Logical Activity Recognition for Understanding of Human Activities

Sun-Sook Lee and Hwan-Seung Yong
Department of Computer Science and Engineering,
Ewha Womans University, 120-750, Seoul, South Korea
caffuchino@ewhain.net, hsyong@ewha.ac.kr

Abstract

Recognizing human activities based on user's contextual information is becoming more important to the success of context-aware pervasive computing. In this paper, we define activities and introduce a six-layer activity model. We also describe our algorithm to perform activity recognition. The key contextual information is location, and other pieces of contextual information are also used in our study. To evaluate our algorithm we collected user data from twenty test users while they performed daily activities. We present a comprehensive experimental evaluation and discuss the results.

Keywords: *Activity recognition, activity model, logical activity*

1. Introduction

Over the past few years, activity recognition has been gaining increasing attention in computer vision, artificial intelligence, sensor networks, and pervasive computing. Most existing works are mainly based on wearable sensors. Attaching sensors to the body or object is a promising and inexpensive technique to acquire data about certain types of human movement. However, most wearable sensors are not applicable in real world applications.

The recent advances in mobile communication devices have garnered much interest in the activity recognition and context recognition areas due to their cheap diverse and powerful incorporated sensors. These sensors include GPS, cameras, microphones, digital compasses, gyroscopes and accelerometers. In particular, smart phones can be used as already deployed sensors to sense, process, store, and transfer context data [1]. Some of the current approaches in activity recognition use a variety of these different sensors to recognize user's activity and the context [2, 3]. Majority of these work have been focused on recognition of physical activities such as walking, sitting, and running.

In our earlier work, we classified user activities into six different types according to their recognition complexity, and developed an activity model to recognize user activity (logical activity) based on the user's current context [4].

In this paper, we present a new algorithm to recognize which logical activity is being performed that is dependent upon user context data such as user profile, user location, user's physical activity, and time of day acquired from mobile phones as such information would suggest what the users are doing. We also describe our datasets, and evaluate our algorithm. Finally, we conclude the paper with future work.

2. Related Work

Activity recognition has recently gained increasing attention as a research topic because of the increasing availability of sensors in mobile devices. Much of activity recognition work is

focused on recognition of physical activity using body worn acceleration sensors. Accelerometers are currently among the most widely studied wearable sensors for activity recognition. The studies using accelerometers for monitoring Activities of Daily Living (ADLs) have shown that it is possible to discriminate between a range of user activities [5, 6]. Lee and Mase [7] proposed a new method for detecting and classifying a person's activity using a combination of a biaxial accelerometer, digital compass, and gyroscope. Their classification technique was based on a fuzzy-logic reasoning method.

With the increasing popularity of smart phones, much of the recent work uses collected data from smart phones to recognize user activity. Kwapisz, Weiss, and Moore [8] used the accelerometer of Android phones to collect accelerometer data and infer the type of activity such as sitting, standing, walking, jogging, and climbing sitting.

Additional studies have focused on inferring human activities from raw GPS data. This work includes an approach based on supervising learning to automatically infer different types of user activity from collected GPS logs [9] and extracting a person's activities and significant places from traces of GPS data using hierarchical conditional random fields [10].

Other approaches take contextual information into account in order to detect activities. The studies using not only data acquired from body-worn sensors but also environmental data from other sensors and user's surrounding environment have been reviewed in [11]

3. Logical Activity Recognition

Recognition of human activities is not simple as the number of activities is very large, and human activities are complex, and varied among individuals. In order to create an efficient algorithm that recognizes activities of daily living, we defined type of 'activity', classified this into six different types, and developed a computational model for defined activities.

3.1. Defining and Modeling Logical Activity

We classified user activity into six different types according to recognition complexity. Figure 1 shows the six-layer activity model we proposed [4]. Gesture and Action are defined as human movement using body parts such as moving the hand down and lifting a left leg while Activity consists of Actions such as taking classes, eating, and listening to music. Action consists of Gestures.
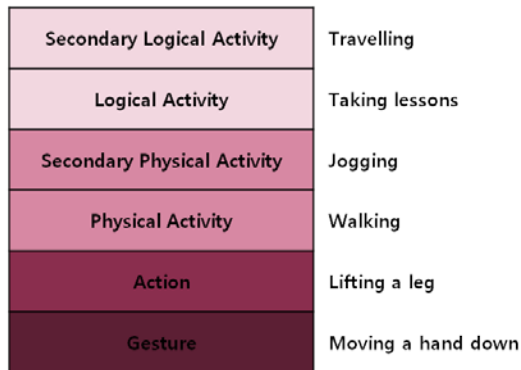

Secondary Logical Activity	Travelling
Logical Activity	Taking lessons
Secondary Physical Activity	Jogging
Physical Activity	Walking
Action	Lifting a leg
Gesture	Moving a hand down

Figure 1. Six-layer Activity Model

Activity is classified to physical activity and logical activity. Physical activity is a simple activity that is composed of only Actions, whereas logical activity involves complex activities

using five senses of human bodies or more interactions with the environment such as tools and other people. For example, cooking is a logical activity as it requires cooking utensils, and calling is also a logical activity as it needs a phone and the interaction from another person to enable the call.

Activity can be identified by the temporal distribution of activities of an individual with contextual information such as location, time duration, a series of actions, user profiles, personal backgrounds and occupation. Thus, we defined secondary logical activity that is the more complex and higher level logical activity, and can be inferred from lower level logical activities in process of time. Secondary physical activity also can be inferred from lower level physical activities in process of time.

For example, an office worker stays in a meeting room for 2 hours as shown in Figure 2. It is reasonable to assume that he/she is attending a business meeting. During the meeting, physical activities such as sitting, standing, or writing can be conducted. After certain amount of time pass, working as a secondary logical activity is recognized as working, involves lower level activities including lunch, business meeting, time at office, etc shown in Figure 2. Therefore, it is possible to recognize logical activities currently being performed based on the contextual information including lower level logical activities, location, time duration, and personal information.
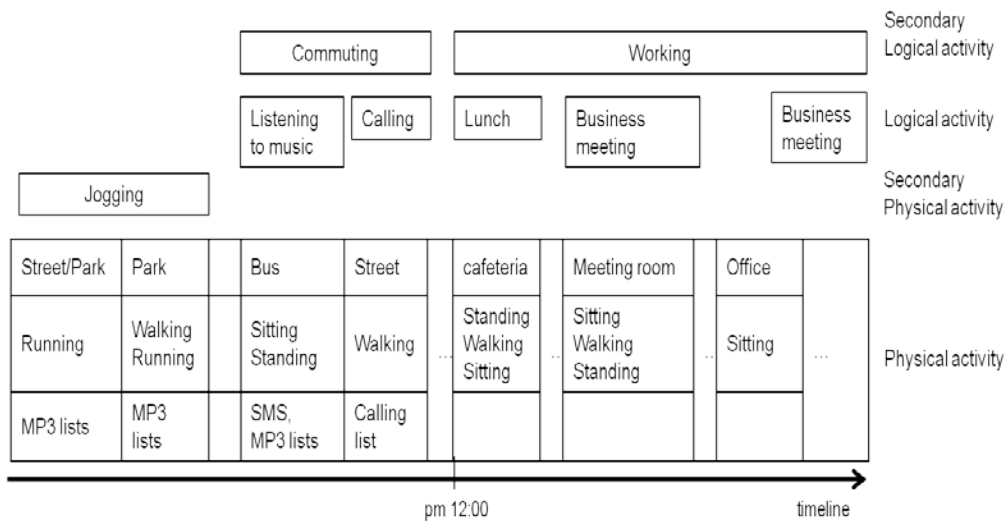

Figure 2. An Example Timeline of an Office Worker

Logical activity is modeled as sequences of user context by characterizing what a user is doing with three components: 1. Location, 2. Time duration, and 3. Contextual features.

First, location information is not user's GPS data in our study. We define location information as a result set from GIS which maps where a user is to a name of the place such as classroom and kitchen. We list these names from WordNet database [12] and refer to GeoWord. GeoWord is location-related words with Artifact domain from WordNet3.0. Table 1 shows examples of GeoWord. The GIS mentioned above is our ongoing research study, and as this paper focuses on the recognition of the user activity, we are not going to discuss GIS in this paper.

Second, time duration is duration of activity time. Activity is recognized when all the important information needed meets the necessary and sufficient conditions for an activity.

Otherwise activity recognition task fails. Some logical activities can be recognized only if they are conducted during certain time thresholds.

Third, contextual features include a series of physical activity, user profile, schedule, SMS history, and MP3 playlist, *etc*.

Table 1. Example of GeoWord

Category	GeoWord Example
Structure/Construction	hotel, toilet, classroom, university, home, bookshop
Way	tunnel, escalator, moving stairway, sidewalk, street
Facility	bus station, zoo, gas station, airport
Transport/Conveyance	train, airplane, car, bus, taxi
Location	car park, farm, lab, field, rice paddy
Furniture	sofa, bed

3.2. Logical Activity Recognition Algorithm

The goal of the activity recognition algorithm is to discover what the user is doing (logical activity). In this study, we consider 10 activities: Taking classes, Doing homework, Commuting, Sleeping, Eating, Cooking, Watching TV, Reading, Socializing, and Exercising.

They are selected because they are performed regularly by many people in everyday life. We felt the majority of logical activities were highly dependent on the location where they were being performed.

Thus we first created a matrix of potential logical activities for location with weight value (w) shown in Table 2 that indicates how certainly logical activities are conducted at certain locations by analyzing training dataset which are mentioned in Section 4.

We also created the second matrix of potential physical activities for locations to indicate how closely correlated between potential physical activities and logical activities shown in Table 3.

Table 2. Example of Matrix of Potential Logical Activities for Location. Columns: 1=Taking Classes; 2= Exercising; 3=Commuting; 4=Eating; 5=Cooking. Value: 1=Certainly; 0.5=Likely; 0=Unlikely

weight value (w)	1	2	3	4	5
Classroom	1	0	0	0	0
Park	0	1	0.5	0.5	0
Library	0	0	0	0	0
Restaurant	0	0	0	1	1

Table 3. Example of Matrix of Potential Logical Activities for Location. Columns: 1=Taking Classes; 2= Exercising; 3=Commuting; 4=Eating; 5=Cooking. Value: 1=Certainly; 0.5=likely; 0=Unlikely

weight value (w)	1	2	3	4	5
Standing	1	1	1	1	1
Sitting	1	1	1	1	1
Walking	1	1	1	1	0.5
Running	0	1	1	0	0

Second, we calculate Logical Activity Candidate Scores (LACS) for each 10 logical activity while input locations are the same. It is composed of four elements:

− Related Location Score (RLS): This score is used to measure the close relationship between locations and logical activities based on the matrix shown in Table 2.
− Related Activity score (RAS): This score is used to measure the close relationship between physical activities and logical activities based on the weight value in Table 3.
− Discarded Activity score (DAS): This score is used to measure how many of the irrelevant input data are discarded for a particular activity.
− Average weight (aw): This calculates an average weight of the weight values (w) of input data.

We calculate the logical activity candidate score using a linear combination of the above three scores with the coefficients aw:

$$LACS = aw * RAS + aw * RLS - (1-aw) * DAS$$

Third, we define Minimum Time (MT) for logical activity which indicates the minimum time duration of certain logical activity being performed. For example, MT for taking classes is 1 hour. If the total duration of user data on which the predicted logical activity is recognized does not satisfy with I hour of MT, then the logical activity is not detected.

Finally, the algorithm recognizes the logical activity with the minimum value of LACS as the final logical activity.

Furthermore, the algorithm iteratively produces secondary physical activities and secondary logical activities based on the logical activity candidates with longer time scale.

4. Evaluation

In this section, we describe our experiments and then present experimental result with corresponding discussions.

In order to collect user data, we asked 20 test users that are aged between 20 and 40 years old for their everyday activities for a week. We had data from 5 female undergraduate students, 5 male undergraduate students, 5 female office workers, and 5 male office workers. The data collection was recorded with the user's name, start and end time of logical activity, label of the logical activity, and kind of the physical activities being performed while performing the logical activity if possible. This creates the data set that is used for training and testing.

We mapped location information to GeoWord and mapped physical activities given by test users into physical activities we define in our study.

User data we collected cannot be directly applied to our algorithm. Thus, we first have to transform the user data into the format for our activity model. To accomplish this we divided the data into 30-second segments and then generated components within each 30-second segment. We chose a 30-second interval as we assumed that it provided sufficient time to exclude several repetitive physical activities that might not influence the recognition task while a user performs a logical activity after we compared the results for a 15-second, 30-second, 45-second, 60-second interval.

In order to evaluate the results, we consider precision of the logical recognition algorithm and the average error rates. Precision refers to number of matches between the recognition correctness inferred by the algorithm and the actual logical activity data given by the user.

Table 4 shows the confusion matrix of one experiment (the pink bar in Figure 3(a)). Figure 3(a) specifies the recognition correctness (precision) associated with each of the logical activities.

Table 4. Confusion matrix of an experiment. (A: Taking Classes B: Doing Homework C: Commuting, D: Sleeping, E: Eating, F: Cooking, G: Watching TV, H: Reading, I: Socializing, J: Exercising)

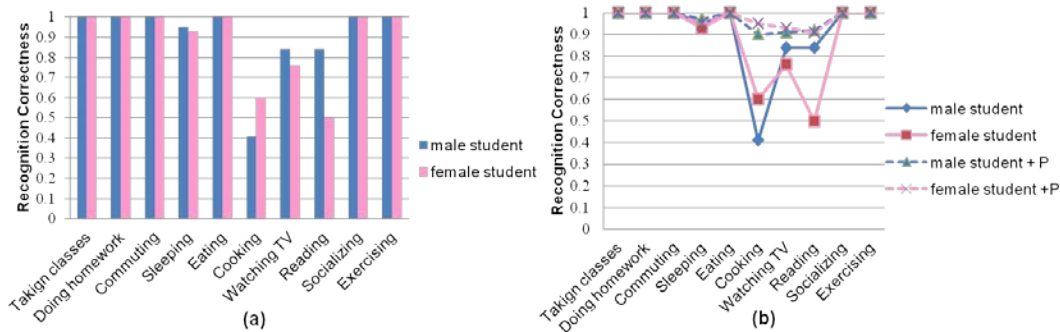|		Predicted Logical Activity									
		A	B	C	D	E	F	G	H	I	J
	A	1800	0	0	0	0	0	0	0	0	0
	B	0	120	0	0	0	0	120	30	0	0
	C	0	0	600	0	0	0	0	0	0	0
Actual	D	0	0	0	2520	0	0	0	0	0	0
Logical	E	0	0	0	0	540	60	0	0	0	0
Activity	F	0	0	0	0	0	0	0	0	0	0
	G	0	0	0	180	0	0	660	30	0	0
	H	0	0	0	0	0	0	0	60	0	0
	I	0	0	0	0	0	0	0	0	180	180

Figure 3. Recognition Correctness Comparison. (a) male and female undergraduate students. (b) comparison after applying user preference feature (P)

The algorithm performs well, given that it is difficult to distinguish between different logical activities that can be conducted in the same location with similar probability. For example, it seems unlikely to distinguish between Cooking and Eating based on the contextual information only. Because both activities can be conducted at kitchen, performing similar physical activities such as Standing and Sitting. This type of activities includes Cooking-Eating and Reading-Watching TV.

The results also indicate that majority of recognition errors are due to confusion between these activities. To limit the confusion between these types of errors, we ran another set of experiments where we add personal preference information to our algorithm. For example when a user prefers Reading to Watching TV, the algorithm applies this preference to the logical activity scores to recognize activities. The resulting recognition correctness is shown in Figure 3(b). Recognition is improved particularly on Cooking, Eating, Watching TV, and Reading, although these activities are still the most difficult to recognize. Because the components used to recognize logical activity are not sufficient to characterize these activities. This problem suggests additional input information be provided.

5. Conclusions

In this paper, we study the activity recognition to detect logical activity. We define this activity and classify into 6 types according to the relative recognition complexity. We also describe an algorithm to recognize logical activities by calculating logical activity candidate scores based on user's contextual information. We further show that activity recognition can be highly accurate.

As a next step, we have been collecting datasets from smart phones for the more efficient algorithm that recognizes more logical activities with better recognition correctness. Also, we realize that it is a difficult task to distinguish between different logical activities that can be conducted on the same location with similar probability.

In order to resolve the problem and improve recognition quality, additional input information such as nearby object information and common sense knowledge will be used in our future work.

Acknowledgements

References

[1] M. Raento, A. Oulasvirta, R. Petit and H. Toivone, "ContextPhone: A Prototyping Platform for Context-Aware Mobile Applications", IEEE Pervasive Computing **(2005)**, Vol.4, No.2, pp.51-59.
[2] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury and A. T. Campbell, "A survey of mobile phone sensing", IEEE Communications Magazine, **(2010)**, pp. 140-150.
[3] T. Choudhury, M. Philipose, D. Wyatt and J. Lester, "Towards Activity Databases: Using Sensors and Statistical Models to Summarize People's Lives", IEEE Data Eng. Bull, vol. 29, **(2006)**, pp.49-58.
[4] S. S. Lee and H. S. Yong, "Logical Activity Modeling for Logical Activity Recognition", Proceedings of 8th International Conference on Multimedia Information Technology, **(2012)**, pp. 26-28.
[5] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola and I. Korhonen, "Activity Classification Using Realistic Data From Wearable Sensors", IEEE Transactions on inf. Tech. in Biomedicine, vol. 10, **(2006)**, pp. 119-128.
[6] N. Ravi, N. Dandekar, P. Mysore and M. L. Littman, "Activity recognition from accelerometer data", Proceedings of 17th Conference on Innovative Applications of Artificial Intelligence, **(2005)**.
[7] S. W. Lee and K. Mase, "Activity and Location Recognition Using Wearable Sensors", IEEE Pervasive Computing, vol. 1, **(2002)**, pp. 24-32.
[8] J. R. Kwapisz, G. M. Weiss and S. A. Moore, "Activity Recognition Using Cell Phone Accelerometers", SIGKDD Explor, Newsl, vol. 12, **(2011)**, pp. 74-82.
[9] Y. Zheng, Y. Chen, Q. Li, X. Xie and W. Y. Ma, "Understanding Transportation Modes Based on GPS Data for Web Applications", ACM Trans, Web, vol. 4, **(2010)**, pp. 1-36.
[10] L. Liao, D. Fox and H. Kautz, "Extracting places and activities from gps traces using hierarchical conditional random fields", International Journal of Robotics, vol. 26, **(2007)**, pp. 119-134.
[11] M. Baldauf, S. Dustdar and F. Rosenberg, "A survey on context-aware systems", International Journal Ad Hoc and Ubiquitous Computing, vol. 2, **(2007)**, pp. 263-277.
[12] WordNet. Princeton University, http://wordnet.princeton.edu.

Authors

Sun-Sook Lee received the M.S degree in engineering in 2008 from Ewha Womans University, Seoul, South Korea. Currently, she is a Ph.D. student of Database Laboratory of Department of Computer Science and Engineering at Ewha Womans University. Her research interests include information retrieval, data mining, ontology engineering and semantic web technologies.

Hwan-Seung Yong is a Professor in Department of Computer Science and Engineering at Ewha Womans University in Seoul, South Korea. He obtained his Ph.D degree in database system from Seoul National University in 1994. His main research interests include data mining, ubiquitous computing, and big data management.