

Identifying Recent Behavioral Data Length in Mobile Phone Log

Iqbal H. Sarker
Dept. of Computer Science
and Software Engineering,
Swinburne University of
Technology, VIC, Australia.

Muhammad Ashad Kabir
School of Computing and
Mathematics, Charles Sturt
University, NSW, Australia.

Alan Colman, Jun Han
Dept. of Computer Science
and Software Engineering,
Swinburne University of
Technology, VIC, Australia.

ABSTRACT

Mobile phone log data (e.g., phone call log) is not static as it is progressively added to day-by-day according to individual's diverse behaviors with mobile phones. Since human behavior changes over time, the most *recent pattern* is more interesting and significant than older ones for predicting individual's behavior. The goal of this poster paper is to identify the recent behavioral *data length* dynamically from the entire phone log for recency-based behavior modeling. To the best of our knowledge, this is the first *dynamic* recent log-based study that takes into account individual's *recent behavioral patterns* for modeling their phone call behaviors.

CCS Concepts

•Information systems → Mobile information processing systems;

Keywords

Mobile data mining; behavior modeling; contexts; recency;

1. INTRODUCTION

Nowadays, mobile phones have become part of our life. Mobile phones can record various types of contextual data related to a user's phone call activities. In the real life, individual mobile phone users may behave differently (Accept | Reject | Missed | Outgoing) in different contexts such as time-of-the-day, days-of-the-week, social situations (e.g., meeting), locations, social relationships (e.g., mother) between individuals [3]. Modeling an individual's phone call behavior utilizing *recent log data* may assist them in their daily activities through capabilities, such as call interruption management, call firewall, call reminder systems.

Since human behavior changes over time, the most *recent pattern* is more interesting and significant than older ones for predicting individual's phone call behavior. To model users' present behavior, a number of research [5, 6] have used the

behavioral patterns of recent mobile phone log, e.g., last 3 months data, to predict the future behavior rather than the patterns derived from the entire historical logs. However, such arbitrary *static* data length is not meaningful to predict future phone call behavior as the time frame of the data length depends on when in the recent past behavior of a user has been changed significantly. Assume that as per log data the user has a call 'reject' behavioral pattern on Monday[10:00AM-12:00PM] as she used to have a regular meeting at that time. Currently, she has no meeting at that time period on Monday and she typically 'accepts' incoming phone calls. So, for this example, the past 'reject' behavioral pattern, even with high evidence according to log data, is not appropriate to predict her future behavior. Therefore, *identifying the optimal length* of log data that encompasses *recent behavioral patterns* is the key for modeling individual's recency-based behavior.

To identify such data length, if we consider only a short length (e.g., last week's data) as indicative of recent behavior, there may not be enough data instances in that time period to infer a valid rule for predicting future phone call behavior. Creating rules based on observations with so little "support" is unlikely to be effective [4]. On the other hand, if we take into account comparatively longer lengths (e.g., last 6 months data) as indicative of recent behavior, we could get greater *support* but it might result a greater *behavioral variations* thus decrease the confidence of some expected rules. As a consequence, we may miss these rules because of low confidence. Therefore, the *optimal length* of log data that reflects the recent behavior of an individual needs to be identified for recency-based behavioral modeling.

2. OUR APPROACH

This section presents our optimal log data length identification approach in step by step. Figure 1 shows an overview of our approach.

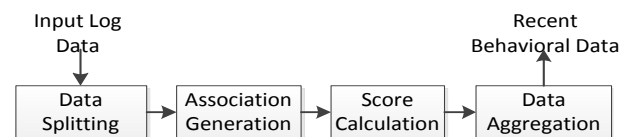


Figure 1: An overview of identifying recent behavioral log data

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MobiQuitous 2017 November 7–10, 2017, Melbourne, VIC, Australia

© 2017 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5368-7/17/11.

DOI: <https://doi.org/10.1145/3144457.3144506>

Step1 (Data Splitting): In this step, we split the entire log into week-wise data as the time-of-the-week is the most important aspect of user behavior in a mobile-Internet portal [2]. We choose weekly basis splitting because individuals' behaviors are not identical for all days in a week (Monday, Tuesday,..., Sunday), may differ from day-to-day that is captured in a week.

Step2 (Association Generation): In this step, for each week-wise data we generate context associations (combination of contexts in datasets) by adding the contexts incrementally according to the precedence of contexts. To identify the precedence of contexts in a dataset, we calculate information gain which is a statistical property that measures how well a given attribute separates training examples into targeted behavior classes [3]. The one with the highest information is considered as the highest precedence context.

Step3 (Score Calculation): Once we have generated the context associations, we then calculate the conflict score based on conflict behavior for each association between two adjacent weeks. For this, we first identify the *dominant behavior* (maximum number of occurrences) [4], as we do not expect always 100% like behavior of a user for a particular association. For instance, say a user 85% rejects, 10% accepts and 5% misses the incoming calls for a particular association of context (e.g., meeting, office), then 'reject' will be the dominant behavior for that association. We start scanning from the most recent week $[n]$ and continue to all previous weeks $[n-1, n-2, \dots, 1]$ one by one until getting significant behavioral variations for the generated associations between the weeks.

Step4 (Data Aggregation): In this final step, we aggregate the week-wise data based on similar behavioral patterns identified by conflict score. For identifying behavioral similarity, we calculate conflict score rather than likelihood as we do not expect similar contexts in each week. If the conflict score of two adjacent weeks is 0% (no conflict), the behavioral patterns are highly similar in these two weeks. If there is a significant change in the conflict score of two adjacent weeks, we stop scanning and set a boundary line for recent similar behavioral patterns. In this way, for some users, recent behavioral patterns are found by aggregating large number of weeks and for some users a smaller number of weeks depending on how the user's behavior changes over time-of-the-week in different contexts.

3. EXPERIMENTS

We have conducted experiments on real phone log datasets of two individual mobile phone users (randomly selected from Massachusetts Institute of Technology (MIT) Reality Mining dataset [1]). We extract 7-tuple information of the call record for each phone user from the datasets: date of call, time of call, call-type, call duration, location, relationship, call ID. These datasets contain three types of phone call behavior, e.g., incoming, missed and outgoing. As can be seen, the users' behaviors in accepting and rejecting calls are not directly distinguishable in incoming calls in the dataset. As such, we derive accept and reject calls by using the call duration. If the call duration is greater than 0 then the call has been accepted; if it is equal to 0 then the call has been rejected [4]. We also pre-process the time-series data in mobile phone log as it is continuous and numeric. For this, we use BOTS technique [4] for producing behavior-oriented time segments.

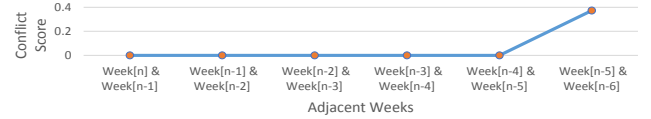


Figure 2: Conflict score for User 1

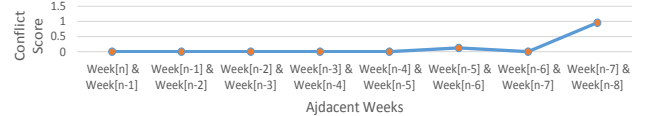


Figure 3: Conflict score for User 2

Figure 2 and Figure 3 show the conflict score between two adjacent weeks starting from most recent week (week n) for User 1 and User 2 respectively. From Figure 2 we found that the behavioral patterns are similar from week n to week $[n-5]$ and a significant variation has been occurred between week $[n-5]$ and week $[n-6]$ for User 1. In other words, the last 6 weeks data is the recent log data that represents the recent behavioral patterns of User 1. Similarly, Figure 3 shows that the last 8 weeks data is the recent log data that represents the recent behavioral patterns of User 2. As the behaviors of different individuals are not identical in the real world, such data length may differ from user-to-user according to their unique behavioral patterns.

4. CONCLUSION

In this paper, we have presented a novel approach to identify recent behavioral data length in mobile phone log dynamically by analyzing individual's behavioral patterns. Although we choose the phone call behavior as an example, our approach is also applicable to other application domains. We believe that our concept of dynamic recent log data helps both the researchers and application developers for predicting behavior of end mobile phone users according to their needs in various real-world applications, such as intelligent call interruption handling, mobile notification management systems, mobile application recommender systems etc.

5. REFERENCES

- [1] Eagle et al. Inferring social network structure using mobile phone data. *National Academy of Sciences*, 2006.
- [2] Halvey et al. Time based patterns in mobile-internet surfing. In *Human Factors in computing systems, Canada*, 2006.
- [3] Iqbal H Sarker et al. An approach to modeling call response behavior on mobile phones based on multi-dimensional contexts. In *Mobile Software Engineering and Systems, Argentina*, 2017.
- [4] Iqbal H Sarker et al. Individualized time-series segmentation for mining mobile phone user behavior. *The Computer Journal, Oxford University, UK*, 2017.
- [5] Lee et al. An adaptive speed-call list algorithm and its evaluation with esm. In *Human Factors in Computing Systems*. ACM, 2010.
- [6] Phithakkitnukoon et al. Behavior-based adaptive call predictor. *ACM TAAS*, 6(3):21, 2011.