Geo-Social Tracking System for Detecting Unusual Behavior: Visualization of Social Activities based on Spatiotemporal Change

Seung Hyun Seo and Chul Chung

Graduate School of Software, College of Information Science and Technology, KAIST, Seoul, KOREA

Hefos@kaist.ac.kr, Hefos@mail.plustech.co.kr, Chul.chung@kaist.ac.kr

Abstract

This paper examines the use of geographical tracking system as a method of detecting juvenile behaviors, while providing an extension to existing a mobile service which has about 30000 users in Korea. The degree of juvenile behaviors has gotten more severe recently with the invention and proliferation of smart phones; they are easily exposed to harmful contents using smart phones. However, we could solve this problem by detecting unusual behaviors, because people exhibit certain behaviors when using devices and contents. In Mobile, since social media is one of the most popular services, we chose it as the research target. In addition, we propose both data visualization and geo-social tracking system for user profiling to examine the underlying meanings of individuals' activities at the social media with spatiotemporal information. With social media becoming more 'geographical', many social networks, including Facebook, provide location based services to assist users' daily lives. In experimental results, we show our visualization methods and weight functions for plotting on the map. Moreover, at the end of this work, we propose how our work could be integrated as an add-on to existing services. Lastly, we discuss our data visualization methods and the constraints we encountered during information gathering. Experimental results in this research are going to contribute to other researches attempting to profile individuals on mobile phones.

Keywords: *Mobile, Social Media, Geo-Tagging, User Profiling*

1. Introduction

Online social media have become more "Geographical" as usage of location information increased for various purposes. Nowadays, more researches on user profiling use locations such as in recommendation [12], link prediction [9, 13] and tie-strength measuring [4, 5, 10, 11]. In fact, location information is continuously applied in various other fields such as target marketing which is based on radius of action, place features and movement paths. Because our research is related to profiling behavior, we have gotten an idea that we could use information in social media.

Moreover, the growing distribution of smart-phone has greatly contributed in proliferating LBS (location based system) and social network [16] which also enabled easier acquisition of geographic information. This definitely implies the growing importance of location information in social media. Also, the importance of geographic information gained from pictures on social media has already been mentioned in other research [8]. Moreover, it seems transparent that location information in social media affects our daily lives through services like "Flickr" and "Eye-Fi". Furthermore, more important point regarding smart phones in our research is that almost all teenagers in Korea own smart phones, use those services, and carry the devices with them at all times. Based on this point of view, we can use smart phones as the resource of gathering behaviors and location information to profile juvenile behaviors.

LBS generally uses GPS. However, other sensors of smart phones such as 3G [2], GSM [20], Wi-Fi, accelerometer [16] were also used for researches related to location and social network as part of user profiling. In our system, user positioning method is based on using location tracking technologies in android system. In addition, to enhance energy consumption in the mobile side, our App adopts Google Best Practice [19] as a benchmark model.

There has been a research on how many and when people take certain actions in social media such as “Like”, “Post” and “Comment” [29]. Also, this research is further stretched to business areas classified as in engagement-rate i.e. the rate at which people watch and engage mostly in social media. For these sort of researches which usually show patterns of user behavior on crowd basis rather than individual basis, we were motivated to challenge ourselves by finding habits of individuals through tracing users’ locations and crawling users’ activities on Social Media, which in this case, Facebook.

Although Facebook tremendously influences our lives, it is difficult to measure its effect over our personal daily lives [21]. Although Facebook has the largest user number among all of social networks, its attempts to expand to location service does not seem successful in spite of their effort such as acquiring gowalla and instagram. We have expected a more intense profiling through merging location information with the exquisite information provided by Facebook. Therefore, we can understand when, where and how many times each user uses social media through quantitatively measuring the dependency on social media.

Due to difficulties in collecting location information from Facebook, we designed an asynchronous geo-social tracking system which gathers activity data from Facebook servers by using Graph-API and receives location information from mobile devices. Since not all Facebook objects involve Check-ins and its objects can be modified by users, we collect location information from mobile phones instead.

Although crowds represent trend, since individuals have their own propensity, they cannot explain each person’s characteristics. Thus, we focus on individual logs and we hypothesize that if a person has habits of social media usage according to situation specifically in different places, users’ habits would be shaped like cluster in the plotted form on the map and chronological form of graph. Then, we could detect unusual behavior through the map. Also, a more exact nature of places, relations and homogeneous groups can be inferred from the social media usage patterns [3]. Finally, our research leads to possibility of detecting unusual behavior by examining location with other information, in this case of Social Media.

2. Related Works

This section explains both the previous researches that our system adopted and the reason why we decided to use it. It also explains how we apply them to our system.

2.1. User Profiling

One of our challenges in user profiling is detecting innate characteristics of individuals. Thus, our system analyzes users’ behavior data on what users prefer in daily lives. Generally, user profiling is divided into knowledge-based and behavior-based [1]. Behavior based profiling uses data derived from observing individuals [22] and knowledge-based profiling uses questionnaires to interpret one’s habits [23]. On the other hand, in information science, user profiling called profiling practice is classified as in target: individual and group [15].

From the two methods, our research adopts behavior based methodology for which questionnaires were not included in the process of patterning historical data. Also, as of targets, we focus on individual profiling rather than group. Thus, our research could be called behavior-based individual profiling.

2.2. Location sensing and social media

As many location based services have been launched and supported by social media or third party services, users can easily compose geo-tagged contents. Most social media have been eliciting geographic information from Check-ins and geo-tagged data and they are essential to understand the radius of actions and where users were [7]. Location based user profiling is continuously expanding by merging with social media properties.

Previous researches propose various ways to collect users' states by using various sensors. For gathering physical location information, there are two typical ways for which we can use to collect geographic information; one is to trace location each period, and the other is to utilize users using either Check-ins or our app voluntarily. In our research, we choose the first method to avoid errors which can be occurred by the voluntary method.

2.3. Data visualization

Data visualization could help us to understand experimental results by analyzing raw data in various ways, from which we could elicit meanings of data and relation between data. Previous researches have already mentioned the importance of data visualization including data transformation [30] so that many researches show their results through visualization method. Moreover, in our research, data visualization is an important part, as we will present users' life log data in various ways of visualization to have users interpret it themselves. Thus, we need to present users with eidetic and intuitive visualization of information.

However, as visualization is often useless for some who lack information that is required to understand its meanings, we have to provide knowledge prior to acknowledging the given information for better understanding [31]. So, our challenge in visualization is to make users capable of handling visualized life log data without prior knowledge.

2.4. Battery problem

Battery problem is the most important thing in mobile phones since they possess numerous on-going sensors which cannot be turned off. There are several user profiling researches on mobile phones with battery problems [14, 17, 18] and one of them proposes four principles for LBS [14]: Sensing Substitution (SS), Sensing suppression (SR), Sensing Piggybacking (SP), Sensing Adaptation (SA).

In our system, we applied SS and SP for other two principles are too expensive. This means if we cannot perfectly construct and apply SR and SA, efficiency of energy consumption cannot be guaranteed since we might result in consuming unnecessary resources on monitoring. Thus, we use location sensing technology in android location framework of which we benchmark the latest best practice well applied above principles publicized in Google I/O 2011.

2.5. Privacy

There are researches on mobile phones covering location privacy in fields of social network [24, 28] and personalized system [27]. As mobile devices have become a large sensor network, it is provoking more serious privacy problems where users do not want to be exposed. It is important for user to prevent disclosure of location information as they are considered private. Previous studies have proposed solutions, such as anonymity model [24], location context aware [26] and server system algorithm [28].

The above researches show some important results that while services are considered useful, users are concerned about privacy. Our users who have used our location-tracking integrated services are also concerned on abusing their private information. For these reasons, because our system is dealing with personalized information related to social network, users' information is not exposed to the others in our system. Also, we pledge not to use our user information for any other purpose apart from research.

3. Geo-Social Tracking System

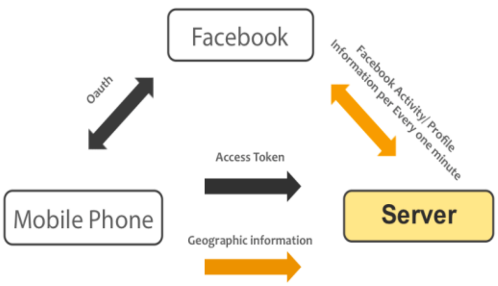

Figure 1. System Flow

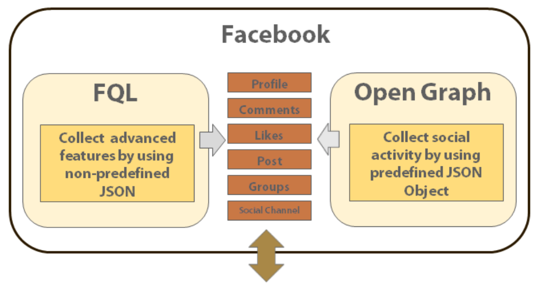

Figure 2. Mothod using in geo-social tracking system

We designed two packages for information gathering and merging to collect different information from each side. Our system has following steps: 1. User run our App and authorize it. 2. Our App gets the Facebook Access Token and sends its information to geo-tagging server. 3. On mobile phones, our App sends geographical information to geo-tagging server. 4. Geo-tagging server receives location information from mobile and collects Facebook activities from Facebook server. 5. Lastly, geo-tagging server merges information and makes it readable as data visualization "Figure 1".

3.1. Mobile

We collect location information from three types of sensors which are GPS, Wi-Fi, and 3G. Since GPS consumes excessive amount of battery, users usually do not turn it on. Thus, we use network sensors mainly for detecting a location. As we choose to trace a location, a user does not need to use our App except in case of authenticating once at first. Moreover, due to both distinctions of system configuration for networking in android system and iOS problem, our system supports android version lower than 2.3.3.

When mobile sends location information, user ID and types of sensors are transferred to server every thirty minutes. Since our system is not limited to which device a user use for running Facebook i.e. PC, Mobile and Tablet, we just need access tokens that will not be temporarily expired and authenticate as server side method in mobile App. There is our App package construction. "Figure 3(a)"

3.2. Geo-tagging server

Facebook suggests Graph-API for information gathering over other methods. However, the crawled information is not shown as in the order of Facebook news feed for it is not ordered in updated time but in created time. Whereas FQL (Facebook Query Language) supports users' news feed in the form of JSON object by querying the updated time of each wall object. For we focus on users' activities which even exist in other walls, we decide to use both of them partially to collect more exact information. We apply both FQL to find out updated object and Open Graph to collect more details by visiting every wall with which a user has been related. "Figure 2"

Server receives access token once when authenticating through mobile and builds database of users. Due to social network's characteristics, the server updates information every thirty minutes when mobile devices send location information "Figure 3(b)". Moreover, our system is under endless loop for data collection which is gathered from Facebook server. The activities which users use actively are as follows:

3.2.1. Post: This object could involve other objects such as Check-ins, Event and Photos. It has its own created time and updated time and it could include links, thumbnails and tags. This is a kind of activity that is written on friends' or user's own wall and is updated by commenting. After certain period, this object would not be updated.

3.2.2. Comment: Comments could include links and tags like Posts and they also have their own created time which affects updated time of related Posts. As mentioned above, after certain period of time, updated time cannot be collected by querying FQL. In this case, in order to collect objects perfectly, Facebook notifies objects that should be considered.

3.2.3. Like: There are two kinds of Like: Post and Comment. This is activated by user with current session key by pressing the like button on a comment or a post. This action doesn't affect up-dated time of object on which user pushes. Since its updated time is not recorded in any other way, we cannot figure out its exact created time.

3.2.4. Social Activity: Social activities from third party are written on either users' or friends' wall through application but they cannot be fully tracked. For example, open graph action which users have customized cannot detect what kinds of actions users do in social channel. Therefore, we can hardly detect what kinds of actions users do on third party services.

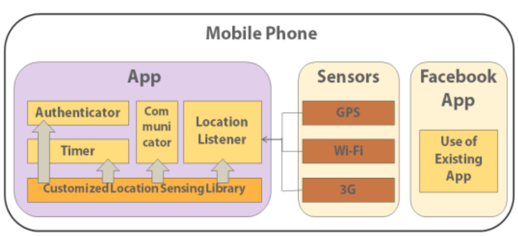

Figure 3. (a) Mobile Phone

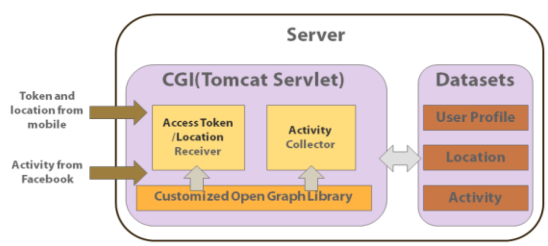

Figure 3. (b) Mobile Phone

4. Implementation Results

In this section, we mention constraints by using Facebook API and the amount of time spent for information crawling. Also, some charts which are shown as in both clustered place in the form of map and in chronological interfaces would represent user's behavior pattern. Lastly, we demonstrate our visualization method to delineate self-cognitive visualization.

4.1. Constraints of implementation

We arrange several constraints when gather information in using Facebook API. We also introduce the way to solve our problem and The problems we could not solve.

4.1.1. Synchronization problem: Facebook has data synchronization problem between Graph API and FQL and these are usually derived from numbers such as like counts and comment counts. In this case, data acquired from Open Graph is more exact.

4.1.2. Server problem: Facebook randomly returns either "400 bad request error" regardless of correct query or "500 internal error" frequently. Exceptional cases cause scheduling problems to occur which are similar to deadlock. So in these exceptional cases, we use LRU (Least Recently used) algorithm.

4.1.3. Mobile SDK problem: In mobile side authentication of Android 2.3.3, Facebook returns unknown errors when the client accepts permissions which are over a particular count. Since Facebook public SDK does not provide support for this problem, we use limited number of permissions.

4.1.4. Contents aging policy problem: Facebook doesn't consider old objects as fresh contents even though they are updated. Thus, querying updated time to crawl all updated objects would not include some of old activities. If we want to overcome this, we will have to query all objects in each wall. So, we choose to collect more accurate times which "Like" buttons were clicked rather than to collect more objects with inaccurate times.

4.2. Visualization

Previous research shows engagement rate in days when users usually use social media [29]. Although this could represent crowd trend, it cannot imply personal characteristics. As expected, each user has a different habit in using social media as against crowd profiling. Also, users have their own preference in action. Although preference of users shows different tendency, it cannot explain where difference comes from and how we can use this information. Thus we use another approach which is to use maps. As we hypothesized, we have plotted spots of using social media on the Google map. Figure 4(a)

Our visualization is designed to help user understand its meanings. For it is difficult to convey intuitive information to all users through one picture, we decide to use users' intuition rather than artificial intelligence. This means that since human beings have superior sense of judgment than machines, users can find routines and unusual actions through their organized activity log. Also, they can realize what kind and type of person they are in social media.

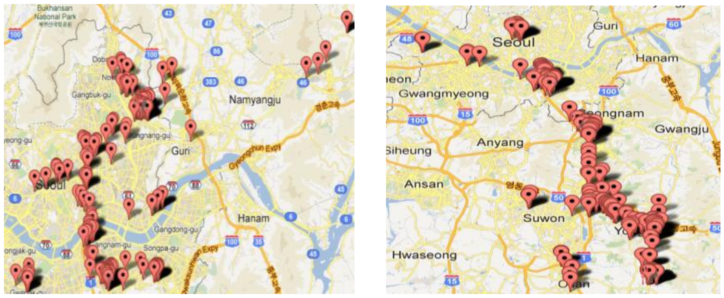

Figure 4. (a) Visualization based on spot user has visitied

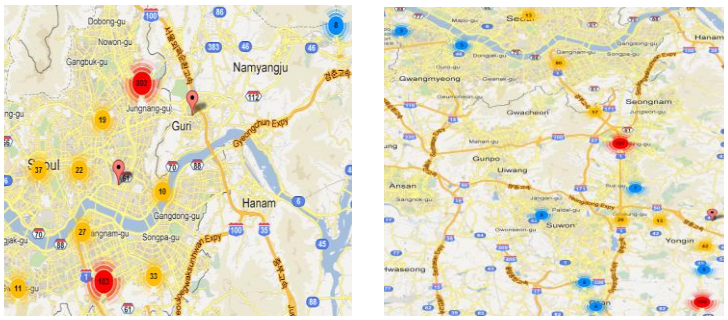

Figure 4. (b) Visualization using cluster with color based on weighted function

Before visualizing usage patterns, to evaluate more precisely, values of usage quantity among users' activities should be converted to absolute values to consider weighting of activities. Also, we must consider the details of activities (*i.e.*, text of Post, text of Comments, text of Post user commenting, *et al.*). For our aim of visualization in this research is to help users recognize pattern which is not routine, we believe that depicting clustered information is more effective in understanding than to show each single information. Figure 4(b)

4.3. Weight function

Our weight formula is defined as the degree of activeness of an action. For example, "Post" is the most positive activity and "Comment" is more positive than "Like". In addition, as users have different habits in usage of social media, we add another factor to the formula which is ratio of activities to reflect users' habits. Since we cannot figure out the ratio of activities in social plug-in, we assume that these activities carry the same proportion as the other three activities.

As mentioned above, activities in social plug-in are estimated by proportion of the other activities: p, l and c are initials of "Post", "Like" and "Comment" (1). So, we have calculated ratio of other activities and applied their ratio to social plug-in. f(x) is weight function related to both s(x) which is ratio of activities and v(x) which is a variable value as subjective measures of motivation. These represent values of each activity in our research (2). w(a) is the sum of social activities' weight in the clustered zone "a" (*i.e.*, specific area) consisted of element "g" on the map. Thus, we can state that this is the sum of number of visits and weighted value which is multiplied by number of visits to specific locations (*i.e.*, q(g)) in a certain area (3).

$$x = \{p, l, c\}, s(x) = \frac{x}{p+l+c} \qquad (1)$$

$$f(x) = (v(x) + (s(x) * (\frac{s}{p+l+c+s}))) * \sqrt{x} \qquad (2)$$

$$w(a) = \sum_{g}\sum_{x}((1 + f(x)) * q(g)) \qquad (3)$$

4.4. Users' feedback

Our research participants have given us several feedbacks. Although they were satisfied to find out how they interacted, they hoped to know more about relation in regard to locations such as whom they have frequently interacted and what kind of text they have posted than just the fact that they have posted. They have also complained about the errors of location sensing, too. Through our feedbacks, we realized that people wish to know about relationships with others alongside knowing themselves. Thus, we could consider that even though they interact with others frequently, they are not aware of whom they are being involved with.

5. Conclusion

5.1. Consequence & Limitation

In this paper, we have described our geo-social tracking system and visualization for analyzing users' habits. Also we examine characteristics of individuals through their daily life log in social media. In addition, our visualization presents social media usage propensity as in geographical and chronological flow and proposes visualization technique for finding meaningful places and unusual behaviors. By using geo-tagging method in social media, we show results of what we can do and the future works that can be done.

Our experimental results give us several knowledge. First, in our method used in Facebook, acquiring more accurate information rapidly requires numerous servers since it takes too long to gather information with a few servers only. Thus, an ideal method would be to allocate one server per each user. Second, we found limitations in Facebook when using Graph API. Third, comparing between users shows that not all of users in social media have the same habits and engagement rate. Also, as it shows users' preferred action, we could infer what type of users they are. Lastly, we show possibility that location information with usage history, in this case especially Facebook, could be used for detecting changing point of behaviors.

Lastly, there are some limitations to overcome in our research. First, we cannot gather information from users who just read instead of doing any other actions. Second, text data are not considered. Third, we don't consider with whom users interact with frequently. Fourth, loss and error of geographic information are not dealt as exceptional case, so we had not revised it. Fifth, not all actions in Facebook are considered.

5.2. Applying to legacy services

Our project has started to detect unusual behavior for youth as an extension of existing mobile phone service in Korea. As extension of this, we have ideas how to apply these experimental results to our products.

First, as one of the detecting methods, we propose map visualization with social media usage log. In our results, if they have some changes for their daily routine, it appears on the map, and the location they have visited would be changed. We could detect weird movement by seeing chronological change of location with Facebook usage log.

Second, quantity of using social media would be different from routine. Because they have average usage quantity of mobile, we can find changed point. This way is useful for someone who has loved ones to care about.

Third, except visualization of activities with geographic information, we have gotten knowledge about which we can infer social connection for location even we do not attend at the same time zone, because our participants have tendency communicating with people who are not in there. Thus, we could use this information to measure social graph and to infer location indirectly.

5.3. Future works

Our research would more advance and expand the future paths by overcoming limitations of our research. 1. By considering wall object with location information, we could measure tie-strength between user and wall object. 2. By processing text data with other information such as which users interact and what users post, it would lead to a better understanding in users' interest and affection in places where they have visited. 3. If it could restore lost information, it would speculate users' characters more precisely. 4. One of our challenges is to improve system performance in time and data accuracy. 5. Adding time where user just read, it would be improved as automatic condition detecting system [6]. 6. If we could gather other logs of application on mobile phone, we can profile abnormal behaviors more exactly than what we have done.

References

[1] S. E. Middleton,N. R. Shadbolt and D. C. De Roure, "Ontological User Profiling in Recommender Systems", ACM, TOIS'04, vol. 22, Issue 1, **(2004)** January, pp. 54 – 88.
[2] R. Keralapura, A. Nucci and Z. -L. Zhang, "Profiling Users in a 3G Network Using Hourglass Co-Clustering", ACM, MobiCom '10, **(2010)**, pp. 341-352.
[3] N. Eagle, A. Pentlandb and D. Lazerc, "Inferring friendship network structure by using mobile phone data", PNAS, vol. 106, no. 36, **(2009)** September 8, MIT Media Lab.
[4] J. -P. Onnela, S. Arbesman, M. C. González, A. -L. Barabási and Christakis, "Geographic Constraints on Social Network Groups", PLOSONE, **(2011)** April, plosone.org.
[5] D. J. Crandall, L. Backstromb, D. Cosleyc, S. Suri, D. Huttenlocher and J. Kleinberg, "Inferring social ties from geographic coincidences", PNAS, vol. 107, no. 52, **(2010)** December 28, Indiana Univ., Cornell University.
[6] M. Nambu, "An Algorithm for the automatic detection of health conditions", Dept. of Gerontology, Nat. Center for Geriatrics & Gerontology, Japan IEEE, vol. 24, **(2005)** July, pp. 38-42.
[7] J. Ghosh, M. J. Beal, H. Q. Ngo and C. Qiao, "On Profiling Mobility and Predicting Locations of Wireless Users", ACM, REALMAN'06, **(2006)**, pp. 55-62.
[8] J. Park, J. Kim, S. Seo, J. Huh and S. Maeng, "GeoWeb: The impacts of Geo-tagged Web Page", KAIST, Samsung Electronics, Kiss 2010, **(2010)**.
[9] S. Scellato, A. Noulas and C. Mascolo, "Exploiting Place Features in Link Prediction on Location-based Social Networks", ACM, KDD'11, **(2011)**, pp. 1046-1054.
[10] J. McGee, J. A. Caverlee and Z. Cheng, "A Geographic Study of Tie Strength in Social Media", ACM, CIKM '11, **(2011)** October, pp. 2333-2336.
[11] J. Cranshaw, E. Toch, J. Hong, A. Kittur and N. Sadeh, "Bridging the Gap Between Physical Location and Online Social Networks", ACM, UbiComp'10, **(2010)**, pp. 119-128.
[12] M. Ye, P. Yin, W. -C. Lee, "Location Recommendation for Location-based Social Networks", ACM, GIS'10, **(2010)**, pp. 458-461.
[13] K. Musiał, P. Kazienko and P. Bródka, "User Position Measures in Social Networks", ACM, SNA-KDD'09, Article no. 6, **(2009)**.
[14] Z. Zhuang, K. -H. Kim and J. P. Singh, "Improving Energy Efficiency of Location Sensing on Smartphones", ACM, Mobisys'10, **(2010)**, pp. 315-330.

[15] http://en.wikipedia.org/wiki/Profiling_practices. wikipedia, profiling practice.
[16] E. Miluzzo, C. T. Cornelius, A. Ramaswamy, T. Choudhury, Z. Liu and A. T. Campbell, “Darwin Phones: the Evolution of Sensing and Inference on Mobile Phones”, ACM, MobiSys’10, **(2010)**, pp. 5-20.
[17] Y. Wang, J. Lin, M. Annavaram, Q. A. Jacobson, J. Hong, B. Krishnamachari and N. Sadeh, “Framework of Energy Efficient Mobile Sensing for Automatic User State Recognition”, ACM, MobiSys’09, **(2009)**, pp. 179-192.
[18] Y. Wang, B. Krishnamachari, Q. Zhao and M. Annavaram, “Markov-optimal Sensing Policy for User State Estimation in Mobile Devices”, ACM, IPSN’10, **(2010)**, pp. 268-278.
[19] http://code.google.com/p/android-protips-location/, **(2011)**, Google I/0.
[20] Y. De Mulder, G. Danezis, L. Batina and B. Preneel, “Identification via Location-Profiling in GSM Networks”, ACM, WPES’08, **(2008)**, pp. 23-32.
[21] E. Bakshy, I. Rosenn, C. Marlow and L. Adamic, “The Role of Social Networks in Information Diffusion”**,** ACM, WWW’12, **(2012)** April 16–20, pp. 519-528.
[22] A. Kobsa, “User Modeling:Recent Work, Prospects and Hazards”, Department of Information Science, University of Konstanz, D-78434 Konstanz, Germany.
[23] F. Sebastiani, “Machine learning in automated text categorization”, ACM, CSUR, vol. 34, issue 1, **(2002)** March, pp. 1-47.
[24] B. Gedik, “Location Privacy in Mobile Systems: A Personalized Anonymization Model”, IEEE, ICSCS’05, **(2005)**, pp. 620-629.
[25] M. N. Gasson, “Normality Mining: Privacy Implications of Behavioral Profiles Drawn From GPS Enabled Mobile Phones”, IEEE Systems, Man, and Cybernetics Society, vol. 41, Issue 2, **(2011)** March, pp. 251- 261.
[26] C. Mancini, K. Thomas, Y. Rogers, l. A. Price, L. Jedrzejczyk, A. K. Bandara, A. N. Joinson and B. Nuseibeh, “From Spaces to Places: Emerging Contexts in Mobile Privacy”, ACM, UbiComp 2009, (2009) September 30 – October 3, pp. 1-10.
[27] A. Kobsa and M. Teltzrow, “Impacts of User Privacy Preferences on Personalized Systems”, Kluwer Academic Publishers, **(2004)**, pp. 315-332.
[28] L. Hu, C. Shahabi, “Privacy Assurance in Mobile Sensing Networks: Go Beyond Trusted Servers”, IEEE, **(2010)**, pp. 613-619.
[29] “Strategies for Effective Facebook Wall Posts: A Statistical Review”, Buddy Media, **(2011)**.
[30] Z. Wen, M. X. Zhou, “Evaluating the use of data transformation for information visualization”, IEEE, vol. 14, Issue 6, **(2008)**, pp. 1309-1316.
[31] M. Chen, D. Ebert, H. Hagen, R. S. Laramee, R. van Liere, K. -L. Ma, W. Ribarsky, G. Scheuermann and D. Silver, “Data, Information, and knowledge in Visualization”, IEEE, vol. 29, Issue 1, **(2009)**, pp. 12-19.

Authors

Seung Hyun Seo

M.S in KAIST, 2013

Staff Engineer at PLUSTECH

Chul Chung

PhD in KAIST. 1989

Professor, Graduate School of Software, KAIST