International Journal of Internet, Broadcasting and Communication Vol.7 No.1 20-22 (2015)
http://dx.doi.org/10.7236/IJIBC.2015.7.1.20

IJIBC 15-1-5

Detecting smartphone user habits using sequential pattern analysis

Lu Dang Nhac[1], Nguyen Thu Trang[2], Nguyen Thi Hau[2], Nguyen Ha Nam[2] , Gyoo Seok Choi[3*]

[1]*Academy of Journalism and Communication*
nhacld@gmail.com

[2]*University of Engineering and Technology, Vietnam National University in Hanoi*
trangnt84@gmail.com,{nguyenhau,namnh}@vnu.edu

[3*]*Department of Computer Science, Chungwoon University, Incheon, Korea*
lionel@chungwoon.ac.kr

Abstract

Recently, the study of smart phone user habits has become a highly focused topic due to the rapid growth of the smart phone market. Indeed, sequential pattern analysis methods were efficiently used for web-based user habit mining long time ago. However, by means of simulations, it has been observed that these methods might fail for smart phone-based user habit mining. In this paper, we propose a novel approach that leads to a considerably increased performance of the traditional sequential pattern analysis methods by reasonably cutting off each chronological sequence of user logs on a device into shorter ones, which represent the sequential user activities in various periods of a day.

Keywords: sequential pattern mining, smart phone, user habits, application log

1. Introduction

In recent years, the use and the functionalities of smart phones are quickly increasing. Analyzing the repeated behaviors of users relying on data collected from these devices becomes an important research field as it brings many business benefits such as the development of a targeted advertising system, oran intelligent recommendation system, etc. An emerging approach, analyzing user logs recorded on smart phone devices, has been successfully applied for detecting user habits by relying on association rule mining algorithms[1, 2]. Moreover, a number of frequent sequential log pattern mining algorithms were proposed and efficiently applied for finding user habits from web-based user transaction data [3, 4]. However, they have not been yet tested on smart phone-based user transaction data. For this reason, we employed these methods on our simulated datasets by simply taking the timely ordered sequences of user logs on a device during a predefined number of days as input, and found no frequent pattern in many cases. This motivated us to investigate an approach of preprocessing the aforementioned input data so that the performance of the traditional sequential pattern mining algorithms can be improved for such smart phone-based user transaction data.

Manuscript Received: Dec, 1, 2014 / Revised: Jan. 12, 2015 / Accepted: Feb. 6, 2015
Corresponding Author: lionel@chungwoon.ac.kr
Tel: +82-32-770-8250
Dept. of Computer Science, Chungwoon University, Incheon, Korea

2. Related works

An algorithm for finding frequent sequential patterns in a web-based transaction database was introduced in 1995 [3]. Considering a database being a set of chronological transaction sequences, it aims to detect all sequential transaction patterns occurring in at least a predefined percentage of the sequences belonging to a particular database. Later, a method combining the depth first traversal of search space with the vertical bitmap representation to store each sequence has been proposed [4]. It was shown to be fast and efficient for databases containing long sequential patterns as it incrementally outputs frequent patterns in online fashion.

3. A method for detecting frequent sequential log patterns

In this work, a database contains chronological sequences of application logs that were recorded on a set of devices. We propose to divide each sequence of application logs belonging to a device, called a primary sequential pattern PS, into an ordered set of non-overlapping subsequences $<SS_1, SS_2,\ldots, SS_k>$ such that the time gap between any two consecutive subsequences is greater than or equal to a predefined value, called a time gap threshold, where $PS = SS_1 \cup SS_2 \cup \ldots \cup SS_k$, $SS_1 \cap SS_2 \cap \ldots \cap SS_k = \varnothing$, $k \in Z^+$.Such subsequences are called *short sequential patterns.* Our goal is to estimate a time gap threshold, called Th, such that we can find as many as possible number of frequently occurred sequential patterns from a set of short sequential patterns obtained by cutting off all primary sequential patterns in the database given the time gap threshold Th. The proposed method is included in the framework of discovering user habits from smart phone user logs as describing in Figure 1.

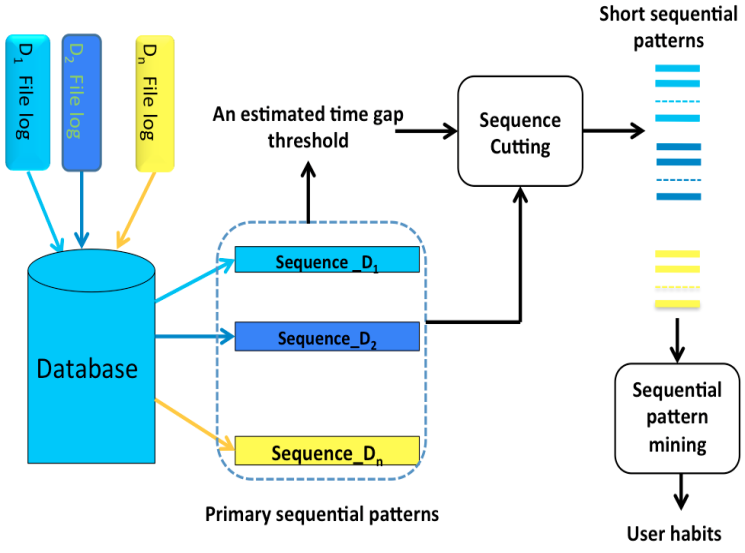

Figure1. A framework for detecting user habits from smart phone user logs

4. Conclusion

We have proposed an efficient approach for preprocessing application logs recorded on a set of smart phones, that lead to a significantly better performance of the traditional sequential pattern mining algorithm as comparing with the approach using the traditional input patterns. Nonetheless, the experiments were just carried out on simulated datasets. Hence, it is expected to perform additional experiments on the real data in the future. Moreover, our future works will consider more details of users 'information, for example their gender, age, etc. so that the detected user habits will be more meaningful for developing an effective marketing strategy of products.

References

[1] H. Cao, T. Bao, Q. Yang, E. Chen, and J. Tian, An effective approach for mining mobile user habits, *In Proc. of the 19th ACM Conference on Information and Knowledge Management (CIKM'10)*, pp. 1677–1680, 2010.

[2] H. Ma, H. Cao, T. Bao, Q. Yang, E. Chen, and J. Tian**,** A habit mining approach for discovering similar mobile users, *The Proceedings of the 21st international conference on World Wide Web*, pp. 231-240. 2012.

[3] R. Agrawal and R. Srikant, Mining Sequential Patterns, *In the Proceedings of the Eleventh International Conference on Data Engineering*, pp. 3-14, 1995.

[4] Jay Ayres, Johannes Gehrke, TomiYiu, and Jason Flannick. SPAM: Sequential PAttern Mining using A Bitmap Representation, *In Proc. of the 8th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 429-435, 2002.