

Smartphone App Usage Log Mining

Woan-Rou Tseng and Kuo-Wei Hsu

Abstract—The rise of smartphones can be connected to the large number of applications, or apps, which can be installed and run by users on smartphones. It becomes important for researchers, smartphone designers, and application developers to know how users use apps on their smartphones. The aim of this paper is to present our work where data mining is applied to smartphone app usage log data with a focus on data preparation. The methods include association rule mining and sequential pattern mining. The results are the discovered rules and patterns that reflect users' real-life app-using behaviors; they can lead to improved user interfaces, and they can also be used to reconstruct the context in which a user used his or her smartphone. We demonstrate an application of data mining that can have an impact on the smartphone industry, and we also demonstrate a data-driven approach that can help us know more about how users use apps on their smartphones.

Index Terms—Association rule mining, sequential pattern mining, smartphone app usage log.

I. INTRODUCTION

During the past few years, people have seen mobile technology developed significantly. Feature phones are mobile phones that provide a richer set of functions, such as camera and multimedia player, and they are no longer restricted to voice communication. Compared to feature phones, smartphones are with more advanced computing capability and communication functionality, and they provide an even richer set of functions, such as full web browser and navigation. Smartphones allow users to install third-party applications, or apps, to customize their phones and to enrich the functions of their phones.

By allowing apps to have better integration with their underlying systems, smartphones have changed many aspects of daily lives of many people. Many interesting apps can be installed and run on smartphones, such as games and applications for social network services (including instant messaging). The rapid development of smartphone apps has been an important driver of smartphone adoption. Therefore, researchers (for example, social scientists [1]), smartphone designers, and application developers would like to know how users use the apps on their smartphones.

Manuscript received August 30, 2013; revised November 11, 2013. The work presented in this paper was supported in part by the National Science Council of Taiwan under Grant Numbers NSC 101-2221-E-004-011 and NSC 102-2221-E-004-013. This paper was also supported in part by the X-Mind Research Group at the National Chengchi University, Taipei, Taiwan, sponsored by "Aim for the Top University Plan" of the university and the Ministry of Education of Taiwan. Their support is gratefully acknowledged.

W.-R. Tseng was with the Department of Computer Science, National Chengchi University, Taipei, Taiwan (e-mail: 100753016@nccu.edu.tw).

K.-W. Hsu is with the Department of Computer Science, National Chengchi University, Taipei, Taiwan (Corresponding author, e-mail: hsu@cs.nccu.edu.tw).

More specifically, for researchers, they are interested in rules or patterns of smartphone app usage that can help them reconstruct the context in which a user used his or her smartphone; for smartphone designers and application developers, they are interested in rules or patterns that can help them provide improved user interfaces. Because such rules or patterns can be used for user behavior analysis, they will also be beneficial to researchers.

Given the rapid growth of the smartphone industry, there is an urgent need for studies that will lead researchers, smartphone designers, and application developers to a better understanding of how users use apps on their smartphones. Most of the related studies reported in the literature are based on questionnaire surveys or lab experiments. However, a limitation of the use of questionnaires concerns assumptions that are too strong to be realistic; limitations of the use of the data collected in a lab experiment include that such data are usually only for a short period of time and also that such data are not very representative of users' real-life app-using behaviors.

Some researchers used computational methods rather than questionnaire surveys or lab experiments to gain a better understanding of how apps are used. Chittaranjan *et al.* applied machine learning with other models to location data and call logs in order to categorize users, and their results would allow apps to better meet each user's needs [2]. Falaki *et al.* investigated the relationship between user activities and power consumption, and they indicated that by knowing how a user interacts with his or her smartphone, user experience can be improved and energy drain can be predicted more accurately [3]. Xie *et al.* proposed a method to analyze app usage data in order to distinguish the operations performed by human users from the ones performed by devices infected by malware [4]. Burguera *et al.* used a crowdsourcing system to collect app traces, and they proposed a method to analyze app traces in order to identify apps infected by malware [5]. However, the above studies are limited to particular aspects of smartphone app usage as they are limited to certain types of apps or certain usage patterns of apps.

To overcome the aforementioned limitations, we use a different approach and apply data mining to log data in which user' real-life app-using behaviors over a long period of time are recorded. Considering the abundant information hidden in smartphone app usage log data, we believe that data mining is an ideal approach to the problem of extracting rules or patterns from such log data. From a computer science perspective, we demonstrate an application of data mining that can have an impact on the smartphone industry. Smartphone designers and application developers can benefit from the discovered rules and patterns and they can provide improved user interfaces. Our focus is on data preparation, which is the foundation of data mining. Once data are in the right format, it becomes straightforward to use existing tools

(possibly with minor modifications) to perform data mining tasks. After all, there is no need to reinvent the wheel if it exists and works. The critical point (and also what this paper presents) is how to find the right tool and how to use it in the right way. From a social science perspective, we demonstrate a data-driven approach that can help us know more about how users used apps on their smartphones. Researchers can use the discovered rules and patterns to reconstruct the context in which a user used his or her smartphone.

Eagle and Pentland collected a data set that captures how participants used certain functions on their feature phones [6]; Akoush and Sameh used the data set to study mobile user movement prediction [7]; Wang *et al.* worked on clustering instant messages on the data set [8]; Farrahi and Gatica-Perez studied the discovery of users' daily routines from the data set [9]. However, our interest in this paper is in smartphone app usage log data, and the flexibility provided by smartphones is greater than that provided by feature phones. Furthermore, this paper is not about users' moving behaviors or daily routines, and it is not limited to instant messages.

Smartphone app usage log mining is similar to but different from web log mining, because users are given a higher degree of freedom in installing and using apps on their smartphones. First, a smartphone app, such as a camera or recorder app, can work as an interface between its users and the environment, but a web page cannot do so. Second, web pages are linked through hyperlinks, but there are no hyperlinks between smartphone apps; in other words, there is no guidance to tell users how to switch between smartphone apps. Third, a session physically exists on and is managed by a web server and/or a web client (such as a browser), but there are no such concepts as sessions when users use apps on their smartphones. These differences make general web log mining techniques unsuitable for smartphone app usage log mining.

Fig. 1 presents the research workflow of the work presented in this paper. The rest of this paper is organized as follows: First, we describe the data preparation procedure in Section II. Then, we describe how to use the classic association rule mining algorithm and the traditional sequential pattern mining algorithm to mine smartphone app usage log data in Section III and Section VI, respectively. Finally, we conclude this paper and describe potential directions for future work in Section V.

II. DATA PREPARATION

Data preparation is important for data mining [10]-[12]. Before we describe the data preparation procedure, we describe the log data used in this paper. Chen *et al.* developed a platform to collect log data in which users' real-life app-using behaviors are recorded [13]. The platform is specifically designed for the Android operating system. The main advantage of the platform is that it is not restricted to some apps; it collects log data for apps that ever use the log service provided by the Android operating system. The main disadvantage of the platform is that it only contains data generated when apps were activated (started) and does not contain data generated when apps were deactivated (ended). The log data used in this paper were collected by the platform. The log data were generated by 25 users (in Taiwan) on 26

smartphones from September 2010 to March 2011.

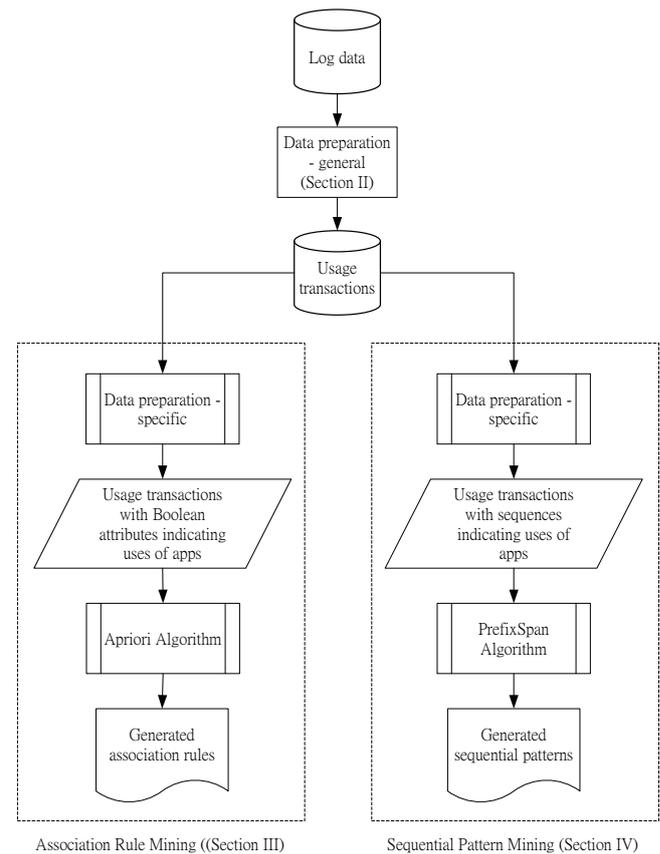


Fig. 1. The research workflow.

In the raw data set, every log entry includes a unique number, the identification number of a user, the identification number of a smartphone, a timestamp, and the package name of the app that was used. The two identification numbers can be used together to identify a unique operating unit. Although the raw data set is a rich data set, it is not suitable for the data mining tasks that we intend to perform. First and foremost, we cannot identify a usage transaction directly from the raw data set, where a usage transaction means a series of a user's app-using behaviors. Because a usage transaction is typically composed of several log entries, each of which can possibly be separated from others by hundreds of unrelated log entries in the raw data set, we first need to perform a series of data cleaning and transformation operations.

Given a database of log entries of smartphone app usage, we partition the database into a set of usage transactions, each of which corresponds to a series of uses of apps made by a user on a smartphone in a period of time. To begin with, we sort log entries in the raw data set by the identification numbers of users and by timestamps. Then, we define a usage transaction T as follows: T is composed of all the log entries corresponding to the same user and the same smartphone with the constraint that the difference in time between two consecutive log entries in T does not exceed 10 minutes. This threshold is determined empirically. Two consecutive log entries in a usage transaction are not necessarily next to each other in the raw data set.

Initially, there are more than 260 thousand log entries in the raw data set. Following the above definition, we obtain 25,856 usage transactions. Furthermore, there are 1,132 apps. Because some apps are much more popular than others, we

focus on the most frequently used apps. More specifically, we consider only apps that appear in more than 0.5% of the log entries in the raw data set. The threshold seems low, but it removes most apps and leaves only 33 apps. Three of the 33 most frequently used apps are launchers. On a smartphone running the Android operating system, the main task of a launcher is to help its user run other apps. It is neither surprising nor interesting to see that launchers are used more frequently than others, so we consider only the other 30 most frequently used apps. In 1,603 usage transactions, none of the 30 most frequently used apps appear. Although we exclude launchers in this paper, we would like to investigate how launchers are used with other apps in the future. The role a launcher plays is so special that we believe such an investigation will be beneficial to other researchers.

Consequently, in the data set prepared for the data mining tasks that will be described later in this paper, every usage transaction record includes a unique identification number, the identification number of a user, the identification number of a smartphone, the start timestamp (or the timestamp of the first log entry that belongs to the usage transaction), the end timestamp (or the timestamp of the last log entry that belongs to the usage transaction), and an array of apps corresponding to the usage transaction. The last part will be in different formats and have different contents: When we apply association rule mining to the data set (Section III), the last part is a set of apps that ever appear in a usage transaction. When we apply sequential pattern mining to the data set (Section VI), the last part is a sequence of apps that were used sequentially in a usage transaction. Attributes related to the used apps are specific attributes, and others are general attributes. Using a data format specific to a task is important for data mining practitioners, while in the literature we can hardly find papers where data formats suitable for smartphone app usage log mining are discussed in detail.

III. ASSOCIATION RULE MINING

Association rule mining was first introduced by Agrawal *et al.* in the early 1990's [14], [15]. Given a database of transactions, where a transaction contains a set of items purchased by a customer, association rule mining [14], [15] is to generate significant rules, and each rule associates the purchase of some items with the purchase of others in a transaction. For example, a rule " $A, B \rightarrow C$ " can be interpreted as follows: If a customer purchased items A and B in a transaction, then he or she would probably purchase item C in the same transaction.

The rule generation process consists of two main steps [15]: The first step is to identify frequent item sets, in each of which items were purchased together with a frequency greater than a pre-specified threshold. Such a frequency is called support, which is the probability that some items would be purchased together (in a transaction) with respect to all transactions. The threshold for support is called minimum support. The second step is to identify rules each of which indicates that if some of the items in a frequent item set were purchased together, then some of the rest of the items would be purchased with a probability greater than a pre-specified threshold. Such a probability is a conditional probability and is called confidence. Similarly, the threshold for confidence

is called minimum confidence. It is clear that minimum support and minimum confidence together determine the significance of the generated rules.

Association rule mining is finding applications in behavior analysis [16], such as plagiarism detection [17], students' enrollment [18], and recommendation [19].

To prepare data for association rule mining, we record what apps were used in a usage transaction, and neither the quantity of each of them nor the order of them would be recorded. Because we consider only the 30 most frequently used apps, we create 30 Boolean attributes to indicate what apps were used and what apps were not used in a usage transaction. Here the order of apps in a usage transaction is not important, so the order of these 30 Boolean attributes does not matter. By adding these 30 Boolean attributes to the data set, we can use the implementation of the classic Apriori algorithm [15] provided by WEKA [20] to generate rules.

Below we report rules that we obtain when using minimum support of 0.01 and minimum confidence of 0.3. In Table I, there are 22 rules sorted by their confidence values, as indicated by the last column. For every app, we only show the last part of its package name. Please note that, in Table I, the arrow " \rightarrow " (used by most implementations of the classic Apriori algorithm) does not imply that the app(s) in the left-hand side would be used before the app(s) in the right-hand side. In what follows, Table I included, katana is the package name of Facebook and pandahome2 is a game, while it is straightforward to map the other package names to apps (and to their main functions).

TABLE I: RULES GENERATED BY THE APRIORI ALGORITHM

Rule	Conf.
htcdialer \rightarrow phone	0.77
contacts \rightarrow phone	0.76
gm, htcdialer \rightarrow phone	0.76
mms, htcdialer \rightarrow phone	0.75
htcdialer, htcontacts \rightarrow phone	0.72
packageinstaller \rightarrow settings	0.71
phone, htcontacts \rightarrow htcdialer	0.63
htcontacts \rightarrow phone	0.57
album \rightarrow camera	0.54
htcontacts \rightarrow htcdialer	0.5
phone, gm \rightarrow htcdialer	0.5
phone, mms \rightarrow htcdialer	0.48
vending \rightarrow settings	0.4
phone \rightarrow htcdialer	0.38
camera \rightarrow album	0.38
htcontacts \rightarrow phone, htcdialer	0.36
camera \rightarrow katana	0.36
pandahome2 \rightarrow whatsapp	0.35
plurk \rightarrow browser	0.32
maps \rightarrow settings	0.32
maps \rightarrow katana	0.32
plurk \rightarrow whatsapp	0.31

Those rules containing HTCDialer and Contacts show the typical use of smartphones (or just mobile phones). Those rules containing Album and Camera show that a user switched between these two apps frequently. Smartphone users use their smartphones as digital cameras more often these days. Those rules containing Camera and Facebook show that photo sharing is an important function for

Facebook. This may also suggest that photo taking can be an important function for a Facebook-like app. If we consider those rules containing Maps and Facebook, and if we also consider that the log data used in this paper were generated between September 2010 and March 2011, then we may say that users took Maps as a compliment to location-based services of Facebook at that time.

Although those patterns indicate relationships between different uses of apps in the same usage transaction, we do not know what app would be used after some app was used. Therefore, we intend to apply sequential pattern mining to the data set. However, for smartphone app usage log mining, if we prepare data for sequential pattern mining in the traditional way and use the traditional sequential pattern mining algorithm directly, we will have patterns indicating the relationships between different uses of apps in different usage transactions, which will not be what we want.

IV. SEQUENTIAL PATTERN MINING

Sequential pattern mining, which was first introduced by Agrawal and Srikant in the mid 1990's, is to discover patterns indicating the sequence of items that were purchased in transactions by a customer [21]. It is to discover patterns indicating that, for example, if a customer purchased A in a transaction, then he or she would purchase B in the successive transaction.

PrefixSpan is an efficient algorithm for sequential pattern mining [22]. In the pattern generation process of the algorithm, a sequence of items is partitioned into a set of prefix items and a set of suffix items, and the suffix items are used to generate the projected databases for the prefix items. The projected databases are used to calculate the frequency of the pattern that begins with the prefix items. The process runs recursively: Each time a new prefix item is considered, and a pattern is grown with one item at a time. Similar to the case in association rule mining where infrequent item sets are removed at an early stage in the rule generation process, the patterns that begin with infrequent prefix items are removed at an early stage in the pattern generation process. The algorithm also uses minimum support as the threshold to remove infrequent patterns; it can generate patterns directly, so it does not need to use minimum confidence to generate rules from frequent item sets. Last but not least, the algorithm is suitable for our work, due to the fact that it prefers dense databases and our database is dense because the contained items (or apps) are all with high values of support.

Sequential pattern mining is also finding applications in behavior analysis. Below are some examples: Choi *et al.* used it in a recommendation system [23]. Chikhaoui *et al.* used it as part of a model proposed to discover and recognize human activities in pervasive environments [24]. Hu *et al.* used it in customer relationship management [25].

In order to prepare data for using sequential pattern mining for smartphone app usage log mining, we start from the data set prepared by the procedure described in Section II. For each usage transaction, we create a sequence where apps are recorded in the order that they were used. For example, if a user used Facebook, Gmail, and Facebook again, then we create a sequence composed of the following elements for the corresponding usage transaction: The first is Facebook, the

second is Gmail, and the third is Facebook. That is, the order of apps in a usage transaction is recorded literally.

Transactions in a data set prepared for being used by the traditional sequential pattern mining algorithm are grouped by users. We do not group transactions by users because our goal is to discover patterns of sequences in a usage transaction. A pattern indicating that a user used Facebook in the morning and Gmail in the afternoon is less interesting than a pattern indicating that Gmail would be used right after Facebook was used; the former is what we will obtain if we directly apply the traditional sequential pattern mining to data in the traditional format, while the latter is what we will obtain if we apply the traditional sequential pattern mining algorithm to data transformed into the format described in this section.

In order to generate patterns indicating what app would be used after some app was used, first we add a sequence of the used apps to each usage transaction, then we make minor modifications (mainly related to data processing) to the implementation of the traditional PrefixSpan algorithm [22] provided by SPMF [26], and finally we apply the modified implementation to the prepared data set. The minimum support is set to 0.01.

The number of patterns generate by a sequential pattern mining algorithm is usually larger than the number of rules generated by an association rule mining algorithm, if we use the same minimum support for both of them and set the minimum confidence to a reasonable value for association rule mining. For example, we observe 527 patterns when we use the minimum support of 0.005 (compared to 77 rules that we observe when using the same minimum support and set the minimum confidence to 0.3 in association rule mining). The reason is that an app can appear several times in a sequential pattern but not in a rule generated by an association rule mining algorithm.

Below we report some of the generated patterns. Inspired by the work presented by Chen *et al.* [27], we report patterns regarding Facebook, Camera, and Album in Table II. In Table II (and in the following tables), patterns are sorted by their support values, as indicated in the last column. Please note that, unlike the case in Section III (association rule mining), the arrow " \rightarrow " in this section implies that the app in the left-hand side would be used before the app in the right-hand side.

TABLE II: SEQUENTIAL PATTERNS REGARDING FACEBOOK (KATANA), CAMERA AND ALBUM

Pattern	Sup.
camera \rightarrow album	0.02
camera \rightarrow katana	0.02
album \rightarrow camera	0.01
katana \rightarrow camera	0.01

From Table II, we can see that the probability of using Camera first and then Album is about twice of the probability of using Album first and then Camera. Similarly, the probability that a user used Camera before using Facebook is about twice of the probability that he or she used Facebook before using Camera. These findings can hardly be made if we simply use association rule mining. They may also suggest that better photo browsing and sharing functions can

be integrated with a Camera-like app.

We report patterns showing how users made phone calls in Table III. Patterns in the table are related to the typical use of smartphones (or just mobile phones). Surprisingly, as we can see from Table III, Contacts and HTCContacts were used to make phone calls by users (who contributed the log data used in this paper) less often than HTCDialer. This may initiate a review and redesign of Contacts-like apps.

TABLE III: SEQUENTIAL PATTERNS RELATED TO MAKING PHONE CALLS

Pattern	Sup.
htcdialer → phone	0.1
htccontacts → phone	0.02
contacts → phone	0.02

We report patterns regarding apps for social networking services in Table IV. Although the support values in the table are small individually, the sum of them is close to 0.1, which means that apps for social networking services were used in about 10% of usage transactions (between September 2010 and March 2011). Undoubtedly, Facebook has been a center for social conversations since its first day. Table IV gives us more insights into how users used apps for social networking services. Users switched between these apps equally frequently. For example, the probability that a user used Plurk (for microblogging) before using WhatsApp (for instant messaging) is about the same as the probability that a user used WhatsApp before using Plurk. Since Plurk was released prior to WhatsApp and the sequential patterns corresponding to these two are almost equal, the patterns reported in Table IV have foreseen the rise of WhatsApp.

TABLE IV: SEQUENTIAL PATTERNS REGARDING APPS FOR SOCIAL NETWORKING SERVICES

Pattern	Sup.
katana → gtalk	0.02
katana → plurk	0.01
katana → whatsapp	0.01
plurk → katana	0.01
plurk → whatsapp	0.01
whatsapp → katana	0.01
whatsapp → plurk	0.01

We report patterns showing what users used before or after using Brower in Table V. Patterns in the table show that Browser was used in more than 10% of usage transactions (between September 2010 and March 2011).

TABLE V: SEQUENTIAL PATTERNS RELATED TO APPS USED BEFORE OR AFTER BROWSER.

Pattern	Sup.
katana → browser	0.03
browser → katana	0.02
gmail → browser	0.02
browser → gmail	0.01
browser → phone	0.01
browser → plurk	0.01
plurk → browser	0.01

We report patterns regarding successive use of an app in Table VI. The following is an example that fits the first

pattern in the table: In a usage transaction, a user used Facebook and went back to the home screen or used another app that was not one of the most frequently used apps, and then he or she used Facebook again; the difference in time between the two uses of Facebook is less than 10 minutes.

TABLE VI: SEQUENTIAL PATTERNS REGARDING SUCCESSIVE USE OF AN APP

Pattern	Sup.
katana → katana	0.04
whatsapp → whatsapp	0.04
browserp → browser	0.03
gmail → gmail	0.03
camera → camera	0.02
katana → katana → katana	0.02
whatsapp → whatsapp → whatsapp	0.02
album → album	0.01
plurk → plurk	0.01
browser → browser → browser	0.01

The generated patterns, as shown above, can be used to help design one-stop apps, by which users could enjoy more functions with fewer switches between apps. For example, social apps can emphasize on not only photo sharing but also photo taking and photo editing; camera or camera-like apps can consider enhancing functions that make easier browsing, viewing, and editing photos, and additionally they can consider enhancing functions that better integrate with social networking services.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we presented our work on smartphone app usage log mining. We applied data mining to the log data, in which there are more than 260 thousand log entries that recorded how 25 users used popular apps (that were used the most frequently) on 26 smartphones in their work and daily lives from September 2010 to March 2011, and we put our focus on data preparation. We described how to prepare data for utilizing the classic association rule mining algorithm and the traditional sequential pattern mining algorithm in order to mine smartphone app usage log data, and we discussed the generated rules and patterns that can help researchers, smartphone designers, and application developers gain a better understanding of how apps are used.

As for our future work, we would like to perform further analysis on the log data in order to have findings even more beneficial to researchers, designers, and developers. We plan to extend the traditional sequential mining algorithm so as to provide a tool that can discover more helpful patterns.

ACKNOWLEDGMENT

The authors gratefully acknowledge the critical review by anonymous reviewers on the manuscript of this paper.

REFERENCES

- [1] M. Raento, A. Oulasvirta, and N. Eagle, "Smartphones: an emerging tool for social scientists," *Sociological Methods Research*, vol. 37, no. 3, pp. 426-454, February 2009.
- [2] G. Chittaranjan, J. Blom, and D. Gatica-Perez, "Mining large-scale smartphone data for personality studies," *Personal and Ubiquitous Computing*, vol. 17, issue 3, pp. 433-450, March 2013.

- [3] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, "Diversity in smartphone usage," in *Proc. 8th Int'l Conf. Mobile Systems, Applications, and Services*, San Francisco, CA, USA, 2010, pp. 179-194.
- [4] L. Xie, X. Zhang, J. Seifert, and S. Zhu, "pBMDS: a behavior-based malware detection system for cellphone devices," in *Proc. 3rd ACM Conf. Wireless Network Security*, Hoboken, NJ, USA, 2010, pp. 37-48.
- [5] I. Burguera, U. Zurutuza, and S. Nadjm-Tehrani, "Crowdroid: behavior-based malware detection system for android," in *Proc. 1st ACM Wkshp. Security and Privacy in Smartphones and Mobile Devices*, Chicago, IL, USA, 2011, pp. 15-26.
- [6] N. Eagle and A. Pentland, "Reality mining: sensing complex social systems," *Journal Personal and Ubiquitous Computing*, vol. 10, issue 4, pp. 255-268, March 2006.
- [7] S. Akoush and A. Sameh, "Mobile user movement prediction using Bayesian learning for neural networks," in *Proc. Int'l Conf. Wireless Communications and Mobile Computing*, Honolulu, HI, USA, 2007, pp. 191-196.
- [8] L. Wang, Y. Jia, and W. Han, "Instant message clustering based on extended vector space model," in *Proc. 2nd Int'l Conf. Advances in Computation and Intelligence*, Wuhan, China, 2007, pp. 435-443.
- [9] K. Farrahi and D. Gatica-Perez, "What did you do today?: discovering daily routines from large-scale mobile data," in *Proc. 16th Int'l Conf. Multimedia*, Vancouver, Canada, 2008, pp. 849-852.
- [10] R. Cooley, B. Mobasher, and J. Srivastava, "Data preparation for mining world wide web browsing patterns," *Knowledge and Information Systems*, vol. 1, no. 1, pp. 5-32, February 1999.
- [11] M. Munk, J. Kapusta, and P. Švec, "Data preprocessing evaluation for web log mining: reconstruction of activities of a web visitor," *Procedia Computer Science*, vol. 1, issue 1, pp. 2273-2280, May 2010.
- [12] A. Rajenderan, "Data preparation for web mining – a survey," *International Journal of Advanced Computer Research*, vol. 2, no. 6, pp. 1-7, December 2012.
- [13] P. M. Chen, C. H. Chen, W. H. Liao, and T. Y. Li, "A service platform for logging and analyzing mobile user behaviors," in *Proc. 6th Int'l Conf. E-Learning and Games*, Taipei, Taiwan, 2011, pp. 78-85.
- [14] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Record*, vol. 22, iss. 2, pp. 207-216, June 1993.
- [15] R. Agrawal, and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th Int'l Conf. Very Large Data Bases*, Santiago, Chile, 1994, pp. 487-499.
- [16] J. Han, H. Cheng, D. Xin, and X. Yan, "Frequent pattern mining: current status and future directions," *Data Mining and Knowledge Discovery*, vol. 15, issue 1, pp. 55-86, August 2007.
- [17] M. Drl k, M. Munk, and J. Skalka, "Usage analysis of system for theses acquisition and plagiarism detection," *Procedia Computer Science*, vol. 3, pp. 866-871, 2011.
- [18] Z. Abdullah, T. Herawan, N. Ahmad, and M. M. Deris, "Extracting highly positive association rules from students' enrollment data," *Procedia - Social and Behavioral Sciences*, vol. 28, pp. 107-111, 2011.
- [19] A. A. Kardan and M. Ebrahimi, "A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups," *Information Sciences*, vol. 219, pp. 93-110, January 2013.
- [20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The WEKA data mining software: an update," *SIGKDD Explorations*, vol. 11, issue 1, June 2009.
- [21] R. Agrawal, and R. Srikant, "Mining sequential patterns," in *Proc. 11th Int'l Conf. Data Engineering*, Taipei, Taiwan, 1995, pp. 3-14.
- [22] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu, "PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth," in *Proc. the 17th Int'l Conf. Data Engineering*, Heidelberg, Germany, 2001, pp. 215-224.
- [23] K. Choi, D. Yoo, G. Kim, and Y. Suh, "A hybrid online-product recommendation system: combining implicit rating-based collaborative filtering and sequential pattern analysis," *Electronic Commerce Research and Applications*, vol. 11, issue 4, pp. 309-317, July–August 2012.
- [24] B. Chikhaoui, S. Wang, and H. Pigot, "ADR-SPLDA: activity discovery and recognition by combining sequential patterns and latent Dirichlet allocation," *Pervasive and Mobile Computing*, vol. 8, issue 6, pp. 845-862, December 2012.
- [25] Y. H. Hu, T. C. K. Huang, and Y. H. Kao, "Knowledge discovery of weighted RFM sequential patterns from customer sequence databases," *Journal of Systems and Software*, vol. 86, issue 3, pp. 779-788, March 2013.
- [26] P. Fournier-Viger, A. Gomariz, A. Soltani, and T. Gueniche. (2013). SPMF: Open-Source Data Mining Platform. [Online]. Available: <http://www.philippe-fournier-viger.com/spmf/>
- [27] P. M. Chen, H. Y. Wu, C. Y. Hsu, W. H. Liao, and T. Y. Li, "Logging and Analyzing Mobile User Behaviors," in *Proc. 2nd Int'l Symp. Cyber Behavior*, Taipei, Taiwan, 2012.



Woan-Rou Tseng received her M.S. degree from the Department of Computer Science at the National Chengchi University, Taipei, Taiwan, in 2013; she received her B.S. degree from the same department in 2010.



Kuo-Wei Hsu earned his Ph.D. from the Department of Computer Science and Engineering at the University of Minnesota, Minneapolis, MN, USA, 2011; he obtained his M.S. degree in Computer Science and Information Engineering from the National Taiwan University, Taipei, Taiwan, 2001; he obtained his B.S. degree in Electrical Engineering from the National Chung Hsing University, Taichung, Taiwan, 1999.

He is currently an assistant professor in the Department of Computer Science at the National Chengchi University, Taipei, Taiwan. Before he entered the Ph.D. program, he worked as an information engineer in the National Taiwan University Hospital, Taipei, Taiwan. His current research interests include data mining, database systems, and software engineering. Dr. Hsu is currently a member of ACM and IEEE.