ICIC Express Letters
Part B: Applications
Volume **9**, Number **6**, June **2018**

ICIC International ©2018 ISSN 2185-2766
pp. **607–614**

ANALYZING MOBILE APPLICATION LOGS USING PROCESS MINING TECHNIQUES: AN APPLICATION TO ONLINE BOOKSTORES

Seungeun Kim[1] and Dongsoo Kim[2,*]

[1]Department of IT Distribution and Logistics
[2]Department of Industrial and Information Systems Engineering
Soongsil University
369 Sangdo-Ro, Dongjak-Gu, Seoul 06978, Korea
sungkim@soongsil.ac.kr; *Corresponding author: dskim@ssu.ac.kr

Received December 2017; accepted March 2018

Abstract. *A vast amount of user logs is recorded while users utilize various mobile applications. As lots of companies pay no attention to log data, logs usually cannot provide useful meaning or insights. In order to improve user experiences and acquire valuable insights, we propose a framework for analyzing mobile application logs using process mining techniques. Previous studies based on surveys or third party applications have limitations such as data collection problems or reliability issues. The proposed framework enables a company to understand the behavior of diverse users and to identify problems by setting up various performance questions. We applied the proposed method to a virtual online bookstore and derived analysis questions in order to helpful insights. The analysis results based on process discovery, pattern discovery and performance analysis of application logs can help improve user interface deployment of the mobile application.*
Keywords: Application log analysis, Process mining, Mobile application, Mobile commerce, User behavior analysis

1. Introduction.

Mobile commerce (m-commerce), which is expanding with increasing penetration rate of smartphones, is a monetary transaction that takes place in mobile devices. According to a survey on the use of mobile shopping in 2015, more than half of Korean smartphone users use mobile shopping, and typically they use mobile applications installed on their devices [10]. However, the application market is very competitive [11]. In addition, one of the m-commerce features is that it uses 'small screen' mobile devices compared to desktop PCs. Therefore, there is a limitation of information amount that can be provided in one page compared to a desktop web page [7]. Consequently, companies are interested in managing User Interface (UI) that can increase the purchase convenience based on effective interaction between applications and users.

There are three types of researches on user application usage, depending on data collection method. First, previous studies conducted researches to improve UI by indirectly investigating users' experiences through surveys and interviews [9]. However, this approach is time consuming and costly, and depends on users' memories. The quality of the results on the data depends on the respondents' willingness to respond. Second, some researchers used the third-party applications in order to obtain permissions of the applications to be analyzed, stored the data in the database, and analyzed the user behavior through the query [8]. Due to the constraints of data collection depending on the server's permission, this limits the analysis scope. The third type is to use application log data directly. Some researchers collected logs of Android and Window platforms and analyzed usage according to application types [4]. This method of using objective and quantitative data does not have risk of data bias. In addition, a large number of logs can be used to discover usage patterns through various criteria.

In this paper, we propose a framework for analyzing m-commerce user behavior using process mining techniques. Because process mining analyzes actual logs generated in the real world, the proposed framework will help m-commerce companies identify UI issues, determine the deployment of application contents, and acquire valuable customer insights.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents a framework for application log analysis using process mining. Section 4 describes an application case of online bookstores using the proposed framework. Finally, Section 5 concludes the paper and suggests future research directions.

2. **Related Work.** This section presents the backgrounds of the proposed work. Basically, process mining aims to discover a process model from event logs. It tells companies, how employees and/or processes are actually working. And it is possible to improve processes for companies [15]. Because of these advantages, process mining is being studied in various fields such as shipyards and mobile games [3,6]. The three main types of process mining are *discovery*, *conformance checking*, and *enhancement*. First, discovery is creating a process model from the event logs, which allows analysts to visualize complex business processes in the real world. Second, conformance checking compares the standard process model of organizations with the discovered process in order to determine suitability. Third, enhancement modifies the current process model to be realistic (Repair) and adds information such as timestamps, resources, and bottlenecks to the model (Extension) [12].

There are different analytical perspectives of process mining: *control-flow*, *time*, *organizational*, and *case perspectives*. Through the process discovery, control-flow perspective focuses on ordering activities. Time perspective focuses on timestamps and frequencies of activities. For example, it enables to analyze bottlenecks, measure utilization of resources and predict the performance of ongoing cases. Organizational perspective concentrates on the information of involved resources, and the relationship between the resources. The results can be structured as a social network. Finally, case perspective can be used to understand decision making through the data involved in the case. In addition, information about customers can be useful in this perspective [13]. In this study, we focus on time and case perspectives.

Previous studies that analyzed application logs have used several techniques. Some researchers used machine learning to analyze user usage patterns [5]. Other researchers analyzed the usage patterns using the iPhone logs [1]. However, the above studies are pattern-oriented researches of user's specific applications. This study not only provides pattern analysis and performance analysis, but also analyzes research questions that companies want to know, focusing on the user's behavior process.

3. **Mobile Application Logs Analysis Using Process Mining.** This section describes the proposed framework of application log analysis and how process mining is used for the analysis. In order for users to have good experiences and to improve the rate of revisit and purchase, m-commerce companies should immediately identify and improve any issues so that users can continue to choose the contents of the application. To prevent user deviation, it is important to collect user behavior data for all users and to find satisfaction and dissatisfaction points, and to take appropriate actions based on the analysis results. In this regard, process mining can extract a process and find points of improvement from the application logs. The proposed framework is composed of 5 steps as shown in Figure 1. Each step is described in detail in the following sections.

3.1. **Defining research questions.** Typically, an analysis begins by defining questions of interest. At first, we should decide what to analyze. These questions need to be discussed with domain experts and m-commerce companies. It also requires that collected logs should resolve the questions to understand user behavior and identify problems. Then a scope of analysis is determined from the converged questions. Examples of questions

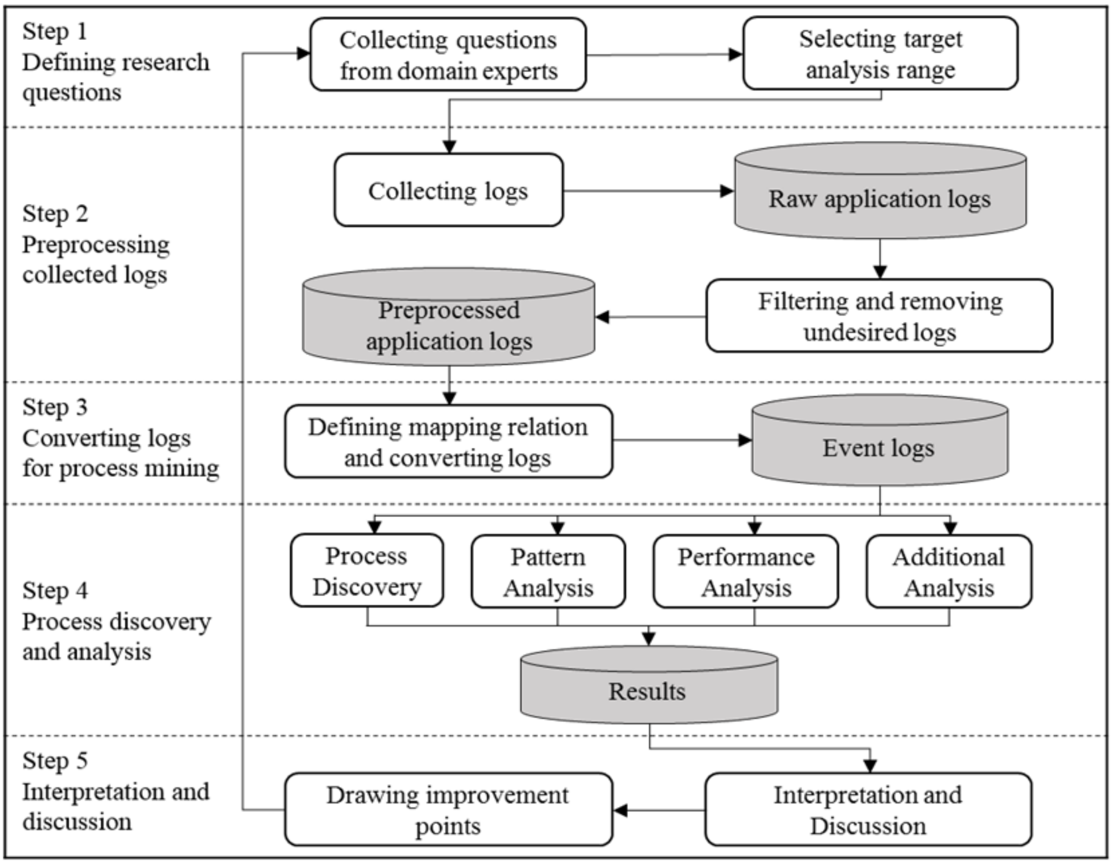


FIGURE 1. Framework for the application log analysis using process mining

include the following. What are the usage patterns of new users? What is the difference between a first-time buyer and a user with a past purchasing history? How do users behave differently based on their membership grade?

3.2. **Preprocessing collected logs.** Application log data can generally be classified into two types. The application installation logs are collected once, when users launch the application for the first time after installing it on their devices. These logs remain unchanged, whereas *application execution logs* change from time to time as users utilize the application. An activity, which is one of application components, is a main class that displays the UI on the full screen and handles events and a new UI on the whole area of the screen. The logs related to application execution are collected based on the activity life cycle. In this study, we analyze application execution logs. We extract the logs from the database and filter them by refining the information related to the user behavior.

3.3. **Converting logs for process mining.** The filtered logs that correspond to the scope of analysis in Step 1 should be converted to event log data format for the proper process mining tool. For instance, a device key is mapped to the *Case ID* of the process mining to distinguish actions such as join, login, product detail page view, and payment. The series of actions can be converted into the *Activity* and the time of occurrence of these activities can be changed to the *Timestamp*. This allows analysts to prepare to discover the entire process of the application users.

Since deriving the whole process is one aspect of process mining, additional analysis needs to be carried out, to obtain meaningful results. This is to perform process mining techniques including application user specific data. Examples of items in logs include customer indicators such as number of visits, number of items in a shopping cart, number of purchases, and behavior of first-time visitors. These log entries are expressed as the

TABLE 1. Mapping from application logs to process mining event logs

Application log	Process mining event log
Device key	Case ID
User actions (join, search, payment, etc.)	Activity
Completed time of each user action	Timestamp
Customer indicators (number of visits, average purchase amount, membership grade, etc.)	Other

'*Other*' attribute in the process mining event log. Table 1 shows the mapping results. After defining the mapping relationship, the data need to be converted to a specific format such as CSV (Comma-Separated Value) and MXML (Mining eXtensible Markup Language).

3.4. **Process discovery and analysis.** This step includes four main analysis methods, that is, process discovery, pattern analysis, performance analysis, and additional analysis.

Process discovery: Analysts can import event logs into a process mining tool such as Disco released by Fluxicon. Then, the tool automatically detects data columns. Because core feature of process mining is discovering the process model from the imported logs, we can obtain an actual process model of application users [2]. Also, we can adjust the level of activity and path parameter values. In order to find the most common process that represents the activities that occur most frequently in the overall data, we control the two parameters.

Pattern analysis: We can identify major patterns using analysis methods such as performance sequence diagram analysis and advanced dotted chart analysis in ProM [14]. Through these analyses, case patterns can be found according to the sequence and frequency of events. By measuring the various patterns and frequencies that appear in the event logs, we can understand the behavior patterns of users.

Performance analysis: This analysis enables the analyst to measure the execution time, maximum, minimum, mean, and median for each activity. Thus, we can find valuable information such as long-lasting, short or almost unseen application pages.

Additional analysis: This is directly related to the scope of the analysis determined in Step 1. In order to understand why users behave differently in applications, we need to analyze customer related data. The case perspective of process mining is related to this analysis. Therefore, it uses *Other* columns. After selecting appropriate techniques according to analysis questions, we conduct proper process mining analysis methods.

3.5. **Interpretation and discussion.** In this step, we combine various analysis results from Step 4. And then, we can recognize answers of the questions. After that, we discuss and interpret with application managers or domain experts, and find improvement points. Finally, the improvement points can be reflected in the early stages of question collection. As a result, the proposed framework has a circular form.

4. **Application Case.** In this section, we present a sample business log and the analysis results to explain the process of implementing the analysis framework in this study. Since the real logs of companies are constrained to obtain and release the data, we randomly generated artificial log data.

4.1. **Explanation of example log data.** We applied the proposed framework to a virtual bookstore application. It consists of 80 activities such as join, login, logout, and membership withdrawal. Table 2 shows a fragment of the sample event logs.

TABLE 2. Event logs of a mobile bookstore application

caseID	Activity	Timestamp (completed)	Membership grade	···
187_G	Login	2017-06-03-22-15	Gold	···
65_VV	Pay	2017-06-03-22-16	VVIP	···
67_G	Reused Market	2017-06-03-22-16	Gold	···
···	···	···	···	···

4.2. **Analysis result.** According to Step 1, analysis questions of this bookstore application can be established as follows. 1) How do users interact with less frequently used shortcut menus on the home screen? And are there customer deviations? 2) How does the frequency of payment, shopping cart, search, and detailed page view activities differ depending on the customer's membership level? 3) How long does it take to view certain content based on membership level? In Steps 2 and 3, randomly generated application log data were converted into CSV and MXML files.

The results of *process discovery*, *pattern analysis*, and *performance analysis* of Step 4 are shown in Figure 2. First, we derived behaviors of bookstore application users as a whole process model. Also, we found the main process by adjusting activity and path parameters to remove activity and path with relatively low frequency. As complex processes are discovered, additional analyses are needed, not just process discovery.

In *pattern analysis*, the performance sequence diagram analysis method was used to derive six patterns with frequency. There are numerous kinds of patterns because some activities having no casual relations can be freely performed according to the purpose of users and some activities can be repeatedly performed. And the advanced dotted chart analysis method was performed using relative time and sorting according to actual duration. The analysis result show that a typical start activity is the login activity. In addition, other users who do not log in to the application have a relatively short process instance.

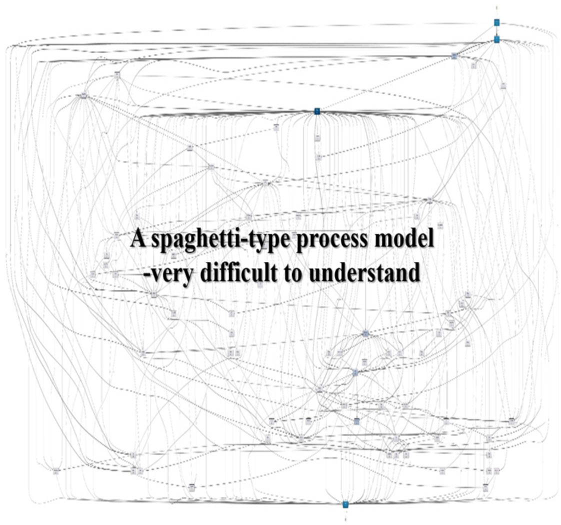
The processing time of activities can be visualized using the basic performance analysis method. The analysis results show that users have spent the longest time in some activities such as writing book reviews and viewing book news. The activity frequency is high in the following order: product details page, login, cart page, and search. The analysis also shows that users rarely use pages such as viewing free gifts on the home screen and checking affiliate card benefits.

Figure 3 shows the results of *additional analysis* to solve the analysis questions in Step 1. We handled the first question that concerns shortcut menus on the application home screen. Through *performance analysis*, we can find pages with low frequencies among 15 shortcut menus. The application home page is the first page to be visited by users. The shortcut menus need to be useful as well as user-friendly. For example, let us consider that a user selects a shortcut menu called 'Today's Book' on the home page. If the content is interesting, users can read the detailed pages and book reviews, which provides the opportunity of purchasing books. Consequently, in this respect, Question 1 compares two shortcut menus for home book cast and home free gifts that are relatively less frequent. Using Disco's attribute filter, we can find that users who touch home book cast menu tend to terminate the application while users who touch home free gifts tend to search products.

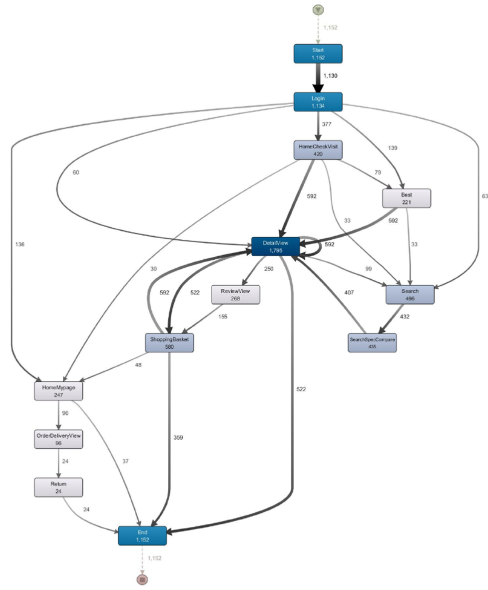
The next question is concerned with specific activities depending on the membership grade. The result shows that purchase activity that directly relates to profits is the highest in VVIP grade, and the rate of searching for products is higher when the membership grade is higher.

The last question is about how stay time differs depending on the membership grade on a particular page. This analysis aims to identify the impact of membership grade on new

Process discovery of application users

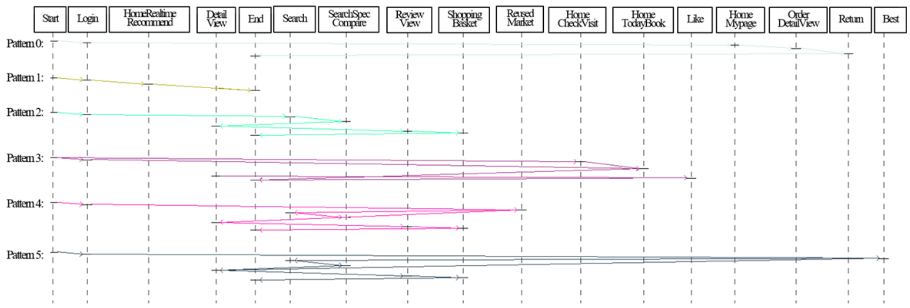


Activities and paths 100%

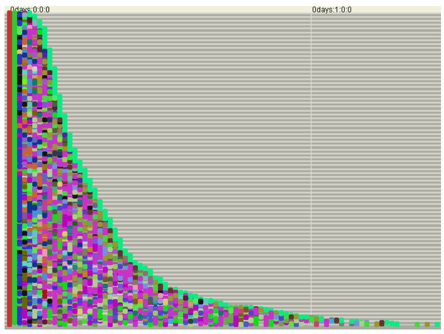


Activities and paths 10%

Pattern analysis

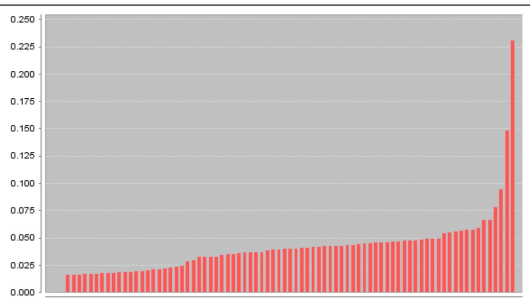


Performance sequence diagram

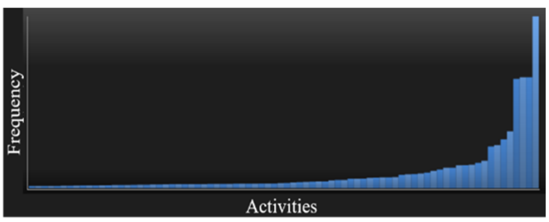


Advanced dotted chart analysis

Performance analysis



Processing (task) bar chart



Histogram for frequency of activities

FIGURE 2. Results of process discovery, pattern analysis and performance analysis

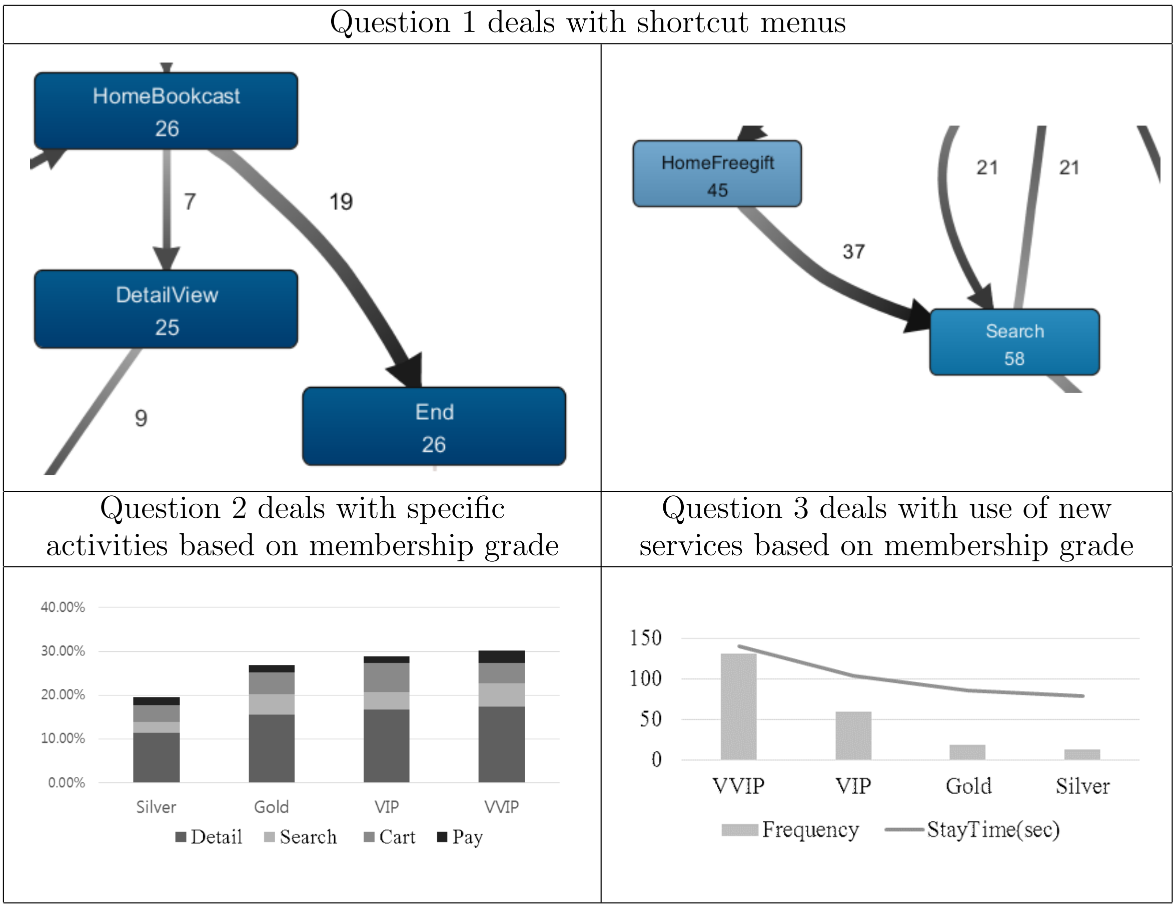


FIGURE 3. Result of additional analysis

services. We compared the stay time based on the membership grade on the Book and Flower page. It is a service that ships books and flowers to encourage regular purchases. The analysis result shows that the higher the membership grade is, the longer the stay of new services is and the higher the frequency is.

4.3. **Discussion.** As part of Step 5, we can discuss the problems of the application and draw up improvement points according to the analysis results. Based on the converged analysis questions, we may suggest that shortcut menus that are less useful or lead to user deviation in the application should be changed or removed. According to Question 2, lots of members of silver grade tend to view product detail pages, but then close the application. Therefore, we may suggest a promotion strategy on the product detail pages in order to increase the purchase rate of silver grade members. Also, we can ensure that the newly designed service is being used by the original target customers.

Taking these results into consideration, the m-commerce application can identify usage patterns depending on the membership grade and analyze the performance of shortcut menus and new services. This analysis is derived from the analysis questions, and it is possible to find other insights based on a variety of questions. Therefore, this study provides an opportunity to improve the UI of the application, since the application needs to ensure that the contents continue to be optimized and tailored to the users.

5. **Conclusions.** In this study, we presented an innovative approach to application user behavior analysis using process mining techniques. The proposed analysis framework can discover patterns of users' behaviors from the process-centric perspective, and analyze the performance of application UI in order to identify problems. M-commerce companies can identify UI issues, determine the deployment of contents, and acquire customer insights using the proposed framework. This work is meaningful because it provides insights

to understand complex and diverse users while overcoming the limitations of specific application user pattern analysis in the previous studies. Also, we widened the application area of process mining.

Due to the limitations of the actual log usage, we created a sample log and performed an experiment. We hope to use real application log data in order to provide more realistic results and various UI performance effects. In the future, we will conduct in-depth research to understand mobile application users through consultation with domain experts.

Acknowledgment. This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A1B05029080). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] B. Park, I. Cho, Y. Cheon and W. Lee, A log analysis of smartphone application usage: Focusing on domestic iPhone users, *Journal of the HCI Society of Korea*, vol.2011, no.1, pp.1051-1055, 2011.
[2] C. W. Günther and A. Rozinat, Disco: Discover your processes, *BPM 2012*, 2012.
[3] D. Lee, J. H. Park and H. Bae, Comparison between planned and actual data of block assembly process using process mining in shipyards, *Journal of Society for e-Business Studies*, vol.18, no.4, pp.145-167, 2013.
[4] F. Hossein et al., Diversity in smartphone usage, *Proc. of the 8th International Conference on Mobile Systems, Applications, and Services*, pp.179-194, 2010.
[5] G. Chittaranjan, J. Blom and D. Gatica-Perez, Mining large-scale smartphone data for personality studies, *Personal and Ubiquitous Computing*, vol.17, no.3, pp.433-450, 2013.
[6] H. Kwon and D. Kim, A method for churn analysis of new users of mobile games using process mining, *ICIC Express Letters, Part B: Applications*, vol.7, no.8, pp.1797-1803, 2016.
[7] J. Nilson and R. Budiu, *Mobile Usability*, MITP-Verlags GmbH & Co. KG, 2013.
[8] J. W. An, D. Moon and D. S. Cho, Design and implementation of usage pattern analysis application for smart phone, *Proc. of Korea Computer Congress*, pp.1811-1813, 2015.
[9] J. Y. Lee and S. I. Kim, A study on development of travel application UI design for improving user experience, *Journal of Digital Convergence*, vol.15, no.2, pp.369-374, 2017.
[10] Korea Internet & Security Agency, *Mobile Internet Usage Survey in 2015*, 2015.
[11] P. Racherla, C. Furner and J. Babb, *Conceptualizing the Implications of Mobile App Usage and Stickiness: A Research Agenda*, http://ssrn.com/abstract=2187056, 2012.
[12] W. M. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer Heidelberg Dordrecht, 2011.
[13] W. M. van der Aalst et al., Process mining manifesto, *International Conference on Business Process Management*, pp.169-194, 2012.
[14] W. M. van der Aalst et al., ProM: The process mining toolkit, *BPM (Demos)*, pp.1-49, 2009.
[15] W. M. van der Aalst, T. Weijters and L. Maruster, Workflow mining: Discovering process models from event logs, *IEEE Trans. Knowledge and Data Engineering*, vol.16, no.9, pp.1128-1142, 2004.