DOI: 10.2298/CSIS121101033H

Activity Inference for Constructing User Intention Model

Myunggwon Hwang, *Do-Heon Jeong, Jinhyung Kim, Sa-kwang Song, and Hanmin Jung

Korea Institute of Science and Technology Information (KISTI)
Daejeon, Republic of Korea
{mgh, heon, jinhyung, esmallj, jhm}@kisti.re.kr
*Corresponding Author

Abstract. User intention modeling is a key component for providing appropriate services within ubiquitous and pervasive computing environments. Intention modeling should be concentrated on inferring user activities based on the objects a user approaches or touches. In order to support this kind of modeling, we propose the creation of object–activity pairs based on relatedness in a general domain. In this paper, we show our method for achieving this and evaluate its effectiveness.

Keywords: Ubiquitous and pervasive computing, context awareness, user intention modeling, lexical cohesion.

1. Introduction

Ubiquitous and pervasive computing (UPC) is a post-desktop model of human–computer interaction in which information processing has been thoroughly integrated into everyday objects and activities[1]. The fundamental basis of UPC is context awareness, which makes it possible for computers to both sense and react to user behaviors based on the user's environment. By context we mean "any information that can be used to characterize the situation of entities (i.e., whether a person, place, or object) that are considered relevant to the interaction between a user and an application, including the user and the application themselves [1]." The aim of UPC is to understand the context, grasp the user's intention, and provide suitable services in proper time. Moreover, understanding the context means to know where the user is and what the user does. To this end, many studies have dealt with user intention modeling to construct rules or relatedness between contexts and user intentions.

Various methods of user intention modeling have been proposed and been shown to perform well under specific conditions [2, 3, 4, 5, 6]. Unfortunately,

[1] Ubiquitous computing - Wikipedia, the free encyclopedia: http://en.wikipedia.org/wiki/Ubiquitous_computing, 22 Sep., 2012.

good performance has been largely domain-bound and dependent on narrow and/or artificial assumptions about intended actions.

User intention modeling must contend with the complexity of *n*:*m* mapping between contexts and intentions. For example, when a subject uses a computer at his desk in an office, a UPC system cannot infer his exact intentions, because the context is related to many intentions. The focus of this study is to improve the effectiveness of multiple mappings for user intention modeling. Assuming that most user intentions involve engagement of an object, suitable services can be provided if the UPC system can determine the activities related to the various objects the user approaches and touches. In a previous paper [7], we proposed a method to overcome the limitations faced in earlier works. In this paper, we employ similar techniques to support more flexible and reliable modeling of user intentions, based on the lexical cohesion (similarity measurement) between nouns and verbs.

Text strings are used to represent all tangible and intangible objects (nouns) as well as human activities (verbs). To collect human activities appropriate for given objects, we employ WordNet, Google n-gram data, and the Dice coefficient. Specifically, WordNet is used for selecting verbs to describe main activities, while Google n-gram data, a massive corpus of collective intelligence, is used to calculate lexical cohesions. Our evaluation of this method suggests that it can provide great convenience in preparing object–activity pairs, through processing of collective intelligence data.

This paper[2] is organized as follows. Section 2 describes work related to our research. In section 3, we explain our proposed similarity measurement between objects and activities. In section 4, we evaluate our method and assess its contribution. Finally, in section 5, we summarize our findings.

2. Related Works

The main of UPC is to detect user intentions accurately and to provide services appropriate to these intentions. Assuming that understanding user intention is essential, research on UPC must correlate strongly with that on human–computer interaction (HCI) and Ambient Intelligence [17]. In addition, to understand user intention, it is mandatory to perform interrelation modeling between user activities and surrounding objects. User intention can vary according to places and objects. Various methods have been proposed to recognize place- and object-dependent intentions, such as data mining methods based on a large corpus [8], machine learning methods [9, 10], ontology-driven methods [11], and information retrieval-based methods [7]. We extended some of the information retrieval-based techniques and applied them to a domain of 17 different activities common to an office area In the previous work [7]. We measured similarities between these intentions and

[2] This paper is an extension of [15] presented at the IMIS 2012 conference. It contains additional content such as more examples for easy understanding and new experimental results with abundant test sets.

surrounding objects based on the n-gram dataset from Google using Bayesian probability. Figures 1 and 2 provide an example output from the application we have developed.

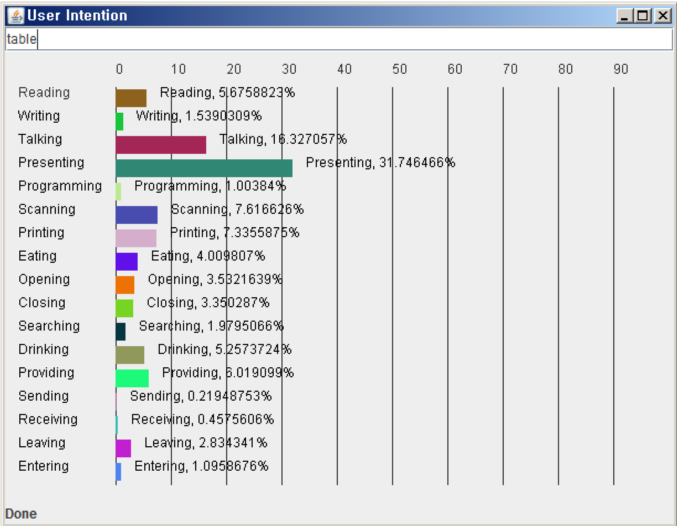

Fig. 1. An application for user intention modeling (object: ‘table’)

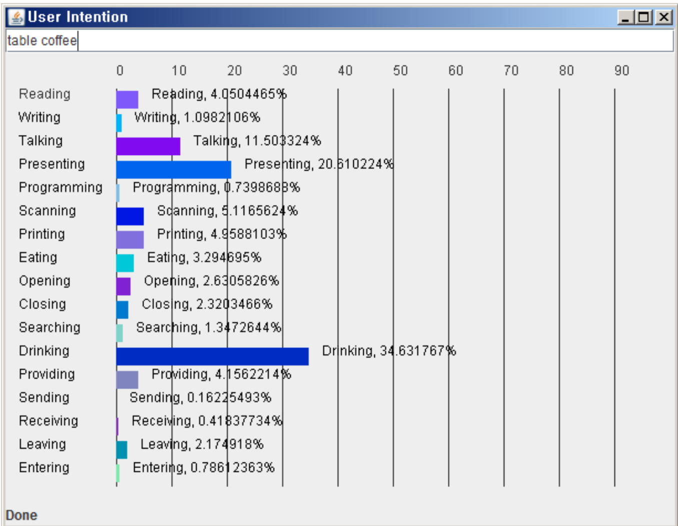

Fig. 2. An application for user intention modeling (objects: ‘table’ and ‘coffee’)

Figure 1 shows the relatedness between the single object ‘table’ and each of our 17 activities. The graphed output shows that when a user approaches or touches a table, a UPC system can expect certain activities more than others, namely, ‘presenting,’ ‘talking,’ ‘reading,’ ‘providing,’ ‘scanning,’ and so on, in order of relatedness. Figure 2 shows the results when two objects, ‘table’ and ‘coffee,’ are approached at the same time. The addition of the ‘coffee’ object to the ‘table’ object provides a further clue for assessing the relatedness of activities, so that in this case the application shows ‘drinking’ as the most likely user intention. Based on this intention, a UPC system can provide appropriate services related to ‘water’ and ‘cup,’ for example.

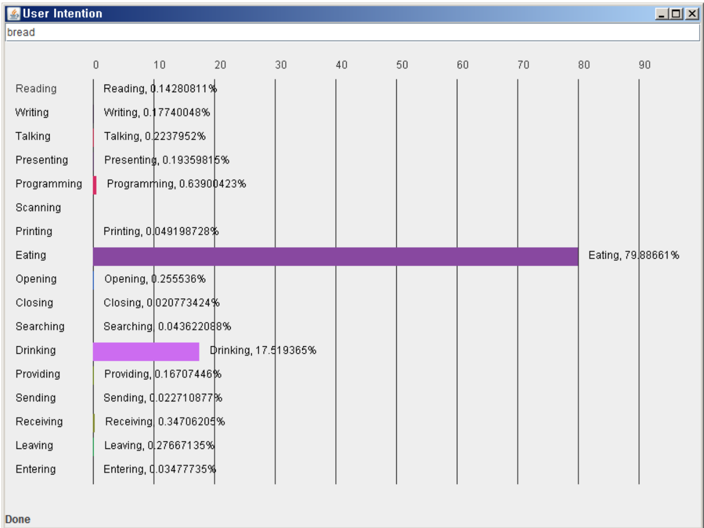

Fig. 3. An application for user intention modeling (object: ‘bread’)

Figure 3 shows a case in which a single approached object, ‘bread,’ relates not only to the likely activity of ‘eating’ but also to the associated activity of ‘drinking’. Thus, a UPC system can expect that someone with a piece of bread likely intends to eat the bread and possibly drink something soon after. Based on results from the previous work, we have found that these kinds of linkages between objects near the user and possible activities involving those objects can be obtained through analysis of large sets of data. However, the work restricted the domain of user intentions to 17 commonplace activities. Therefore, we concentrate on making unrestricted matrix between activities and objects in this research.

3. Inferring Human Activities

At the core of our system, and of future UPC environments, there must be a base of fundamental data for user intention modeling. Figure 4 illustrates the proposed architecture for context-aware applications in UPC. The components presented in this paper are shown in grey.

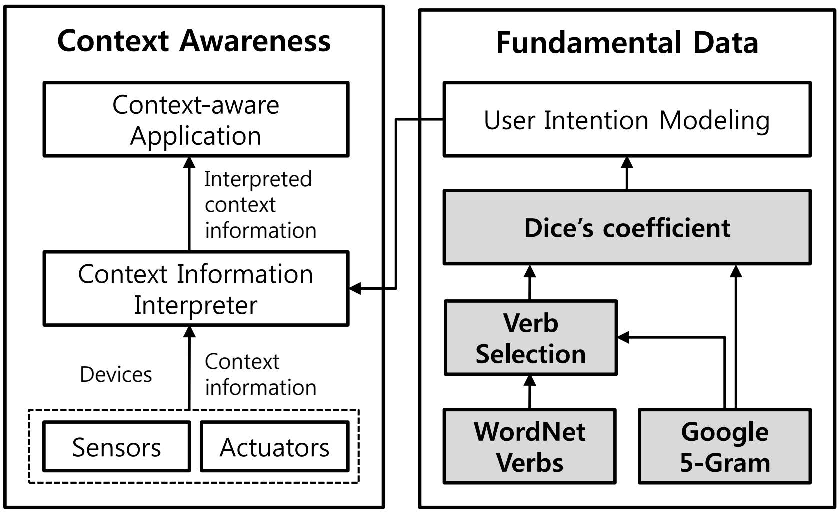

Fig. 4. Architecture for proposed context awareness

3.1. Data Resources

As mentioned in section 1, we used two primary data resources: WordNet and N-Gram. Here we describe these resources in greater detail.

WordNet. We used WordNet[3] to select verbs that describe the human activities needed for modeling most user intentions. Developed at Princeton University, WordNet is a lexical database of English based on psycholinguistics. It has been continuously expanded since 1985 [12, 13] and contains nouns, verbs, adverbs, adjectives, and their definitions, along with semantic networks and subject categories. Though our original work in this area [15] used both nouns and verbs from WordNet, the system proposed here uses only the verbs describing major human activities. This was done because such human activities form a relatively stable set compared to the set of nouns, which is constantly expanded by new products, new issues, social phenomena, etc [18, 19].

[3] WordNet (A lexical database for English): http://wordnet.princeton.edu/

Google N-Gram. We used the 5-gram data from Google n-gram as a broad corpus for similarity measurements between objects and activities. Because n-gram was made by the collective intelligence of people all over the world, we could be confident of a certain degree of objectivity to our measurements. Google n-gram provides token types ranging from a 'unigram' to a '5-gram' of constituent tokens extracted from a huge amount of documents on January, 2006 [14]. Some statistics and examples from N-Gram are listed in Table 1. The data is provided by LDC (Linguistic Data Consortium)[4] [7]. For our system, we chose the 5-gram as the best collection to represent lexical cohesions and specific contexts with semantics. We considered both the words in each token and the count of those words to be important factors for activity inference.

Table 1. Statistics and examples of Google n-gram

Token types	Number of tokens	Token examples (count)
Unigram	13,588,391	Mobile (94,162,727) Phone (151,683,102)
Bigram	314,843,401	Mobile phone (9,414,886) Smart phone (965,621)
Trigram	977,069,902	talking about phone (326) talking on phone (4,144)
4-gram	1,313,818,354	talking on mobile phone (3,888) talks with mobile phone (57)
5-gram	1,176,470,664	protection for your mobile phone (4,372) talking on mobile phone with (271)

3.2. Selection of Activities

WordNet contains 11,488 verbs in total. As a fundamental design principle, we wanted to employ all of these verbs in our system. However, we recognized that this amount of data would likely cause large processing delays. For this reason, we focused on a small subset of verbs describing common human activities. If certain verbs were used to describe human activities frequently, they were also expressed in text. Based on this assumption, we wrote a simple program to calculate the frequencies of verbs in a Google 5-gram. The basic steps of this program are as follows:

Step 1. All WordNet verbs are prepared. Each verb is used for

Step 2. In order to minimize word variances between WordNet and Google 5-gram, all the words are stemmed by Porter stemmer[5].

Step 3. Stop words (including 'be' verbs) are removed.

Step 4. Document frequencies (*df*) are calculated based on the total counts of tokens for each verb. For example, in table 1, the verb 'talk'

[4] LDC (Linguistic Data Consortium): http://www.ldc.upenn.edu/

[5] Porter Stemming Algorithm: http://tartarus.org/martin/PorterStemmer/

also implies ‘talking’ and ‘talks’ and thus yields a total count of 5 in the example data.

Step 5. Verbs are filtered by their *df*. We set a threshold value for *df* at 10,000 to remove less common human activities. This filters out all but the most frequently used 20% of activities.

When executed, this program returned a result set of 2,727 WordNet verbs: $vs = \{v_i,\ 0 < i <= n\}$, where *vs*, *v*, and *n* indicate verb set, a verb, and |*vs*| (2,727 in this work).

3.3. Similarity Measurement

This section describes how we measure similarity as a basis for calculating relatedness between our set of human activities and world objects. An object can be related to various activities in different degrees. For example, the single object ‘table’ might relate to activities such as ‘presenting,’ ‘talking,’ ‘reading,’ and so forth (as in figure 1), whereas the two objects ‘table’ and ‘coffee’ together might relate more strongly to a smaller range of activities (as in figure 2). Google 5-gram contains almost all verbs and nouns, and provides occurrence counts for every token. All world objects and all human activities are expressed in nouns and verbs found in text documents, and if an object is related to an activity deeply, the pair will occur with greater frequency than others will. Therefore, we can enumerate the relatedness by measuring similarities (lexical cohesion) between nouns and verbs using Dice’s coefficient.

$$relatedness(v_i, n_j) = \frac{2 \times Occ(v_i \cap n_j)}{Occ(v_i) + Occ(n_j)} \tag{1}$$

where, v_i and n_j represent a verb and a noun, respectively, and *Occ* is the occurrence of the token. To illustrate, we provide a sample measurement of relatedness in table 2.

Table 2. Examples of similarity measurement for the noun ‘Book‘

v_i	$Occ(v_i)$	n_j	$Occ(n_j)$	$Occ(v_i \cap n_j)$	relatedness
Read	317,579,294	Book	541,966,689	5,216,567	0.01214
Search	721,112,089			4,532,558	0.00718
Print	177,593,953			1,921,587	0.00534
Write	170,529,264			810,037	0.00227

Table 2 shows examples of calculating the relatedness of the verbs ‘read,’ ‘search,’ ‘print,’ and ‘write’ to the noun ‘book.’ From these results, we can conclude that people generally intend to read when they access or touch a book. Based on the full set of such measurements, we constructed a relatedness matrix between 2,727 verbs and 755,312 nouns (the resulting file size exceeds 4 GB).

4. Evaluation

User intention modeling attempts to identify the activities in which a user might engage by analyzing the nearby objects in which he/she shows interest. To accomplish this, our research focuses on modeling user intention based on the relatedness of nearby nouns a set of verbs. This section provides some experimental data showing how reliably our system selects verbs appropriate to given nouns.

Figure 5 shows the output of an application we designed to evaluate reliability. The application receives one or more nouns as an input and returns the top 20 verbs according to relatedness. In the case of multiple nouns, the application calculates relatedness for each verb. The output verbs are then evaluated by their appropriateness to the input noun. Table 3 shows a sample evaluation of the top 20 verbs related to the single noun ‘door’ extracted from our system.

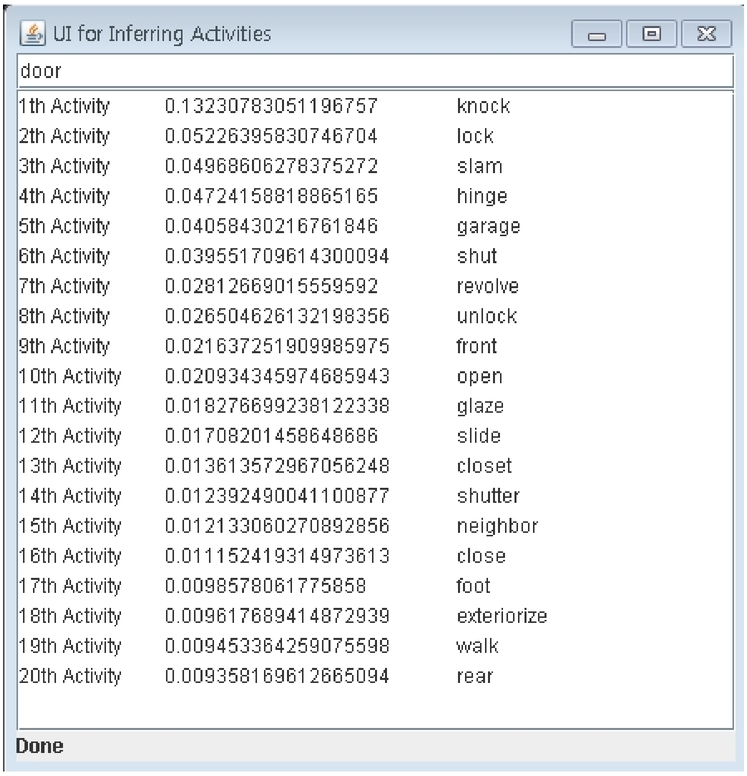

Fig. 5. Application output of the top 20 activities for the object ‘door’, according to the proposed system.

Table 3. Top 20 verbs for 'door' and their appropriateness (1 means appropriate and 0 means inappropriate).

Verbs	Eval.	Verbs	Eval.	Verbs	Eval.
Knock	1	Unlock	1	Neighbor	0
Lock	1	Front	0	Close	1
Slam	1	Open	1	Foot	0
Hinge	1	Glaze	0	Exteriorize	0
Garage	0	Slide	1	Walk	0
Shut	1	Closet	0	Rear	0
Revolve	0	shutter	0		

To evaluate our system for single noun inputs, we chose 200 nouns that represent tangible and intangible objects familiar in daily life. To evaluate our system for multiple noun inputs, we chose an additional 100 objects. Multiple noun input sets were created by combining a noun from our set of single noun inputs (e.g., 'computer java' and 'book bookshelf') with specific statuses or compound nouns (e.g., 'copy machine,' 'opened window' and 'table' + 'copy machine'). Table 4 provides a few examples.

Table 4. Examples of object selection

Count of clues	Examples of clues
Single object	Computer, printer, bookshelf, water, scanner, server, door, coffee, money, java, …
Multiple objects	Computer java, opened door, copy machine, table printer, coffee water, apache server, received email, … table copy machine,

To calculate the reliability of our verb output sets, we use equation (2):

$$reliability = \frac{cnt(appr.)}{cnt(appr.) + cnt(inappr.)} \quad (2)$$

Table 5 provides the results of our evaluation:

Table 5. Evaluation results on data reliability

	Evaluation for single nouns	Evaluation for multiple nouns
Appropriate	1,888	1,284
Inappropriate	2,112	716
Reliability (%)	47.2	64.2

From these results, it is clear that multiple noun inputs yield higher reliability than do single noun inputs by a difference of 17%. The reason for this is that single nouns have more ambiguity. For example, the noun 'window' has at least two possible meanings: (1) a tangible, transparent opening in a wall or door and (2) an intangible object offered by graphical user interfaces. Even if we use the branded, capitalized term 'Windows' to indicate the latter of these two meanings, the stemmer used by our system converts the term to the basic and ambiguous form 'window'. Another example is the term 'java', which may refer to coffee or to a popular programming language.

Such ambiguities have a negative impact on reliability and overall results for single noun inputs are somewhat discouraging. However, we believe the system can be significantly improved by limiting the domain of activities based on location. For example, if someone touches a car key in an office, his/her immediate intention is probably to leave the office, but if he/she touches a car key inside a parking garage, his/her intention is probably to unlock a car and drive away. Such place-bound improvements have already been demonstrated in previous work [7]. We believe that this kind of domain restriction can help to construct a far more useful relatedness matrix between our 2,727 verbs (main activities) and 755,312 world objects. We intend to make these improvements in the future and to base further applications on a more refined relatedness matrix.

5. Conclusion

Inferring human activities is essential to the future of UPC and supporting a well-structured user intention model is the natural starting point to make such inference possible and reliable. To this end, we have leveraged the collective intelligence represented in WordNet and Google 5-gram to find and measure similarities between objects and activities. To test the reliability of our inferences, we created an application and used 200 single objects and 100 multiple objects as test inputs. Even though the overall evaluation was somewhat not satisfied to our expectation, we were able to identify clear paths for improvement using place-bound restrictions of activity sets. We expect that the result is useful for preparing ground data for actual user intention modeling in order to choose wide-ranging and reliable object-activity pairs.

References

1. Dey, A. K. and Abowd, G. D.: Towards a Better Understanding of Context and Context Awareness. In Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing, pp. 304-307, 1999.
2. Gu, T., Pung, H. K. and Zhang, D. Q.: Toward and OSGi-Based Infrastructure for Context-Aware Applications. IEEE Pervasive Computing, Vol. 3(4), pp. 66-74, 2004.
3. Huang, P. C. and Kuo, Y. H.: A Reliable Context Model for Context-aware Application. Systems Man and Cybernetics, pp. 246-250, 2008.
4. Gui, F., Guillen, M., Rishe, N., Barreto, A., Andrian, J. and Adjouadi, M.: A Client-Server Architecture for Context-Aware Search Application. Network-Based Information Systems, pp. 539-546, 2009.

5. Gu, T., Pung, H. K. and Zhang, D. Q.: A Middleware for Building Context-Aware Mobile Services. Vehicular Technology Conference, Vol. 5, pp. 2656-2660, 2004.
6. Zhan, Y., Wang, S., Zhao, Z., Chen, C. and Ma, J.: A Mobile Device Oriented Framework for Context Information Management. Information, Computing and Telecommunication, pp. 150-153, 2009.
7. Hwang, M., Choi, D. and Kim, P.: Information Retrieval Techniques to Grasp User Intention in Pervasive Computing Environment. In Proceedings of International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, pp. 186-191, 2011.
8. Chen, Z., Lin, F., Liu, H., Liu, Y., Ma, W. Y. and Wenyin, L.: User Intention Modeling in Web Application Using Data Mining. World Wide Web: Internet and Web Information Systems, Vol. 5, No. 3, pp. 181-191, 2002.
9. Jansen, B. J., Booth, D. L. and Spink, A.: Determining the User Intent of Web Search Engine Queries. In Proceedings of the 16th international conference on World Wide Web, pp. 1149-1150, 2007.
10. Jung, S., Lee, C., Kim, K. and Lee, G. G.: Hybrid approach to user intention modeling for dialog simulation. In Proceedings of the ACL-IJCNLP Conference Short Papers, pp. 17-20, 2009.
11. Jeon, H., Kim, T. and Choi, J.: Ontology-based User Intention Recognition for Proactive Planning of Intelligent Robot Behavior. In Proceedings of International Conference on Multimedia and Ubiquitous Engineering, pp. 244-248, 2008.
12. Fellbaum, C.: WordNet: An Electronic Lexical Database, MIT Press.
13. Hwang, M., Choi, C. and Kim, P.: Automatic Enrichment Method of Semantic Relation Network and its Application to Word Sense Disambiguation. IEEE Transactions on Knowledge and Data Engineering, Vol. 23, No. 6, pp. 845-858, 2011.
14. Brants, T. and Franz, A.: Web 1T 5-gram Corpus Version 1.1 (LDC2006T13), April 2006.
15. Hwang, M., Jeong, D. H., Kim, J., Song, S. K. and Jung, H.: Similarity Measurement between Objects and Activities for User Intention Modeling. In Proceedings of 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, pp. 103-107, 2012.
16. Jung, J. J.: Contextualized mobile recommendation service based on interactive social network discovered from mobile users. Expert Systems with Applications, Vol. 36, No.9, pp. 11950-11956, 2009.
17. Augusto, J. C. and J. M. Paul.: Ambient Intelligence: Concepts and applications. Computer Science and Information Systems, Vol. 4, No. 1, pp. 1-27, 2007.
18. Jung, J. J.: Boosting Social Collaborations Based on Contextual Synchronization: An Empirical Study. Expert Systems with Applications, Vol. 38, No. 5, pp. 4809-4815, 2011.
19. Eagle, N. and Pentland, A.: Reality mining: Sensing Complex Social Systems. Personal and Ubiquitous Computing, Vol. 10, No. 4, pp. 255-268, 2006.

Myunggwon Hwang received the B.S. degree in Computer Engineering, the M.S. degree in Computer Science, and the Ph.D. degree in Computer Engineering from Chosun University. He is a senior researcher in the Department of Computer Intelligence Research at Korea Institute of Science and Technology Information (KISTI). His research focuses on Semantic Web Technologies, Semantic Information Processing and Retrieval, Word Sense Disambiguation and Knowledge Acquisition.

Do-Heon Jeong received the M.S. in Information Science from Yonsei University in 2003. He is in charge of analytics service development team in the Department of Computer Intelligence Research, Korea Institute of Science and Technology Information (KISTI). His research focuses on Informetrics, Semantic Web and Text Mining.

Jinhyung Kim received the M.S. degree and the Ph.D. degree in Computer Science from Korea University respectively. He is a postdoctoral researcher in the Department of Computer Intelligence Research at the Korea Institute of Science and Technology Information (KISTI). His research focuses on Semantic Web Technologies, Semantic Information Processing and Retrieval, Word Sense Disambiguation and Knowledge Acquisition.

Sa-kwang Song received his B.S. degree in Statistics in 1997 and his M.S. degree in Computer Science in 1999 at Chungnam National University, Korea. He received his Ph.D. degree in Computer Science at Korea Advanced Institute of Science and Technology (KAIST). He is currently a senior researcher in the Department of Computer Intelligence Research, Korea Institute of Science and Technology Information (KISTI). His research interest includes Text Mining, Natural Language Processing, Information Retrieval, Semantic Web, and Health-IT.

Hanmin Jung works as the head of the Department of Computer Intelligence Research and chief researcher at Korea Institute of Science and Technology Information (KISTI), Korea since 2004. He received his M.S. and Ph.D. degrees in Computer Science and Engineering from POSTECH, Korea in 1994 and 2003. His current research interests include technology intelligence based in the Semantic Web and text mining technologies, human-computer interaction (HCI), and natural language processing (NLP).

Received: September 01, 2012; Accepted: March 25, 2013