O_c, O_p) = \frac{P(C_c = \text{yes}, C_p, O_c, O_p)}{\sum_{A_c=\text{yes,no}} P(C_c, C_p, O_c, O_p)},$$
$$\text{where } P(C_c, C_p, O_c, O_p) = P(C_p)P(O_p)P(C_c|C_p, O_p)P(O_c|C_c). \quad (5)$$

An expert DBN computes the observation probability of the corresponding activity state using a number of continuing sensory observations.

Semantic network is a directed graph which represents semantic relations (links) among concepts or common sense knowledge denoted by nodes [24]. It can be used for common sense representation and case-based reasoning. For example, Kofod-Petersen et al. [25] built semantic network based on Activity Theory and applied it to recognize human behaviors. This paper uses a semantic network to define sematic relations among sensor data and a user's activities to select candidate activities from the raw data as shown in Fig. 5. Table 3 summarizes the components of the semantic network. There are 21 kinds of sensor nodes and 11 context nodes (activities) in the semantic network.

The activation values of sensor nodes in the semantic network are determined by the data collected from the sensors. Equation (6) denotes the degree of activation of the sensor nodes.

$$a_l = \sum_{t=1}^{n} f(s_t)/n \quad (6)$$

where a_l means the activation value of the sensor node, s_t is collected sensor data at time step t, n is window size, and f denotes the feature extraction function. After the activation of the sensor nodes, the context nodes are activated through the links among context nodes and sensor nodes. The activation values (the degree of probable occurrence for the node) of a node are propagated to other nodes with different weights according to the directions and the types of the links. There are six types of links in the proposed semantic network as {PartOf, Cause, Prevent, FollowedBy, HardToFollow, Prerequisite}, where the first three of them are used for defining relations between a sensor node and a context node, and the rests are connections between context nodes.

The propagated activation value from sensor node to context node a_{hl} is calculated according to Eq. (7).

$$a_{hl} = \sum f(a_l),\ f(a_l) = \begin{cases} w_{\text{PartOf}} a_l, & \text{when } n_l \to n_h = \text{`PartOf'} \\ w_{\text{Cause}} a_l, & \text{when } n_l \to n_h = \text{`Causes'} \\ -w_{\text{Prevent}} a_l, & \text{when } n_l \to n_h = \text{`Prevents'} \end{cases} \quad (7)$$

where weight w depends on the three types of relations (PartOf, Cause, and Prevent). The activation value for the sensor node a_l is estimated using simple rules with feature vectors.

The activation value between a context node and other context node is propagated through Prerequisite, FollowedBy, and HardToFollow relations as the following Eq. (8):

$$a_{hh} = \sum f(a_h),\ f(a_h) = \begin{cases} w_{\text{FollowedBy}} a_h, & \text{when } n_{h1} \to n_{h2} = \text{`FollowedBy'} \\ w_{\text{Prerequisite}} a_h, & \text{when } n_{h1} \to n_{h2} = \text{`HasPrerequisite'} \\ -w_{\text{HardToFollow}} a_h, & \text{when } n_{h1} \to n_{h2} = \text{`HardToFollow'} \end{cases} \quad (8)$$

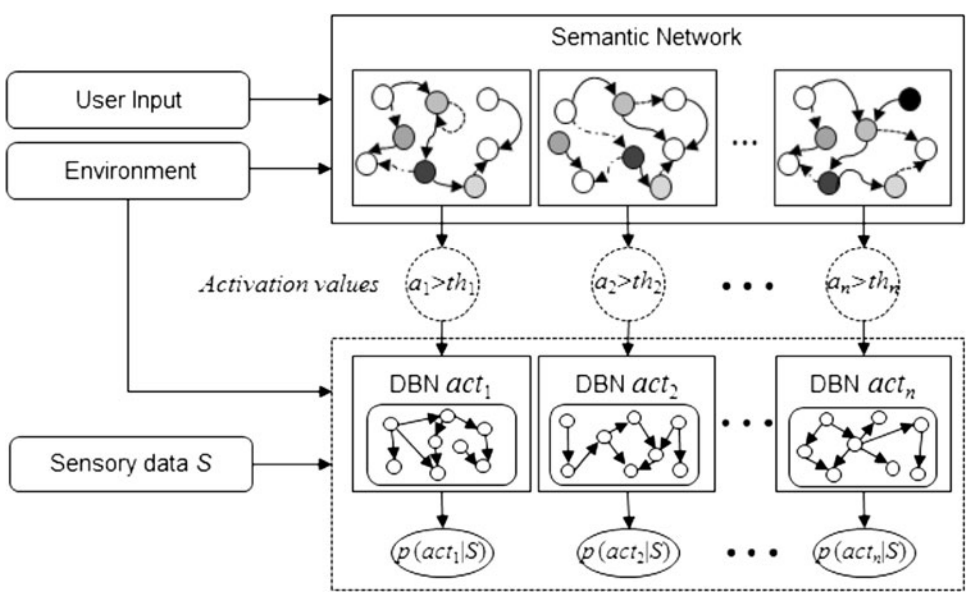

Fig. 5 The semantic network based DBN selection

Table 3 Semantic network components

Type	Semantic network nodes
Sensor	
Time	Time:Morning, Time:BeforeNoon, Time:Noon, Time:Evening
User input	Object:Computer, Object:Keyboard, Object:Mouse, Object:Projector, Environment:Bright, Environment:Noisy, Location:Outdoor, Location:Playground, UserState:Tired, Schedule:Meeting
XBus Sensors	Sensor:NoHeadMotion, Sensor:NoRightArmMotion, Sensor:NoLeftArmMotion
Web	Weather:Sunny, Weather:Rainy, Weather:Snowy, Temperature:Warm
Context (activities)	Sleeping, Eating, TeethBrushing, Driving, Working, Meeting, Calling, Drinking, Urinating, Exercising, Resting

In this paper, the weight values were defined as $\{w_{\text{PartOf}}, w_{\text{Cause}}, w_{\text{Prevent}}, w_{\text{Prerequisite}}, w_{\text{FollowedBy}}, w_{\text{HardToFollow}}\} = \{0.5, 0.6, -0.2, 0.7, 0.3, -0.5\}$. The w values are manually designed by considering the strength of influence with prior knowledge. The previous state of the context can affect the activation of the context. It is defined as a_{hr} which have three types between previous and current context nodes: not executed (inactivated), rejected (the output probability of the inference model is less than a threshold), and accepted. Let $P(t)$ be the output value of the context inference model at time t. Here, a_{hr} is calculated as the following Eq. (9):

$$a_{hr} = \begin{cases} m_{\text{not\_executed}} P(t'), & \text{where the context is not executed,} \\ -\sum_{t=1}^{n} \left(1 - \dfrac{1}{1+\mathrm{e}^{P(t)}}\right) / n, & \text{where the context is rejected,} \\ \sum_{t=1}^{n} \left(\dfrac{1}{1+\mathrm{e}^{P(t)}}\right) / n, & \text{where the context is accepted.} \end{cases} \tag{9}$$

If the DBN model for a context was not executed at the previous time t, $P(t')$ the output value at the latest execution time t' is used with the momentum variable $m_{\text{not\_executed}}$. If the recognition engine rejected the context, the activation value is decreased by the inverse sigmoid function. Otherwise (accepted state), the sigmoid function is used to increase the activation value.

The activation value of a context node a_h is estimated as follows:

$$a_h = a_{hl} + a_{hh} + a_{hr}. \tag{10}$$

Here, a_{hl} and a_{hh} are the activation values propagated from the sensor nodes and other context nodes, respectively, while a_{hr} is estimated from the previous recognition state of the corresponding context (activation from goal).

3.3 Training DBNs using GA

In order to construct a DBN, a model structure has to be identified first, and conditional probability tables (CPTs) are then estimated.

Since time-series variables often lead to complexity problems, the proposed system applies GA to finding optimal structure among the large number of variables, while it trains the CPTs using an expectation–maximization (EM) algorithm. GA is a search technique based on the evolution of biological system [26]. According to a predefined evaluation criterion called fitness (the recognition rate for the training dataset in this paper), the fittest individuals survive into the next generation after some modifications using crossover and mutation.

Figure 6 illustrates the overall process to train DBN model using GA. First, we construct a random initial population where the structure of DBN is represented as a chromosome. There are four kinds of nodes in the DBN as shown in Fig. 7.

Current context C_c, the target node for activity recognition, is connected from C_p, the past state of the C_c, and observations of the past time O_p. C_c is also linked to observations at the current time O_c. Each DBN model includes only one C_c (a user's behavior) as an OVA DBN. Figure 8 illustrates a fixed-length binary string for the chromosome to represent a DBN for a C_c.

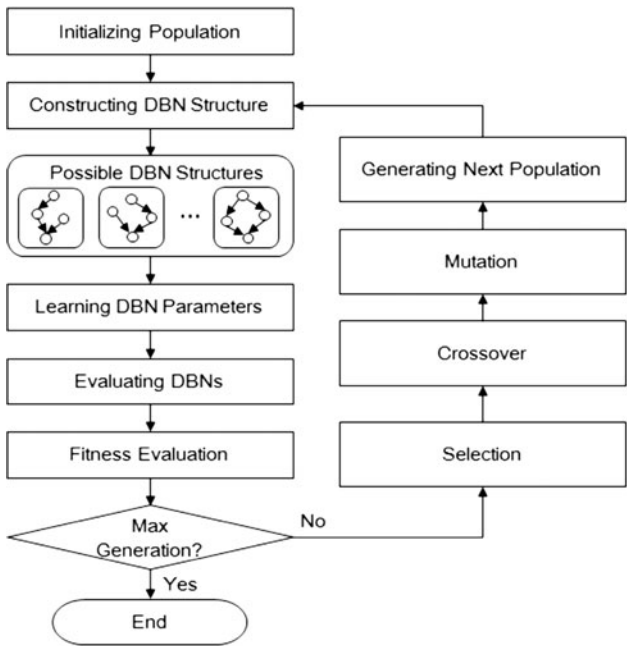

Fig. 6 DBN learning process using GA

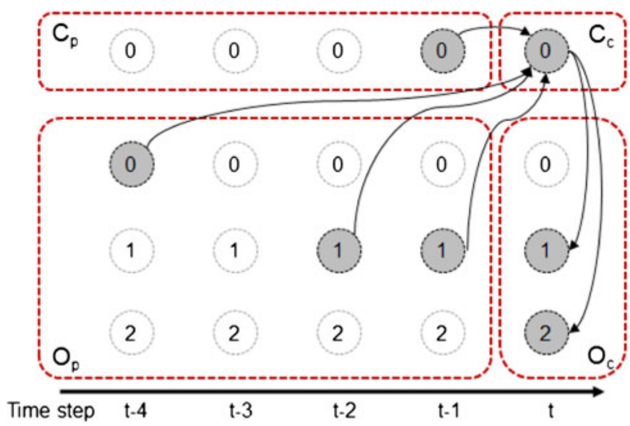

Fig. 7 Four kinds of nodes in the DBN

In order to calculate the fitness of a DBN structure, the performance of the DBN is evaluated for a given dataset. The fitness of an OVA DBN is evaluated by the Eq. (11).

$$f = 0.5 \times \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} + 0.5 \times \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}} \quad (11)$$

where TP denotes the number of true positive, and FP is the number of false positive. TN and FN are the numbers of true negative and false negative, respectively. The fitness function considers both true positive rate and true negative rate. There were some studies to use the ratio of positive and negative samples to train an optimal classifier from imbalanced data [27, 28].

In the selection phase, a roulette wheel method is adopted to select chromosomes to be crossed over and mutated. The crossover process recombines the genetic material of two parent chromosomes by exchanging pairs of their genes. Here, two-point crossover is used where the first point is chosen from C_p and O_p, while the second point is picked from O_c. We used elitism which reserves two slots in the next generation for the highest scoring chromosome of the current generation, without allowing crossover and mutation.

4 Experiments

This section presents the experiments conducted to evaluate the proposed system. We evolved DBN models and evaluated the performance of the context recognition using the collected data.

4.1 Sensor data collection

In order to verify the performance of the proposed system, we collected raw data for 11 activities {Sleeping, Eating, TeethBrushing, Driving, Working, Meeting, Calling, Drinking, Urinating, Exercising, Resting} out of daily life as shown in Table 4. For a training set of the DBNs, one in five in the dataset is used, and the others are for test set. The participants are five male graduate students and they freely collected the data. After the data collection, the target activities are selected as frequent activities in their daily life.

4.2 Evolution of DBNs

Eleven OVA DBNs for context recognition were trained with the following parameters.

- Number of generations: 50
- Number of populations: 50
- Crossover rate: 0.8
- Mutation rate: 0.1
- Selection rate: 0.8
- Elitist preserving rate: 0.02

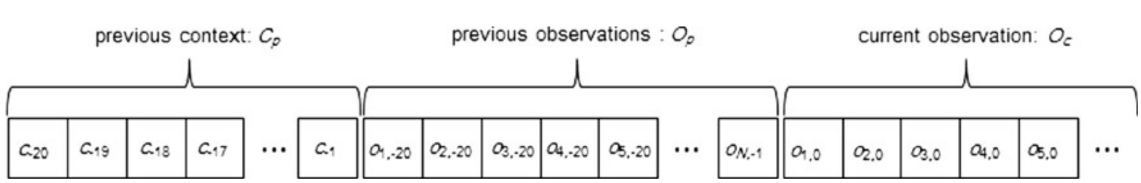

Fig. 8 Chromosome for DBN structure

Table 4 Collected data and size

Activity	Size (min)	Activity	Size (min)
Sleeping	475	Calling	431
Eating	708	Drinking	90
TeethBrushing	315	Urinating	307
Driving	526	Exercising	298
Working	1,933	Resting	106
Meeting	314		

In this paper, the evolved DBNs are evaluated using precision to recognize 11 contexts. We have used EM algorithm to set the conditional probability tables of nodes in a given graph structure. Figure 9 shows the fitness changes of six DBNs through the evolutionary process. In this figure, x axis denotes the number of generations, and y axis represents the fitness of the trained DBNs.

The result divides the evolution into two types: convergence and oscillation. In the convergence case, the DBN can reach the optimal network structure in 5–6 generations as Meeting, Calling, and Drinking in Fig. 9. Driving, Working, and Exercising belong to oscillation case in this figure. Although they showed good fitness in early generations, they cannot maintain the good DBN structure for long period. The oscillating activities are mostly long-term activities. It implies that the evolved DBNs are more appropriate to recognize short-term activities such as Calling and Drinking.

4.3 DBN selection and inference

According to the domain knowledge and some constraints, the semantic relations among the contexts can be represented as shown in Tables 5 and 6. Based on the definitions of the variables and relations, we designed the SN which consists of 33 nodes (11 context nodes and 21 sensor nodes) and 34 relations (10 PartOf, 8 Cause, 2 Prevent, 7 FollowedBy, 5 HardToFollow, and 2 Prerequisite) as shown in Fig. 10.

There are many variables for each DBN model which has the average 4,584 conditional probabilities to recognize the current context. The selective inference approach of the proposed system can efficiently deal with the problem by reducing the cost of inference. Each DBN module is required to estimate the current state of a specific context in the different environmental settings. For instance, 'TeethBrushing' activity can be activated by the other context 'Eating' and time constraints. If there are any changes in the context nodes and sensor nodes, it is necessary to recalculate the probability for 'TeethBrushing.' In a similar manner, the other context nodes require probabilistic inference only when the contexts related to each DBN is

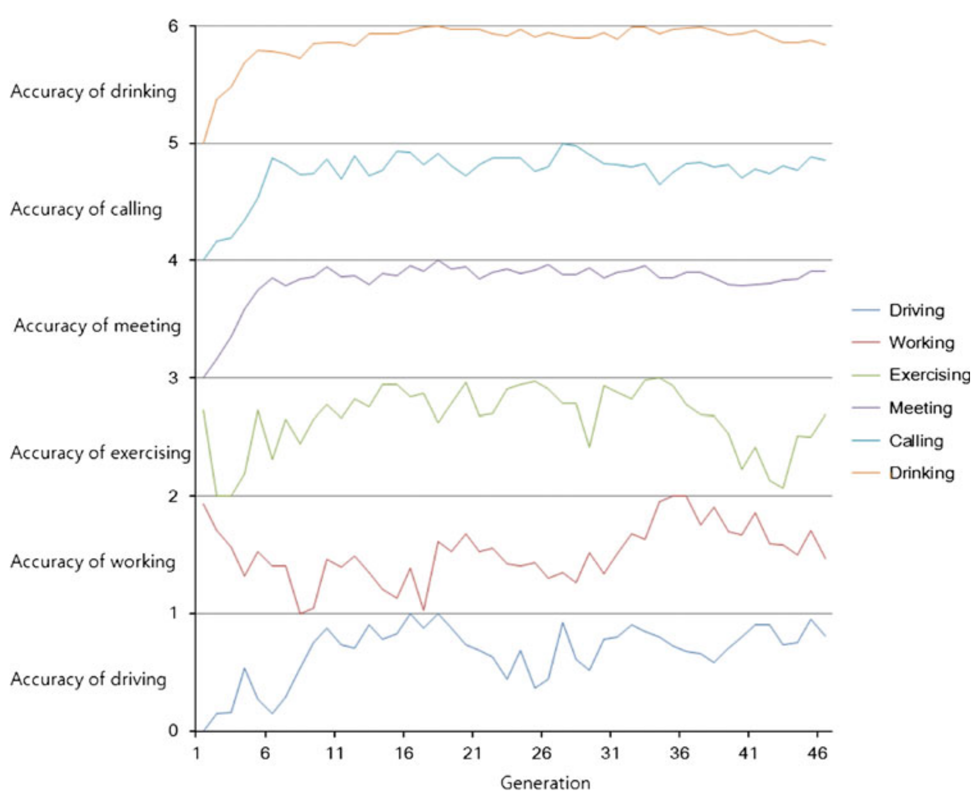

Fig. 9 Evolution of DBNs

Table 5 Semantic relations among the context nodes

	S	Ea	T	Dv	W	M	C	Dk	U	Ex	R
Sleeping (S)	–	▲	–	–	–	–	–	–	–	–	–
Eating (Ea)	–	▼	–	–	–	–	–	–	–	–	–
TeethBrushing (T)	–	▲	–	–	–	–	–	–	–	–	–
Driving (Dv)	–	–	–		–	–	–	–	–	–	–
Working (W)	–	–	–	–	–	–	–	–	–	–	–
Meeting (M)	–	–	–	–	–	–	–	–	–	▼	–
Calling (C)	–	–	–	▲	–	▲	–	–	–	▼	–
Drinking (Dk)	–	–	–	–	–	–	–	–	–	O	–
Urinating (U)	–	▲	–	–	–	–	–	▲	▼	–	–
Exercising (Ex)	–	▼	–	–	–	–	–	–	–	–	–
Resting (R)	–	▲	–	–	O	–	–	–	–	–	–

▲, FollowedBy; ▼, HardToFollow; O, HasPrerequisite; –, Unrelated

Table 6 Semantic relations among the context nodes and sensor nodes

	F	C	K	M	P	NHM	NRM	NLM	O	Mo	N	T	S	W	E	Ra	Sn
Sleeping	–	–	–	–	–	▲	▲	▲	–	–	–	O	–	–	–	–	–
Eating	–	–	–	–	–	–	–	–	–	–	O	–	–	–	–	–	–
TeethBrushing	–	–	–	–	–	–	–	–	–	O	–	–	–	–	–	–	–
Driving	–	–	–		–	–	–	–	O	–	–	–	–	–	–	–	–
Working	▲	▲	▲	▲	–	–	–	–	–	–	–	–	–	–	–	–	–
Meeting	–	–	–	–	▲	–	–	–	–	–	–	–	–	–	–	–	–
Calling	–	–	–	–	–	–	–	▲	–	–	–	–	–	–	–	–	–
Drinking	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Urinating	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Exercising	–	–	–	–	–	–	–	–	▲	–	–	–	O	O	O	▼	▼
Resting	–	–	–	–	–	–	–	–	–	–	–	O	–	–	–	–	–

F Forenoon, *C* computer, *K* keyboard, *M* mouse, *P* projector, *NHM* NoHeadMotion, *NRM* NoRightArmMotion, *NLM* NoLeftArmMotion, *O* outdoor, *Mo* morning, *N* noon, *T* tired, *S* sunny, *W* warm, *E* evening, *Ra* rainy, *Sn* snowy, ▲ PartOf, ▼ Prevent, O Cause, – Unrelated

activated. This approach reduces the number of inference for each DBN module dramatically.

In order to evaluate the effectiveness of the proposed method, we compared standard Bayesian network [17] (one of recognition methods introduced in Sect. 2) with it. Figure 11 shows the result of the comparison. All BNs in Fig. 11 means the standard inference using all models. The other methods (20 selection + BN, 10 selection + BN, 1 selection + BN) have the different number of selection and only one BN inference. In our method, a BN model is selected when the activeness of an activity node in a semantic network in Fig. 10 is greater than the predefined threshold. If any node cannot be activated, the threshold gets lower and the selection process is repeated. The number of selection in Fig. 11 represents the repeating number of the selection process. Although the selection process is repeated several times, the cost is much lower than the standard BN inference.

To show the performance of the DBN modules, we evaluate the precision of inference models on 11 activities as shown in Fig. 12. The result of the evaluation is summarized in Table 7 as a confusion matrix. The table includes 'Failed' state which means the classifier cannot recognize any activities. In other words, the state means that the probability of the result of any Bayesian network models cannot exceed the decision threshold. We remained the confusing case as 'Failed' state to reduce false positives. When the activity recognition results are analyzed, true positives and false positives are considered more importantly than true negatives and false negatives [29], because there are so many activities in a real life and the inferred activities are a small subset of all activities. The precision of most activities is higher than 0.9, but three activities of Calling (C), Urinating (U), and Resting (R) show worse performance.

As well as the evaluation of the precision, a male subject further conducted the scenario based test to check the usability.

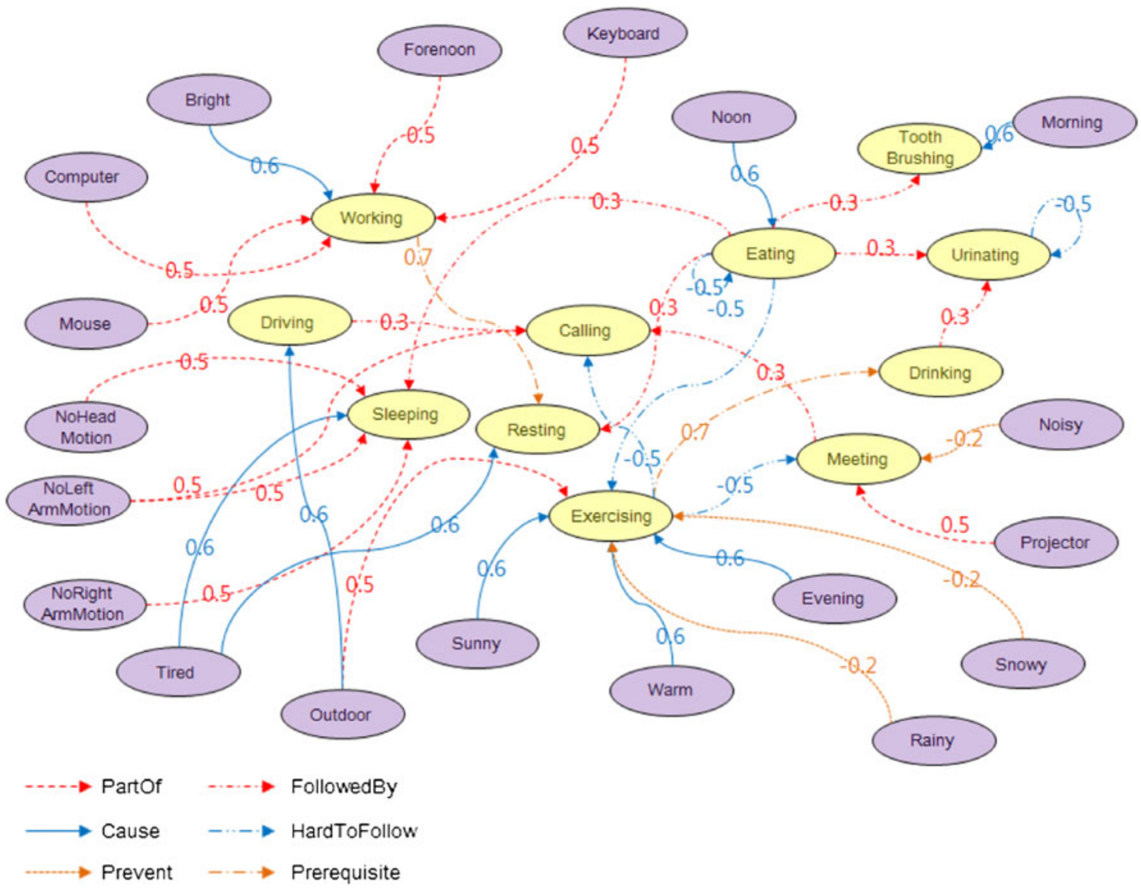

Fig. 10 Semantic network constructed with behaviors and sensing information

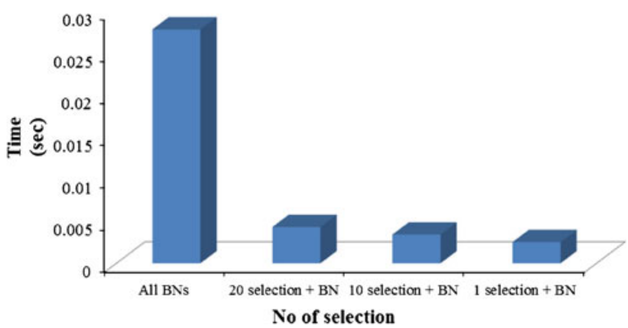

Fig. 11 The time required to perform activity inference

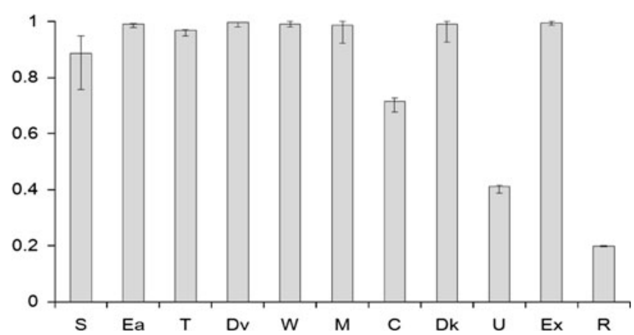

Fig. 12 Precision of inferred activities

The test requested him to do some activities which were recognizable in this system. The data set of the test includes the changes in the sensor values for each activity. The data set can be downloaded and tested in the author's homepage (http://sclab.yonsei.ac.kr/resources/datasets/multimodal.xlsx).

5 Concluding remarks

For the multi-modal context-aware system with multiple wearable sensors, the effective model construction and the efficient inference algorithm for the system are essential. This paper presents a system using an evolutionary process to train the context models and a selective inference to reduce the cost of inference. A genetic algorithm is used to optimize the structures of the DBN models. The selective inference using semantic network provides a method to reduce the cost of activity recognition by selecting an appropriate DBN model effectively. The proposed system was evaluated with real data. The result of the experiments shows the superiority of the method.

Table 7 Confusion matrix for user activity inference

	S	Ea	T	Dv	W	M	C	Dk	U	Ex	R	F (Failed)
Sleeping (S)	468	0	2	0	0	0	0	0	0	0	0	5
Eating (Ea)	0	396	0	0	0	0	90	0	104	0	0	118
TeethBrushing (T)	0	3	99	0	0	0	0	0	0	0	0	213
Driving (Dv)	0	0	1	329	0	0	0	0	1	0	4	191
Working (W)	0	0	0	1	1,180	0	0	0	30	0	0	722
Meeting (M)	0	0	0	0	0	269	0	0	0	0	0	45
Calling (C)	0	0	0	0	0	0	241	0	31	0	0	159
Drinking (Dk)	0	0	0	0	0	0	0	6	0	0	0	84
Urinating (U)	0	0	0	0	0	0	0	0	120	0	0	187
Exercising (Ex)	0	0	0	0	0	0	0	0	0	135	0	163
Resting (R)	25	0	0	0	0	0	0	0	3	0	1	77

In further studies, data collection with larger number of participants is necessary to evaluate the proposed system in general cases. Since users have different behavioral patterns for each other, we will also consider their variations to build the realistic semantic models as the future work. Furthermore, a comprehensive comparative study should be conducted with the relevant methods mentioned at the related works.

Acknowledgments The authors would like to thank Dr. J.-K. Min, Mr. S.-I. Yang and Dr. J.-H. Hong for their help to implement the system used in this paper. This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program.

References

1. Lester J, Choudhury T, Kern N, Borriello G, Hannaford B (2005) A hybrid discriminative/generative approach for modeling human activities. Paper presented at the proceedings of the 19th international joint conference on artificial intelligence, Edinburgh, Scotland
2. Choudhury T, Borriello G, Consolvo S, Haehnel D, Harrison B, Hemingway B, Hightower J, Klasnja PP, Koscher K, LaMarca A, Landay JA, LeGrand L, Lester J, Rahimi A, Rea A, Wyatt D (2008) The mobile sensing platform: an embedded activity recognition system. Pervasive Comput IEEE 7(2):32–41. doi:10.1109/MPRV.2008.39
3. Ogris G, Stiefmeier T, Lukowicz P, Troster G (2008) Using a complex multi-modal on-body sensor system for activity spotting. Paper presented at the proceedings of the 2008 12th IEEE international symposium on wearable computers
4. Roy N, Gu T, Das SK (2010) Supporting pervasive computing applications with active context fusion and semantic context delivery. Pervasive Mobile Comput 6(1):21–42. doi:10.1016/j.pmcj.2009.10.002
5. Van Laerhoven K, Aidoo KA, Lowette S (2001) Real-time analysis of data from many sensors with neural networks. In: Proceedings of the fifth international symposium on wearable computers, 2001, pp 115–122. doi:10.1109/iswc.2001.962112
6. Wang Y, Lin J, Annavaram M, Jacobson QA, Hong J, Krishnamachari B, Sadeh N (2009) A framework of energy efficient mobile sensing for automatic user state recognition. Paper presented at the proceedings of the 7th international conference on mobile systems, applications, and services, Kraków, Poland
7. Krause A, Ihmig M, Rankin E, Leong D, Gupta S, Siewiorek D, Smailagic A, Deisher M, Sengupta U (2005) Trading off prediction accuracy and power consumption for context-aware wearable computing. Paper presented at the proceedings of the ninth IEEE international symposium on wearable computers
8. Murao K, Terada T, Takegawa Y, Nishio S (2008) A context-aware system that changes sensor combinations considering energy consumption. Paper presented at the proceedings of the 6th international conference on pervasive computing, Sydney, Australia
9. Hwang K-S, Cho S-B (2009) Landmark detection from mobile life log using a modular Bayesian network model. Expert Syst Appl 36(10):12065–12076. doi:10.1016/j.eswa.2009.03.002
10. Hong J-H, Yang S-I, Cho S-B (2010) ConaMSN: a context-aware messenger using dynamic Bayesian networks with wearable sensors. Expert Syst Appl 37(6):4680–4686. doi:10.1016/j.eswa.2009.12.040
11. Bao L, Intille S (2004) Activity recognition from user-annotated acceleration data. doi:http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.68.5133
12. Brezmes T, Gorricho J-L, Cotrina J (2009) Activity recognition from accelerometer data on a mobile phone. Paper presented at the proceedings of the 10th international work-conference on artificial neural networks: part II: distributed computing, artificial intelligence, bioinformatics, soft computing, and ambient assisted living, Salamanca, Spain
13. Huỳnh T, Blanke U, Schiele B (2007) Scalable recognition of daily activities with wearable sensors. Paper presented at the Proceedings of the 3rd international conference on Location-and context-awareness, Oberpfaffenhofen, Germany
14. Jatoba LC, Grossmann U, Kunze C, Ottenbacher J, Stork W (2008) Context-aware mobile health monitoring: evaluation of different pattern recognition methods for classification of physical activity. In: 30th annual international conference of the IEEE on Engineering in Medicine and Biology Society. EMBS 2008, 20–25 August 2008, pp 5250–5253. doi:10.1109/iembs.2008.4650398
15. Godfrey A, Conway R, Meagher D, ÓLaighin G (2008) Direct measurement of human movement by accelerometry. Med Eng Phys 30(10):1364–1386. doi:10.1016/j.medengphy.2008.09.005
16. Blum M, Pentland A, Troster G (2006) InSense: interest-based life logging. Multimedia IEEE 13(4):40–48. doi:10.1109/mmul.2006.87
17. Maurer U, Rowe A, Smailagic A, Siewiorek D (2006) Location and activity recognition using ewatch: a wearable sensor

platform. In: Yang C, Julio A (eds) Ambient intelligence in everyday Life. Springer, Berlin, pp 86–102

18. Oliver N, Garg A, Horvitz E (2004) Layered representations for learning and inferring office activity from multiple sensory channels. Comput Vis Image Underst 96(2):163–180. doi:10.1016/j.cviu.2004.02.004
19. Harrison BL, Consolvo S, Choudhury T (2010) Using multimodal sensing for human activity modeling in the real world handbook of ambient intelligence and smart environments. In: Nakashima H, Aghajan H, Augusto JC (eds), Springer, Berlin, pp 463–478. doi:10.1007/978-0-387-93808-0_17
20. Ermes M, Parkka J, Mantyjarvi J, Korhonen I (2008) Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. IEEE Trans Inf Technol Biomed 12(1):20–26. doi:10.1109/titb.2007.899496
21. Oliver N, Horvitz E (2005) S-SEER: selective perception in a multimodal office activity recognition system. Paper presented at the proceedings of the first international conference on machine learning for multimodal interaction, Martigny, Switzerland
22. Schmidt A, Aidoo KA, Takaluoma A, Tuomela U, Laerhoven KV, Velde WVd (1999) Advanced Interaction in Context. Paper presented at the proceedings of the 1st international symposium on handheld and ubiquitous computing, Karlsruhe, Germany
23. Ming L, Rozgic V, Thatte G, Sangwon L, Emken A, Annavaram M, Mitra U, Spruijt-Metz D, Narayanan S (2010) Multimodal physical activity recognition by fusing temporal and cepstral information. IEEE Trans Neural Syst Rehabil Eng 18(4): 369–380. doi:10.1109/tnsre.2010.2053217
24. Marinov M, Zheliazkova I (2005) An interactive tool based on priority semantic networks. Know Based Syst 18(2–3):71–77. doi:10.1016/j.knosys.2004.08.001
25. Kofod-Petersen A, Cassens J (2006) Using activity theory to model context awareness. In: Second international workshop on modeling and retrieval of context. MRC 2005, Edinburgh, July 31–August 1, pp 1–17. doi:10.1007/11740674_1
26. Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning, 1st edn. Addison-Wesley Longman Publishing Co., Inc., Boston
27. Ling CX, Sheng VS (2008) Cost-sensitive learning and the class imbalance problem. Encyclopedia of machine learning. Springer, Berlin
28. Thai-Nghe N, Gantner Z, Schmidt-Thieme L (2010) Cost-sensitive learning methods for imbalanced data. In: The 2010 international joint conference on neural networks (IJCNN), 18–23 July 2010, pp 1–8. doi:10.1109/ijcnn.2010.5596486
29. Tapia EM (2008) Using machine learning for real-time activity recognition and estimation of energy expenditure. Massachusetts Institute of Technology, Cambridge