KeyGraph-based Social Network Generation for Mobile Context Sharing

Myeong-Chun Lee
Department of Computer Science
Yonsei University
Seoul, Republic of Korea
lmspring@sclab.yonsei.ac.kr

Young-Seol Lee
Department of Computer Science
Yonsei University
Seoul, Republic of Korea
tiras@sclab.yonsei.ac.kr

Sung-Bae Cho
Department of Computer Science
Yonsei University
Seoul, Republic of Korea
sbcho@cs.yonsei.ac.kr

Abstract—We propose a KeyGraph-based context sharing method in mobile environment. With the recent advancement of mobile sensors, a variety of mobile applications become vehicles for improving our lives. Context sharing system which shares the user behaviors, emotion, and location is one of the promising fields for the social network service. It is a difficult problem to determine whether a user will share the personal information or not. In typical social network models, users are grouped in communities, and nodes of the same community have strong social links between each other. However, some nodes also have social links outside their “home” community. They have social relationships with users of different groups. Most systems concentrate on generating internal “home” community regardless of outside social relation. In this paper, we classify the personal information into two types. First type is the information to be shared with “home” community only. Second type is the information to be shared with as many people as possible. We utilize KeyGraph algorithm to select a home community for sharing the personal contexts. KeyGraph extracts two types of people who have strong social relationships in a community and have social links with many different communities. In order to show the feasibility of the proposed method, we conduct experiments to extract the user communities from Bluetooth data and implement a real-time context sharing application.

Keywords—KeyGraph, Context-aware, Context sharing, Social Network

I. INTRODUCTION

Mobile context sharing service provides user’s high-level contextual information to their social group. Because of the availability of the mobile sensors, it provides more information than web-based social network systems. The shared personal information makes people closer and helps them meet each other easily. Before sharing the contextual information, it is necessary to generate the semantic information from mobile sensors. Context can be any information such as current location, behavior, etc [1]. With the recent advancement of mobile technology, a number of sensors (Light, Orientation, GPS, Accelerometer Bluetooth, etc) in a mobile device are equipped and those enable the context aware. GPS sensor can recognize the user location and accelerometer can recognize the user behaviors. We can also get more accurate and detailed information from the integration of sensor data.

Although various studies for mobile sharing method have been introduced, most of them only focused on generating “home” community [2][3][4][5][6][7][8][9]. When “home” communities are generated, they share context or content within “home” community. They do not share outside of “home” community because they already assume there is no social relationship between the communities. However, some cases have social relationship between communities. For instance, a friend group ‘A’ and another friend group ‘B’ could have a relationship if both are students in the same class. Sharing context between communities makes a system more flexible and precise. If context is shared between communities, the degree of context information could be also different since internal “home” community has closer relationship than relationship between communities.

This paper proposes a KeyGraph-based context sharing method. KeyGraph was originally used to extract keywords in a document [10]. It finds a keyword (event) using their linked co-occurrence relation. Even if the keyword is not frequent in a document, KeyGraph can find key information based on co-occurrence structure. Ohsawa applied the method to predict the earthquake [11]. Kim et al. proposed a KeyGraph-based content management system. They used KeyGraph to extract key events in integrated mobile information [12]. we use KeyGraph to generate social communities and to connect between relevant communities. Co-occurrence of Bluetooth logs between subjects generate the clusters which represent social groups. Key person links the communities, and inferred context is shared by defined criterion. Additionally, we implement a prototype application to infer a user’s behaviors and to share the information with users. A user can get context information of their social group members from a mobile phone in real-time.
The rest of the paper is organized as follows. In Section II, we review some related works for context sharing and content sharing. Since the objective of content sharing is similar to context sharing, we examine both of them. In Section III, we present details of the proposed method, and Section IV shows experiments to evaluate it. Finally, we summarize this paper and present future works in Section V.

### II. MOBILE CONTEXT SHARING

Context sharing includes context awareness as well as a method to share personal information. Context awareness is a process to recognize or infer the user situations from lower-level information. A user can acquire high-level information such as activities and emotions by high-level context inference process. Table I summarizes some related works about context sharing. The related works are categorized in terms of sharing method, system architecture, and shared information.

Before sharing the context, setting a standard between privacy and sharing is necessary. Christin et al. introduced a privacy bubble that shared pictures with other users [4]. It is created by the user and continued until a pre-defined certain time arrives. The contents can be accessed to people who are within the bubble. Park et al. introduced Bayesian network to define a relationship between users [5]. To infer the relationship, various variables such as private relation, work relation, contact relation, emotion relation, etc. are needed.

Sorathia et al. and Lee et al. shared the personal context by using the phonebook and synchronization [6, 7].

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Sharing method</th>
<th>System architecture</th>
<th>Sharing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. Christin et al.</td>
<td>2012</td>
<td>Privacy bubble</td>
<td>Application</td>
<td>Content</td>
</tr>
<tr>
<td>H. Park et al.</td>
<td>2011</td>
<td>Bayesian Network</td>
<td>Server</td>
<td>Context</td>
</tr>
<tr>
<td>K. Sorathia et al.</td>
<td>2009</td>
<td>Phonebook</td>
<td>Phone to Phone</td>
<td>Context, Content</td>
</tr>
<tr>
<td>J. Lee et al.</td>
<td>2009</td>
<td>SMS based Synchronization</td>
<td>Phone to Phone</td>
<td>Context</td>
</tr>
<tr>
<td>J. Yamamoto et al.</td>
<td>2009</td>
<td>Context sharing message broker</td>
<td>Context model based</td>
<td>Content</td>
</tr>
<tr>
<td>C. Dorn et al.</td>
<td>2007</td>
<td>Context access control, subscription and control</td>
<td>Server</td>
<td>Context</td>
</tr>
</tbody>
</table>

Most of the methods define a standard for sharing the personal information by using their own method. They utilize the indirect information such as SMS, phonebook, call logs, etc. However, it is necessary to use direct information which can present a physical relationship. Our proposed method helps us solve these problems by using KeyGraph and Bluetooth. Bluetooth enables the physical relationship and can infer closeness directly. Some Bluetooth sensor-based methods have...
been already successfully applied to create mobile social networks. Reyna et al. suggested BluePartner to support mobile social networks and to promote human relationships [13]. Pietiläinen et al. proposed MobiClique to create social network using Bluetooth sensor [14]. The common idea to build a social network is that they use Bluetooth sensor. It is an efficient method since people do not need any annotation and it is simple to create direct relationship.

III. METHOD

A. Server-Client Architecture

We design a platform as illustrated in Figure 1. The server generates high-level contexts and selects people to share contexts from the personal sensor data. In preprocessing step, location mapping is conducted. Since GPS information is recorded as numeric, it has to be converted to semantic labels. The pre-defined locations are mapped with GPS coordinates. Then, preprocessor generates primitive context such as semantic location and a user’s transportation modes. After that, Bayesian network (BN) infers high-level contexts such as activities from the primitive contexts. High-level contexts are stored in database and shared by a standard which is created from KeyGraph.

B. Bayesian Networks for Context Inference

Bayesian network (BN) is a probabilistic graphical model that can be used to uncertain environments [15]. It is represented as a directed acyclic graph that nodes correspond to variables and arcs correspond to probabilistic dependencies between the nodes. For modeling Bayesian network, variable selection, network structure generation, setting probability, and model evaluation steps are needed. The variables and states for modeling in our platform are shown in Table II.

<table>
<thead>
<tr>
<th>Variables</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Dawn, Night, Evening, Afternoon, Morning, Midnight, Noon</td>
</tr>
<tr>
<td>GPS speed</td>
<td>Fast, Middle, Slow</td>
</tr>
<tr>
<td>Transportation mode</td>
<td>Walking, Running, Staying</td>
</tr>
<tr>
<td>Location</td>
<td>Student hall, School, College of engineering, Library, Etc.</td>
</tr>
</tbody>
</table>

Table II. VARIABLES AND STATES FOR BAYESIAN NETWORK MODEL

Figure 2 shows a model to infer ‘eat’ activity. The nodes designed consists of primitive state nodes, middle nodes, and a result node. The raw sensor data determine the primitive state nodes. If the primitive state node depends on a user’s transportation mode, the input state value can be calculated by average acceleration. The input states for GPS speed and time are determined from pre-defined threshold. The middle nodes between result and primitive state nodes decrease the complexity of the model by minimizing the size of Conditional Probability Table (CPT). A final result is calculated from middle nodes as probability values.

We design three more models for inferring a user’s behaviors such as walking, running, eating, studying, and staying. From the probabilistic inference, contexts are automatically recognized without any annotation.

C. Keygraph-based Context Sharing

The original KeyGraph extracts a key term from co-occurrence between terms in a document. In this paper we use KeyGraph to extract two types of key people who have strong social relationship in a community and have many different communities. The social groups are connected from key people and between groups share context that is different from the internal context information. Bluetooth data is collected to apply the method since it is well suited for measuring the physical proximity.

Conceptually, KeyGraph consists of the following two steps. The first step finds some clusters based on association between users. To calculate association, we use co-occurrence of Bluetooth logs. The total amount of Bluetooth logs are represented as $S = \{d_1, d_2, \ldots, d_n\}$ where $d_n$ means Bluetooth logs which are collected during $n$ days. Equation 1 shows the strength of association between users.
\[
\text{assoc}(w_i, w_j) = \sum_{d \in s} \min(|w_i|_d, |w_j|_d)
\]

where \(w_i\) and \(w_j\) denote the user ID and \(|w_i|_d\) and \(|w_j|_d\) denote occurrence frequency of user ID. We assume an occurrence if a user ID is indicated for five minutes in the Bluetooth log. The minimum number between \(w_i\) and \(w_j\) represents the co-occurrence frequency and summation of the co-occurrence during \(n\) days represents the association between users. Then, we sort all of associations in a descending order and link all users serially. Linked users are shown as a graph. The number of users to constitute a graph is defined as a heuristic method. Some users in the graph have more than two links when two unlinked users have stronger association strength than linked users. In this case, previous link is disconnected and the users who have stronger association strength are connected with each other. It makes a group into two groups called clusters or foundations. Since the association is computed once a day, the clusters are updated every day.

In the second step, key people who connect the groups are extracted. To do it, we compute co-occurrence between the cluster and the user ID as shown in Equation 2.

\[
\text{based}(w, g) = \sum_{d \in s} |w|_d|g - w|_d
\]

where \(g\) is a cluster. Equation 3 represents the association between users that is occurred in all day (16 days in our data) and the clusters are computed. \(|g - w|_d\) is defined as shown in Equation 4.

\[
\text{neighbors}(g) = \sum_{d \in s} \sum_{w \in g} |w|_d|g - w|_d
\]

\[
|g - w|_d = \begin{cases} 
|g|_d - |w|_d & \text{if } w \in g, \\
|g|_d & \text{if } w \notin g.
\end{cases}
\]

Key\((w)\) in Equation 5 is to find a key person which is high probability to link the clusters based on a conditional probability. The division of \(\text{based}(w, g)\) by the \(\text{neighbors}(g)\) indicates the rate of the occurrence frequency of user ID \(w\) in the neighborhood of user ID in \(g\). The high ranked key values are then extracted as key person candidates.

\[
\text{key}(w) = 1 - \prod_{i}(1 - \frac{\text{based}(w, g)}{\text{neighbors}(g)})
\]

Finally, we extract a key person by computing the strength between the users who have high frequency occurrence \((HF_i)\) and high \(key(HK)\) value \((HK_f)\). The person who has the highest \(\text{column}(HF_i, HK_f)\) value is selected as a key person which links the clusters.

\[
\text{column}(HF_i, HK_f) = \sum_{d \in s} \min(|HF_i|_d, |HK_f|_d)
\]

IV. EXPERIMENTS

Eleven graduate students collected the sensor-based logs during sixteen days. And there are two working groups who work together in a same place. The Bluetooth logs were recorded every 5 minutes and 18,349 logs were collected for the experiment. Figure 4 shows connection information between subjects. Due to the battery consumption or battery change, the data collection was occasionally stopped.

![Figure 4](image)

To infer the participant contexts, we also collected the GPS, accelerometer, GSP speed, and time logs. The server side system has written in C# and C++ and the client application has implemented on Nokia Lumia 900 for user interaction and data collection. Figure 5 shows Lumia 900 and data collection interface respectively. The sensor data is collected every 0.5 second.

![Figure 5](image)

Figure 6 depicts a KeyGraph result using Bluetooth log. As can be seen in the Figure 6, subject 9 is selected as a key person and two groups are generated. Notice that the member of group1 consists of master students \((1, 7, 11)\) while group2 consists of doctoral students \((2, 3, 5)\) except 4. However, subject 4 usually works with doctoral students. Another thing is subject 6, 8, and 10 have not included in any groups. Actually, they do not have close relationship with others since they are not full-time students.
Based on KeyGraph in Figure 6, we share the inferred contexts by dividing the sharing information into two categories. The reason for dividing the sharing information is privacy protection since the location context is more important than behavior contexts. While the internal groups can be shown in both location contexts and behavior contexts, the groups with the outside which is linked by the key person only share the behavior contexts. The interface we support is indicated in Figure 7 (a) is a location context map which can check the location of other user in real-time. The context is updated every three seconds and is indicated if the name on the map is selected. It can be confirmed in Figure 7 (b). Previous contexts logs are also provided as shown in Figure 7 (c).

![KeyGraph result](image)

Fig. 6. KeyGraph result

V. CONCLUSION AND FUTURE WORK

In this paper, we have introduced a mobile context sharing method to generate social groups automatically. We utilized Bluetooth log as KeyGraph data since Bluetooth sensor can efficiently record the connection relation of people. Bayesian networks infer the user contexts by GPS and accelerometer sensors. The system is implemented based on server-client architecture for efficient data management and to reduce the running time. We have confirmed the potential of KeyGraph from the experiment. It has created a social group without any annotation and it has generated a standard for sharing the contexts according to KeyGraph result. Though call logs and SMS have a privacy problem to collect the logs, it will give more reasonable result if the logs are used with Bluetooth data. In the near future, we are planning to support some contents (movie, picture, etc.) and more contexts. We are also implementing the web-based interface to construct more flexible system.

ACKNOWLEDGMENT

This research was supported by MICROSOFT KOREA, INC., (Microsoft Korea) as a project, “Mobile Context Sharing for Social Relationship with Microsoft Windows Phone 7.”

REFERENCES