2012 IEEE 12th International Conference on Data Mining Workshops

Mining Temporal Profiles of Mobile Applications for Usage Prediction

Zhung-Xun Liao*, Po-Ruey Lei†, Tsu-Jou Shen*, Shou-Chung Li* and Wen-Chih Peng*
*Department of Computer Science,
National Chiao Tung University, HsinChu, Taiwan
Email: {zxliao.cs96g, acrt3820, dreampilot.cs00g, wcpeng}@nctu.edu.tw
†Department of Electrical Engineering,
Chinese Naval Academy, Kaohsiung, Taiwan
Email: kdboy1225@gmail.com

***Abstract*—Due to the proliferation of mobile applications (abbreviated as Apps) on smart phones, users can install many Apps to facilitate their life. Usually, users browse their Apps by swiping touch screen on smart phones, and are likely to spend much time on browsing Apps. In this paper, we design an AppNow widget that is able to predict users' Apps usage. Therefore, users could simply execute Apps from the widget. The main theme of this paper is to construct the temporal profiles which identify the relation between Apps and their usage times. In light of the temporal profiles of Apps, the AppNow widget predicts a list of Apps which are most likely to be used at the current time. AppNow consists of three components, the usage logger, the temporal profile constructor and the Apps predictor. First, the usage logger records every App start time. Then, the temporal profiles are built by applying Discrete Fourier Transform and exploring usage periods and specific times. Finally, the system calculates the usage probability at current time for each App and shows a list of Apps with highest probability. In our experiments, we collected real usage traces to show that the accuracy of AppNow could reach 86% for identifying temporal profiles and 90% for predicting App usage.**

***Keywords*-mobile application; temporal profile; prediction; data mining;**

I. Introduction

Smart phones have became an important smart device in people's daily life. We use them to communicate with friends, check emails, take pictures, and play games. Concurrently, we can install many kinds of mobile applications (abbreviated as Apps) in our smart phone and invoke them for individual purposes. However, according to our observation, the average number of Apps of each device is about 70 to 80 and that of some devices which even exceed 150. We realize that as the number of Apps in a user's smart phone increases, users will spend an increasing amount of time looking for and launching the Apps they want to use.

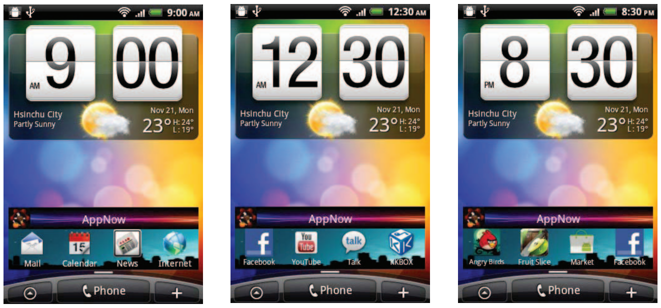
To deal with this problem, we have designed an AppNow widget which can dynamically predict users' App usage through mining temporal profiles from the users' previous usage behavior. For example, Figure 1 shows different prediction results at different times in one day. In Figure 1(a), the time is 9:00 a.m. and AppNow shows that the user is intending to start work by checking calender, emails, and so on. In Figure 1(b), the time goes to 12:30 p.m. and AppNow indicates that the user is about to communicate with friends using social network services. In Figure 1(c), the user is likely to play games at home when the time is 8:30 p.m. However, there are two challenges when designing the AppNow widget: 1) connecting the relation between Apps and their launched times, and 2) calculating the usage probability through comparing the App launched times and current time. First, Apps are not always launched at the same time. For example, if a user checks Facebook approximately once every one and a half hours, the usage time could be around 9:00, 10:30, 12:00, and so on. Therefore, to connect the relation between time and Apps usage, we proposed a temporal profile to summarize the usage history of each App. Second, since the launched times of an App may not exactly match the current time, we have to model the usage probability over the time different between the launched times and current time.

To the best of the authors' knowledge, although there are many research works focusing on smart phones [1], [2], [3], [4], [5], [6], [7], [8], there are no existing works that explore predicting usage behavior, let alone developing a widget on smart phones. On the other hand, current prediction algorithms on location, purchasing, and co-authoring [9], [10], [11], [12] do not create the relation with the aspect of time, such that they cannot be applied to solve the novel problem of predicting the App usages.

A. System Framework

The system flow of the AppNow widget is shown in Figure 2, where AppNow possesses three main components, a usage logger, a temporal profile constructor and an App usage predictor. The usage logger records the launched time and App ID on every App launch. The temporal profile constructor builds a temporal profile for each App. We summarize and investigate the usage history for each App into a set of (period, {specific time}) tuples which is, therefore, called a temporal profile for that App. The usage predictor calculates the probability of using each App at the current time. The AppNow widget then shows the 4 Apps with the highest probability.

978-0-7695-4925-5/12 $26.00 © 2012 IEEE
DOI 10.1109/ICDMW.2012.11



(a) Morning (b) Noon (c) Night

Figure 1. An example of the AppNow widget on a smart phone.

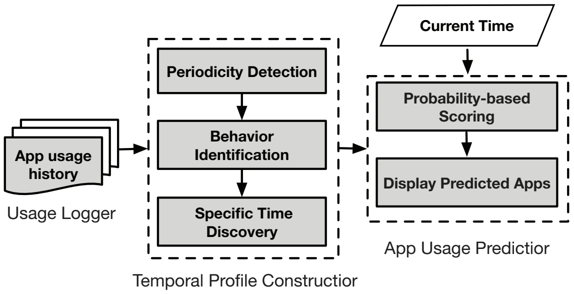
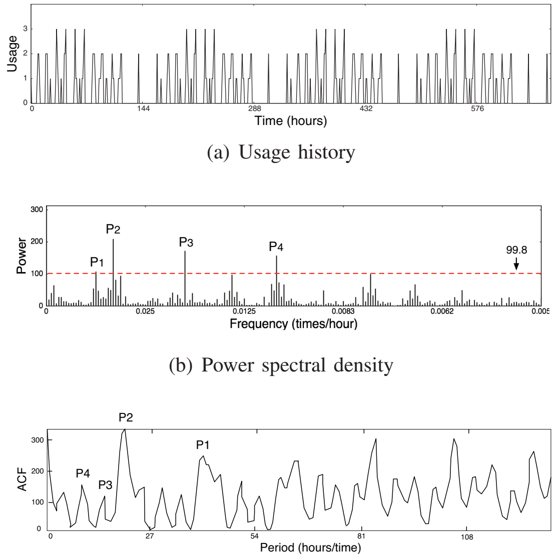


Figure 2. The system flow of AppNow.



(a) Usage history

(b) Power spectral density

(c) Discovered periods

Figure 3. An example of periodicity detection.

B. Demonstration of the System

The AppNow widget is developed on Android based smart phones. Users do not need to set any parameters. The system automatically logs the App usage behavior and updates the temporal profiles. When the widget becomes active (i.e. shown on the screen), the predictor updates the displayed four Apps regarding current time. Users can execute the Apps by directly touching the App icon in the AppNow widget. However, we neither move nor re-organize the placement of Apps in users' smart devices, so when the AppNow widget cannot provide the correct Apps, the user can still find them in their original place.

II. Design and Implementation

A. Mining Temporal Profiles

A temporal profile is a set of (p, T) tuples, where p is the period of usage and T is the set of usage specific times in the periodicity p. The discovery of temporal profiles consists of three steps: periodicity detection, behavior identification and specific time discovery.

First, we detect the periodicities of Apps by the idea proposed in [13]. For each App, we adopt Discrete Fourier Transform (DFT) to find the power spectral density (PSD). The App usage history is represented as $AH(app) =< a_1, a_2, ..., a_N >$, where a_i represents the number of launches at time unit i. For example, Figure 3(a) plots the usage chart of one App usage history for 4 weeks, where the length of one time unit is 1 hour, and therefore, N is 672. In addition, Figure 3(b) depicts the periodogram after applying DFT to Figure 3(a). In Figure 3(b), the dashed line is an automatically adjusted threshold which is obtained by using the dynamic cut approach [13]. The main concept of dynamic cut is that the power of a significant frequency should be higher than the maximum power derived from a random sequence. The random sequence is generated by shuffling the original $AH(app)$ and we claim that there should be no significant frequency in the random sequence. Finally, autocorrelation is adopted to verify the periodicities by a more accurate estimation. In Figure 3(c), we map the frequency to period, and we can see that the mapped period P_2 corresponds to 24 hours in this case.

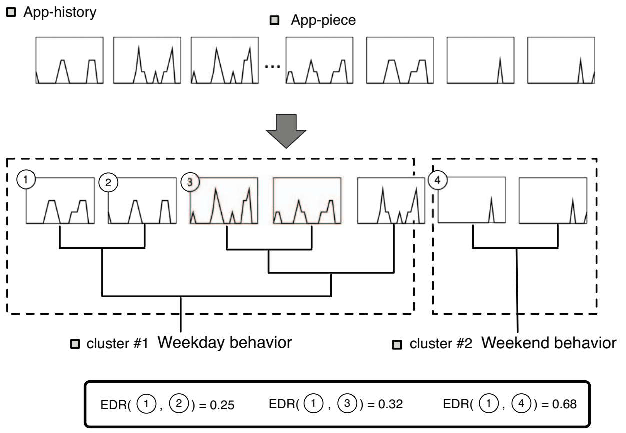


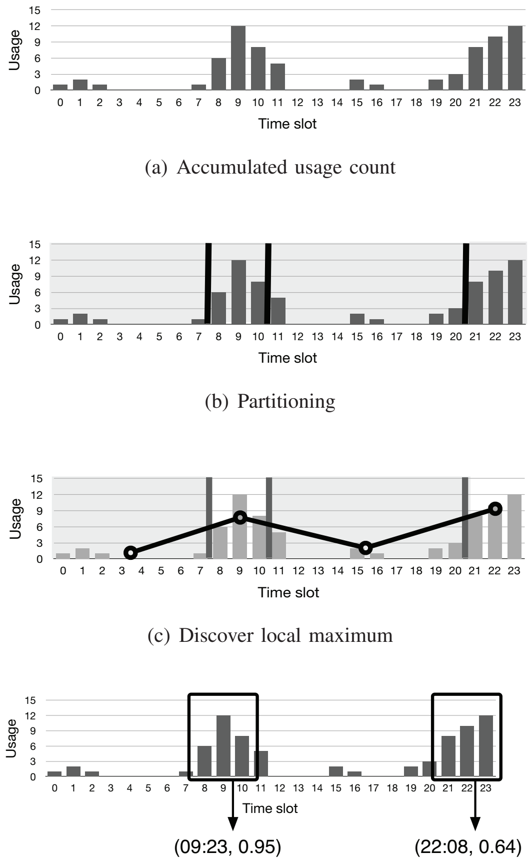
Figure 4. An example of behavior identification.

Second, after the periodicity detection step, for an App with its periodicity set $\{p_1, p_2, \ldots\}$, we further identify multiple behaviors for an individual periodicity, p_i. Since different behaviors may share the same periodicity, we separate them in this step. Here, we perform a hierarchical clustering to identify the user's multiple behaviors. Figure 4 shows an example of behavior identification, where we first decompose the App-history into several pieces according to the derived periodicity, and then we utilize EDR [14] to calculate the distance of two pieces. A hierarchical clustering would separate the pieces into different groups which are viewed as the user's multiple behaviors. In Figure 4, we observed that the two groups belong to weekday behavior and weekend behavior respectively.

Since the exact usage time would be slightly shifted, in the third step, the specific times of each group identified in the previous step is composed by the mean and variance of previous usages. In Figure 5(a), we take an example by plotting the accumulated usage in 24 hours which is the periodicity P_2 in periodicity detection. We first separate the temporal space into partitions such that the variance of usage in each part could be minimized. Intuitively, the partitions could be derived by a greedy algorithm. In Figure 5(b), the partitions are [0,8], [8,11], [11,21], and [21,24]. Then, we calculate the usage of each part and derive their local maximums to be the usage times. As shown in Figure 5(c), [8,11] and [21,24] are local maximums. Finally, the specific times are the means of usage in the local maximums as depicted in Figure 5(d). We use a tuple of $(mean, variance)$ to represent each specific time. Eventually, $(24, \{(09:23, 0.95), (22:08, 0.64)\})$ is added to the App's temporal profile.

B. App Usage Prediction

By deriving the App temporal profile which consists of the usage periods and the corresponding specific times, we can predict the possible App usage by calculating the usage probability of each App. We propose a probability-based scoring model, which is based on Chebyshev's inequality from probability theory [15], to formulate the usage probability of each App. Equation 1 shows the Chebyshey's inequality. It shows the probability of that the time difference between the current time and the specific time is not less than λ. Therefore, we can use Equation 2 to calculate the score, where CTime is current time, STime is the specific time, and $Var[STime]$ is the variance of $STime$.



(a) Accumulated usage count

(b) Partitioning

(c) Discover local maximum

(d) Calculate mean and variance

Figure 5. An example of specific times discovery.

$$P(|x - E[x]| \geq \lambda) \leq \frac{Var[x]}{\lambda^2} \tag{1}$$

Finally, the probability of launching an App is the maximum score from all specific times in its temporal profile, which is formulated as $MAX(Score(CurrentTime, SpecificTime_i))$, for each specific time in the temporal profile.

$$Score(CTime, STime) = \frac{Var[STime]}{|CTime - STime|^2} \tag{2}$$

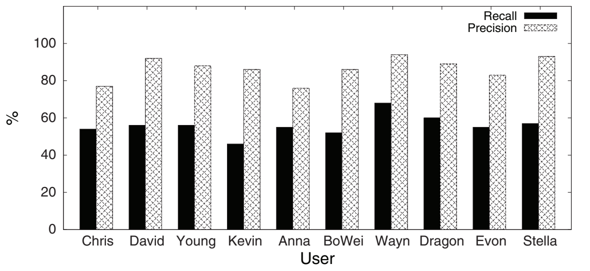
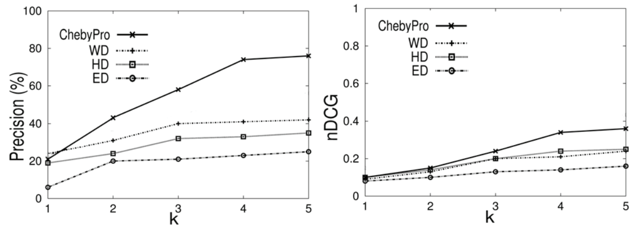


Figure 6. Precision and recall of specific times discovery.



(a) Precision (b) nDCG

Figure 7. The comparison of different prediction approaches.

III. Experimental evaluation

We installed a monitoring logger in 10 smart phones to collect the usage traces from July to December 2012. The dataset was separated into three parts: July to August, September to October, and November to December. For each part, the first month is the training date while the second month is the testing data. The overall performance is, therefore, the average performances of the three parts.

Based on the collected traces, we first evaluate the correctness of the discovered temporal profiles. We check if Apps are launched at the specific times indicated in their temporal profile. Figure 6 shows the precision and recall results for each user. As can be seen in Figure 6, the average precision is 86.4% and the average recall is 55.9% when the AppNow widget shows 4 Apps.

Then, we compare the accuracy of usage prediction among AppNow and three frequency-based baseline methods, WD, HD, and ED. Figures 7(a) and 7(b) depict the precision and nDCG score with different k (the number of Apps shown in AppNow). The baseline methods select the top-k most frequently used Apps in different time intervals. The interval of WD is a whole day (24 hours), HD is half a day (12 hours) and ED is 8 hours. As can be seen in Figure 7, our proposed AppNow outperforms the three frequency-based methods in terms of both precision and nDCG score.

IV. Conclusions and future work

We develop an AppNow widget to predict App usage from mining the App temporal profile. The temporal profile summarizes and investigates the usage history of an App. When the AppNow widget is activated, it calculates the usage probability for each App through a proposed probability-based scoring model. We collected real usage traces from 10 smart phones for 6 months. We evaluated the accuracy of both the temporal profiles and the App prediction. The results show that AppNow outperforms three frequency-based methods. In addition to predicting App usage, AppNow can also help recommend new Apps for users, since the predicted Apps can reflect the user's semantic activity. For example, when the predicted Apps are composed of the Apps related to games, we can infer that the semantic activity of the user is gaming and can recommend other games for that user.

References

[1] A. Thiagarajan, J. Biagioni, T. Gerlich, and J. Eriksson, "Cooperative transit tracking using smart-phones," in *Proc. of SenSys*, 2010, pp. 85–98.

[2] E. Kanjo, J. Bacon, D. Roberts, and P. Landshoff, "MobSens: Making Smart Phones Smarter," *IEEE Pervasive Computing*, vol. 8, no. 4, pp. 50–57, 2009.

[3] N. Pears, D. Jackson, and P. Olivier, "Smart Phone Interaction with Registered Displays," *IEEE Pervasive Computing*, vol. 8, no. 2, pp. 14–21, 2009.

[4] A. Toninelli, R. Montanari, O. Lassila, and D. Khushraj, "What's on Users' Minds? Toward a Usable Smart Phone Security Model," *IEEE Pervasive Computing*, vol. 8, no. 2, pp. 32–39, 2009.

[5] C.-W. You, C.-C. Wei, Y.-L. Chen, H.-H. Chu, and M.-S. Chen, "Using mobile phones to monitor shopping time at physical stores," *IEEE Pervasive Computing*, vol. 10, no. 2, pp. 37–43, 2011.

[6] B. Priyantha, D. Lymberopoulos, and J. Liu, "Littlerock: Enabling energy-efficient continuous sensing on mobile phones," *IEEE Pervasive Computing*, vol. 10, no. 2, pp. 12–15, 2011.

[7] D. Quercia, G. D. Lorenzo, F. Calabrese, and C. Ratti, "Mobile phones and outdoor advertising: Measurable advertising," *IEEE Pervasive Computing*, vol. 10, no. 2, pp. 28–36, 2011.

[8] C. Zhu and W. Sheng, "Motion- and location-based online human daily activity recognition," *Pervasive and Mobile Computing*, vol. 7, no. 2, pp. 256–269, 2011.

[9] S.-M. Huang, C.-F. Tsai, D. C. Yen, and Y.-L. Cheng, "A hybrid financial analysis model for business failure prediction," *Expert Syst. Appl.*, vol. 35, no. 3, pp. 1034–1040, 2008.

[10] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou, "A hybrid prediction model for moving objects," in *Proc. of ICDE*, 2008, pp. 70–79.

[11] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "Wherenext: a location predictor on trajectory pattern mining," in *Proc. of KDD*, 2009, pp. 637–646.

[12] D.-A. Chiang, Y.-H. Wang, and S.-P. Chen, "Analysis on repeat-buying patterns," *Knowl.-Based Syst.*, vol. 23, no. 8, pp. 757–768, 2010.

[13] M. Vlachos, P. Yu, and V. Castelli, "On periodicity detection and structural periodic similarity," in *Proc. of SDM*, 2005.

[14] L. Chen, M. T. zsu, and V. Oria, "Robust and fast similarity search for moving object trajectories," in *Proc. of SIGMOD*, 2005, pp. 491–502.

[15] B. Ostle and L. C. Malone, "Statistics in research: basic concepts and techniques for research workers," 1988.