Extracting Meaningful Contexts from Mobile Life Log

Youngseol Lee and Sung-Bae Cho
Dept. of Computer Science, Yonsei University
Shinchon-dong, Seodaemun-ku, Seoul 120-749, Korea

tiras@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

Abstract. Life logs include people's experiences collected from various sources. It is used to support user's memory. There are many studies that collect and store life log for personal memory. In this paper, we collect log data from smart phone, derive contexts from the log, and then identify which is meaningful context by using a method based on KeyGraph. To evaluate the proposed method, we show an example of the meaningful places by using contexts and GPS logs collected from two users.

1 Introduction

Several researchers have recently developed life logging system. Life log includes people's experiences which are collected from various sensors and stored in mass storage device. It is used to support user's memory and satisfy user's needs for personal information. If he wants to inform other people of his experience, he can easily share his experience with them by means of providing his life log.

Here, we use smart phone for collecting life log (GPS, SMS, call, charging, MP3, photos taken, images viewed, and weather information). Smart phone is a mobile device that includes color LCD screen, mass storage, large memory, and communicative function by using Wi-Fi, Bluetooth, and infrared. It also has a variety of software such as scheduler, address book, media player, and e-book. Mika Raento developed a framework for collecting contexts from smart phone [1], which collects GSM Cell ID, Bluetooth, GPS data, phone data, SMS data, and media information that are transmitted to the server. The contexts could be provided for other contents as additional information. Panu collects log data from mobile devices, and extracts features by preprocessing the log data [2]. The mobile device uses GPS log, microphone, temperature, moisture, and light sensor.

In life logging system, data mining is one of important issues, which identifies more important information than others. Life log system collects all information related to user's experience, which includes more data as time passes. Therefore, it is practically difficult to explore and search large amount of life log. It would take enormous time and be boring to explore all life log data.

In order to prevent from wasting time, it should be possible to check the priority of data and explore more important data. If enormous amount of data are collected from various sensors, it would be effective to explore only meaningful data.

In this paper, we collect log data from smart phone, derive contexts from the raw log, and then identify which is more meaningful context. To evaluate the proposed method, we show an example of the meaningful places by using contexts and GPS log data.

2 Related Works

Context-based video retrieval system recorded life-log video with contexts, and used contexts to search parts of video [3]. User could query related to contexts, and then the system response the candidates of video data. However, if video data are accumulated for a long time and there are too many video data, user would have difficulty to search necessary part of video. It would be more helpful to extract the candidates of meaningful contexts.

MyLifeBits Project is one of the implementations of personal record database system [4]. Personal information is collected by PC, SenseCam and so on, and stored in MS SQL server with relationships among personal information. However, user would have difficulties to explore and search contents because of large amount of personal data.

KeyGraph-based mobile contents management system was suggested to manage user's information in mobile device, which extracted important information using KeyGraph algorithm and provided searching or exploring contents [5]. The problem of the system is using only log data. If analysis and inference of the data was added to the system, it would give better performance.

3 Proposed Method

The whole system consists of four parts: collecting logs, preprocessing logs, inferring contexts using Bayesian networks, and data mining, as shown in Figure 1.

3.1 Log Collection

In this paper, logs are collected from smart phone. The definition of logs is user's all information collected from smart phone. The kinds of logs are GPS, SMS, call, MP3, photo, viewer, charge information from smart phone, and weather information from Internet. Table 1 summarizes the information that could be collected from smart phone. Places visited by user are extracted from GPS log. Records for telecommunication are extracted from call and SMS log. Weather information from Internet can affect user's emotion and status. MP3 music affects user's emotion. Memorable event is derived from photo record. Record of image viewer provides information about user's favorite photo or picture.

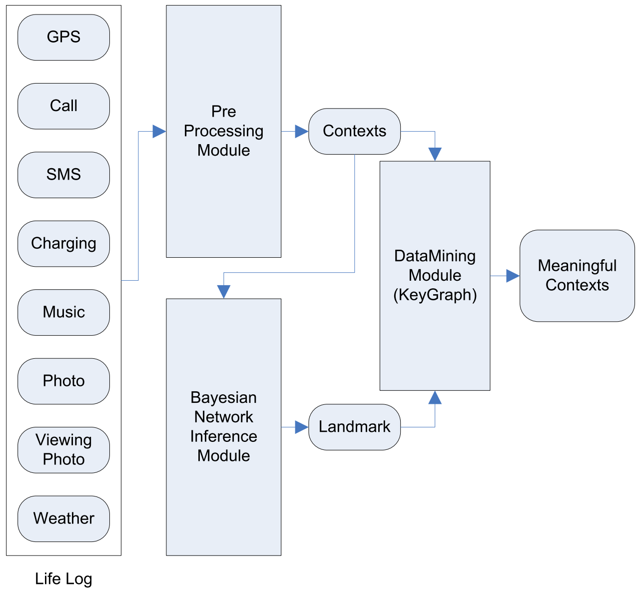

Figure 1. System overview

Table 1. Logs collected from smart phone

Log Type	**Information from Log**
GPS	Longitude, Latitude, Speed over ground, Course over ground, Date, Time
Call	Phone number, StartTime, EndTime, Transmission/Receipt/Absence
SMS	Phone number, Send/Receive, Time
Picture Viewer	FileName, StartTime, EndTime
Photo	FileName, GenerationTime
Weather	Weather, Distance (km), Cloud (%), Temperature (°C), discomfort index (%), sensory temperature (°C), Rainfall (mm), Snowfall (cm), Humidity (%), the direction of the wind, wind speed (m/s), sea-level pressure (hPa)
MP3	Title, StartTime, EndTime, Position
Charging	Percentage of charging, Charging status, Time

Table 2 shows the period of collecting logs. Period of collection varies according to log type. Viewing picture and listening MP3 music are collected at execution of the application. Log data of SMS, call, photo are collected once for a day. GPS location data and information of charging are collected per thirty seconds.

Table 2. Collection of log data and period of collection

Log Type	Collection of Log	Period of Collection
GPS	Collection from GPS module	per second
Call	Collection from stored records for telecommunication	per day
SMS	Collection from stored records for SMS messages	per day
Picture Viewer	Collection from records for using image viewer	per execution of image viewer
Photo	Collection from records for generation of photo file	per day
Weather	Collection from Internet	per day
MP3	Collection from MP3 player	per execution of MP3 player
Charging	Collection from background program	per 30 seconds

3.2 Log Preprocessing and Context Generation

Some information must be transformed to semantic information. For example, GPS coordinate is a real value composed of longitude and latitude. Its value is not intuitive to the human, and so it is transformed to nominal value, such as school, hospital, etc. Similarly, other value is transformed from real value to nominal value. Listening one song more than 5 times can be transformed to 'listening frequently one song'. This transformation is required to input the context to Bayesian network inference module for landmark.

In low level data, frequent or rare data could be meaningful to user. Therefore, statistical analysis of the following items is executed.

- Daily frequency of events
- Sum of daily events
- Elapsed time from the last events occurred.

Table 3. Increase and Decrease of Density

Log Type	Increase of Density	Decrease of Density
GPS	If (GPS event) density++	If (one hour after the event) density--
Call	If (Call event) density++	If (one hour after the event) density--
SMS	If (SMS event) density++	If (20 minutes after the event) density--
Viewing Photo	If (Seeing photo event) density++	If (5 minutes after the event) density--
Photo	If (Photo event) density++	If (30 minutes after the event) density--
Weather	No density	No density
MP3	If (Listening music event) density++	If (30 minutes after the event) density--
Charging	No density	No density

Density of the events is also meaningful to user. Table 3 shows the criteria of increase or decrease of density of the events. If events occurred continuously, the value of density gets increased, leading to the moment of high density of events. The additional information of logs, frequency, density, etc is context. These contexts could be used for user modeling or providing service.

3.3 Landmarks derived from Contexts

Contexts can be useful basis of inference of landmarks that are semantic information or events to remain user's memories. In order to infer appropriate landmarks, Bayesian network module is used. In our previous work, we attempted to extract landmarks through Bayesian network inference module [6], which made an inference of user's emotions and activities according to various life log captured from mobile devices. In this paper, they are used to make context-clusters because landmark represents the major change of contexts.

3.4 Data Mining by KeyGraph

In this paper, KeyGraph is selected to extract the meaningful contexts and the relations among contexts from the context-groups. Probability model or ontology and other rule-based methods can be used to distinguish meaningful contexts from whole contexts. These methods are using frequency or probability of context based on heuristics in order to estimate context. They shall miss infrequent and unknown but personally meaningful contexts. KeyGraph is a chance discovery algorithm that can estimate the importance of context by considering relationships of contexts. Each segmented context-group contains contexts that share user's common status, and clusters are formed from the contexts that frequently share user's common status. Contexts in a cluster have common characteristics. If a context links to more than two clusters, it is regarded as a turning point in user's contexts. The context is meaningful in user's life.

Frequent contexts are selected, and then the relationship among contexts is calculated by summing up the frequencies of contexts in a context-group. $|X|_S$ is a frequency of a context X in a context group S, and *relationship*(w_i, w_j) is the strength of relationship between contexts w_i and w_j.

$$relationship(w_i, w_j) = \sum_{S \in D} \min(|w_i|_S, |w_j|_S) \qquad \textbf{(1)}$$

Links that connect between the contexts are selected in strong relationship (*relationship*(w_i, w_j)) order. If all frequent contexts cannot be connected, the graph structure is divided into more than two clusters. If a context links to more than two clusters, it is regarded as a meaningful context.

clusterOne in equation (2) is the sum of the strength of relationships between a context w and a cluster g.

(**2**)

clusterAll in equation (3) is the sum of the strength of relationships between a context *w* and all clusters in all segmented place-sequences.

$$clusterAll(g) = \sum_{S \in D} \sum_{w \in S} |w|_S |g - w|_S \quad \textbf{(3)}$$

$$|g - w|_S = \begin{cases} |g|_S - |w|_S \ (w \in g) \\ |g|_S \ (w \notin g) \end{cases} \quad \textbf{(4)}$$

clusterOne and *clusterAll* are needed for *clusterValue* in equation (5), which is the probability of the strength of relationship between a context and a cluster.

$$clustervalue(w) = 1 - \prod_{g \subset G} (1 - \frac{clusterOne(w, g)}{clusterAll(g)}) \quad \textbf{(5)}$$

The Higher *clusterValue* of the place is, the higher the importance of the context is in all place-sequences. Finally, the contexts that have high *significance* value in equation (6) are selected as meaningful contexts.

$$significance(w_i, w_j) = \min(|w_i|_S, |w_j|_S) \quad \textbf{(6)}$$

where w_i is a place of high frequency value, and w_j is a place of high cluster value.

4 Experimental Results

In order to evaluate the usefulness of the proposed method, we show the extraction of meaningful places from visited places by using GPS logs, contexts and landmarks that two undergraduate students collected for two weeks. Location is one of important contexts for many systems. For example, life-log video system records video data by using a small wearable camera [7], which stores location as an index to retrieve necessary video data. MyLifeBits collects quite a lot of information about a person and constructs personal database [4], which also collects location as one of useful information. Therefore, we extract meaningful places from mobile life log as an example of extracting meaningful contexts.

4.1 Log Collection

We used logs that two undergraduate students had collected for two weeks to evaluate the proposed method. Figure 2 depicts the number of records in the log data.

We also requested that users report their life. It is required that users record their visiting places, visiting frequencies, and their activities. The user does not usually memorize all places for daily life. If user records one place in his report, the place is the memorable and meaningful place for him. Therefore, the places extracted from the proposed method should be compared with reports. Other contexts except place need the same criteria.

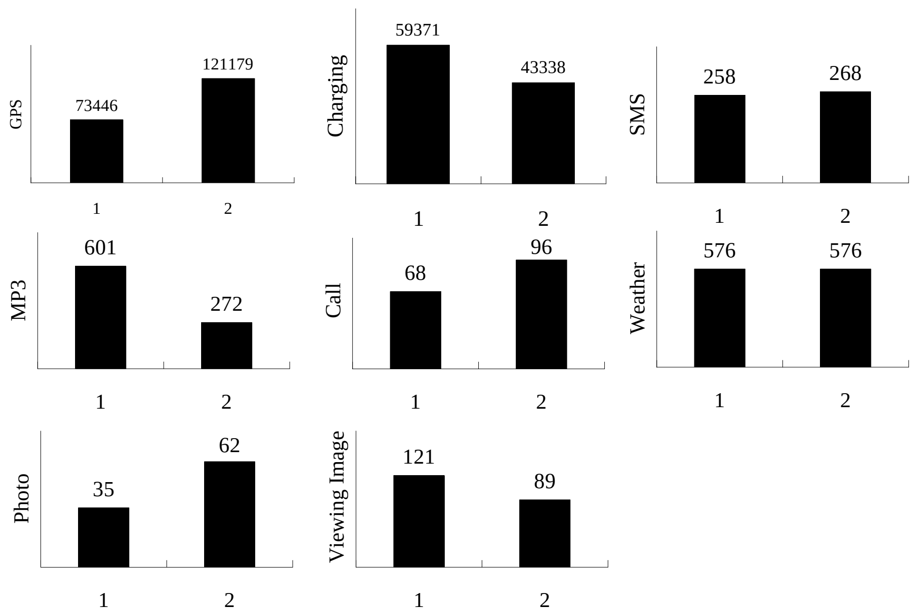

Figure 2. Logs from two undergraduate students

4.2 Log Preprocessing and Context Generation

Our system uses log data collected from smart phones. GPS data is numerical that contains longitude and latitude. Getting visited places needs the process that converts GPS data to place information. Figure 3 shows the process. In addition, all logs are required to analyze frequency and density, and so on.

4.3 Inference from Contexts

Landmarks are extracted from 39 Bayesian networks. 19 of them are designed for place/activity inference, 13 are for user's emotion/condition inference, 5 are for circumstance situation inference, and 2 are for event inference [6]. The total number of landmarks extracted from 2 undergraduate students is 1522. But not a few of them coincided with each other, so that the number of sentences is quite fewer than that of landmarks.

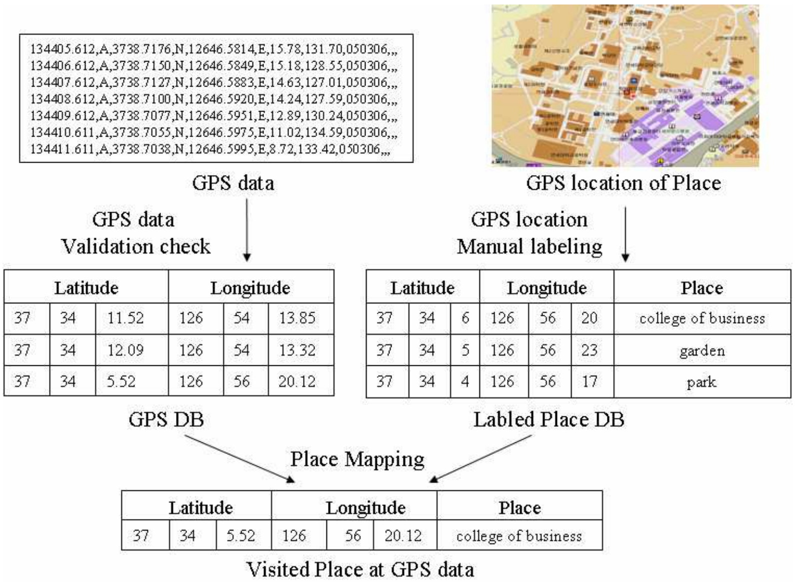

Figure 3. Labeling GPS data and converting to place name

4.4 Data Mining

In order to define the problem clearly, only contexts of place and landmarks are used to extract meaningful places from all contexts and landmarks. Polaris [8] is used to visualize the results. The result of running KeyGraph of student 1 is like Figure 4. {university map, Yonsei engineering research center, Industry-University research center, soccer stadium, movie theater, physical education center, Yonsei eagle statue, amphitheater} are relatively frequent places. {department store, bridge, college of liberal art, college of theology, laboratory} are selected as rare and meaningful places for student 1.

'department store' and 'laboratory' are represented only two times in data samples, but visiting department store is meaningful event in shopping, and laboratory is an important place for our experiment in analyzing her log data. We can find rare and meaningful places.

Student 2 is similar to student 1. Student 2 provides the most GPS data set. Figure 5 shows the result. {parking lot, Yonsei engineering research center, college of engineering, cinema, coffee shop, Chinese restaurant, square, crossing, global lounge, flower garden, college of music, main auditorium, amphitheater} are relatively high frequency places, and {department store, electronic market} are key places. Especially, two places (department store and electronic market) are selected as key places. These places are related to shopping. Department store is related to usual shopping, and electronic market is related to buying a cell-phone with her boyfriend. They are meaningful places visited in two weeks.

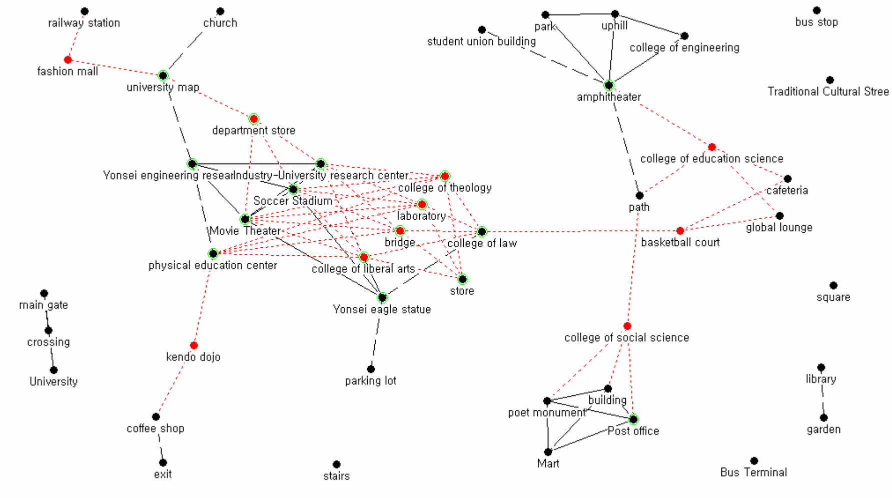

Figure 4. Graph generated with places (student 1)

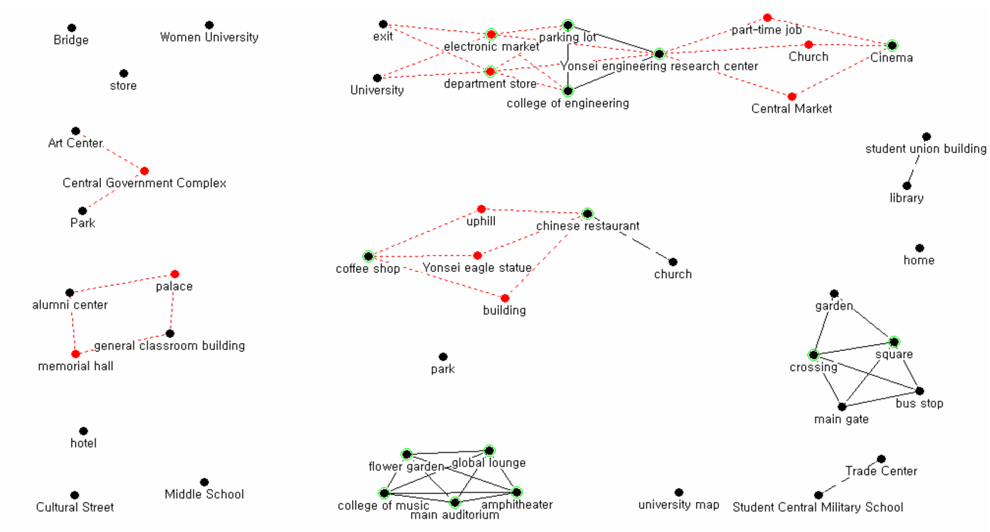

Figure 5. Graph generated with places (student 2)

5 Summary and Future Work

We have proposed a method to extract meaningful contexts from life logs using KeyGraph with landmarks. Here, the landmarks are extracted through Bayesian network inference module and represents user's action or emotion in real life. In order to evaluate the proposed method, it is applied to a sequence of places visited by two undergraduate students for two weeks. As a result, it can extract meaningful places for users. What remains to be studied are collecting more data from more students and extracting more meaningful one from other contexts, such as SMS texts, MP3 use, Call log, photos taken, and so on.

Acknowledgments. This work was supported in part by LG Electronics and MIC (Korea) under ITRC.

References

1. Raento, M., Oulasvirta, A., Petit, R., and Toivonen, H.: ContextPhone: A Prototyping Platform for Context-aware Mobile Applications. IEEE Pervasive Computing, Vol. 4. No. 2. (2005) 51–59
2. Panu, K., Jani, M., Juha, K., Heikki, K., and Esko-Juhani, M.: ContextPhone: Managing Context Information in Mobile Devices. IEEE Pervasive Computing, Vol. 2. No. 3. (2003) 42–51
3. Aizawa, K., and Hori, T.: Context-based Video Retrieval System for the Life-log Applications. In Proc. of MIR 2003, ACM, (2003) 31–38
4. Gemmell, J., Bell, G., and Lueder, R.: MyLifeBits: A Personal Database for Everything. Communications of the ACM, Vol. 49. No. 1. (2006) 88–95
5. Kim, K.-J., Jung, M.-C., and Cho, S.-B.: KeyGraph-based Chance Discovery for Mobile Contents Management System. Knowledge-based and Intelligent Engineering Systems (KES Journal), (2007) (To appear)
6. Cho, S.-B, Kim, K.-J., Hwang, K.-S., and Song, I.-J.: AnyDiary: Daily Cartoon-Style Diary Exploits Bayesian Networks. IEEE Pervasive Computing, Vol. 6. No. 3. (2007) 66-75
7. Aizawa, K., Tancharoen, D., Kawasaki, S., and Yamasaki, T.: Efficient Retrieval of Life Log based on Context and Content. Proceedings of the 1st ACM workshop on Continuous Archival and Retrieval of Personal Experiences, New York, USA (2004) 22-31
8. Okazaki, N., and Ohsawa, Y.: Polaris: An Integrated GPS Log Miner for Chance Discovery. In Proceedings of the Third International Workshop on Chance Discovery and Its Management, Crete, Greece (2003) 27-30
9. Ohsawa, Y.: Keygraph as Risk Explorer in Earthquake-sequence. Journal of Contingencies and Crisis Management, Vol. 10. (2002) 119-128
10. Blighe, M., Le Borgne, H., O'Connor, N., Smeaton, A.F., and Jones, G.: Exploiting Context Information to Aid Landmark Detection in SenseCam Images. Communications of the ACM, Vol. 49.No. 1. (2006) 88-95
11. Ohsawa, Y., Benson, E., N., and Yachida, M.: Keygraph: Automatic Indexing by Co-occurrence Graph based on Building Construction Metaphor. Proc. Advanced Digital Library Conference (IEEE ADL'98) (1998) 12-18
12. Horvitz, E., Dumais, S., and Koch, P.: Learning Predictive Models of Memory Landmarks. CogSci 2004: 26th Annual Meeting of the Cognitive Science Society, Chicago (2004) 1-6

13. Abowd, D. G., Dey, K. A., Brown, J. P., Davies, N., Smith, M., and Steggles, P.: Towards a Better Understanding of Context and Context-Awareness. Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing (1999) 304-307