An Analysis of the Transitions between Mobile Application Usages based on Markov Chains

Charles Gouin-Vallerand

LICEF Research Center, Télé-Université du Québec

5800 St-Denis Boul.

Montreal, QC H2S 3L5 Canada

charles.gouin-vallerand@teluq.ca

Neila Mezghani

LICEF Research Center, Télé-Université du Québec

5800 St-Denis Boul.

Montreal, QC H2S 3L5 Canada

neila.mezghani@teluq.ca

Abstract

By knowing more the usage of mobile application on smart phone, it is possible to predict which applications will be used next. This kind of information is particularly interesting from a recommendation system point of view. In this paper, we present an analysis of the transitions between application usages on the data collected by the University of Cambridge's Device Analyzer project. Among other results, we conclude that the transition probabilities between application usages are distinct information from the probabilities of usage based on periods of time, such as application launches in a same hour and in a same day.

Author Keywords

Mobile applications usage; Data mining; Markov chain; Recommender system

ACM Classification Keywords

H.2.8 Database Applications: Data Mining; H.2.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

In their activities of daily life, a large portion of humans is using intensively software application several times a day: emails, games, communications, news, web browsers, etc. Mobile phones are becoming the first type of electronic devices regarding the usage. For

instance, Britons in 2013 spent more time on their mobile phone each month (an average of 41 hours, 42 minutes) than navigating on the web with PC [1]. As mobile applications are more and more intensively used, analyzing their usage became a research field, where usage models can help to improve the mobile usage experience, for instance, by classifying applications on the mobile phone [2] or recommending applications based on the user's context [3].

Recently, the Device Analyzer project [4] from the University of Cambridge proposed an Android OS application that is logging internal information about Android phones (e.g., battery level, cell tower ID, received SMS, etc.). Published on the Google Play store, more than 17000 devices, all across the globe, installed the Device Analyzer app on their smart phone. In the context of our researches, we are working toward recommendation algorithms that propose applications [5] to users based on their current context. Numerous data provided by the Device Analyzer project bring interesting information on apps usage and usage context, which can be used to recommend apps to mobile phone users. More specifically, we are interested to model the transition between application usages and use these relations to recommend apps to the users. Among other data, this modeling work can help to find causal relations between mobile application usages.

Therefore, in this paper, we are presenting the results of an analysis of mobile applications usage based on sporadic usage transitions/chains provided by the Device Analyzer system. We analyzed the log of thirty devices with more than three months of data each and computed three Markovian models for each device

corresponding to: (i) occurrence/probability of transition between application utilizations; (ii) occurrence/probability of application usage in a same hour; (iii) occurrence/probability of application usage in a same day. These three models are compared, using correlation methods, to verify if the first model (transition between apps) is a valid approach to predict a next application usage. In other words, we verified if the last used application information is valid to predict an application usage or if the occurrence of usage in a same period of time would provide similar information.

In the next section, we present a review of the related work (Section Related Work), followed by a short description of the analyzed data from the Device Analyzer project (Section Data model). Then, we present the methodology we used to analyze these data and how we computed the Markov chains (Section Methodology) followed by a presentation of the results (Section Results). Finally, we conclude this paper by presenting our future work based on the result of this paper.

Related Work

A number of research projects worked on understanding the modalities of mobile application usages. For instance, Verkasalo et al. [6] concluded that there is a strong relation between locations and app usage. Shin et al. [2] propose a system that analyzes the context and history of a device for classifying the different installed apps depending on their probabilities of use. One of their conclusions is that the app transition data are one of the most important contextual information to predict which app will be used next. Similar in their approach, Ke et al. [7] propose a system that predicts app usage based

"312";"89393313";"app|recent|0";"uk.ac.cam.deviceanalyzer/.gui.SplashActivity";"1"

"313";"89393313";"app|recent|1";"com.android.vending/.AssetBrowserActivity";"1"

"314";"89393313";"app|recent|2";"com.sec.android.app.sbrowser/.SBrowserMainActivity";"1"

"315";"89393313";"app|recent|3";"com.android.email/.activity.MessageListXL";"1"

"316";"89393313";"app|recent|4";"com.sec.android.app.clockpackage/.ClockPackage";"1"

"317";"89393313";"app|recent|5";"com.android.contacts/.activities.PeopleActivity";"1"

"318";"89393314";"app|recent|6";"com.android.contacts/.activities.DialtactsActivity";"1"

"319";"89393314";"app|recent|7";"com.android.mms/.ui.ConversationComposer";"1"

Sample 1 Example of last started apps data from the Device Analyzer project

also on the mobile phone context. In their case, they focused on five contextual information: last used app, hour of the day, day of the week, location and the user profile. The two kinds of information about the time are correlated with the location of the phone.

On the other hand, about using contextual information and application usage for recommending mobile apps, the AppJoy [8] uses app usage patterns, found by data mining months of data, to recommend mobile apps. In [3], Boëhmer et al. propose a system that uses contextual information, such as the location, accelerometer data and application usage to recommend apps from a custom Android OS app store.

From these related works, the last app usage or usage transition information stands out from the different contextual information available from mobile phone. However, in these previous works this type of information has not been corroborated with other usage data such as the relation between last used apps and other apps used in a same period of time.

Data Model
The Device Analyzer project proposes an Android application which log on the University of Cambridge's server, several types of data on smart phone usage: phone location, battery energy level, connected and visible Wifi networks, etc. In the context of our research, we especially focused on using the data on the ten last started apps on the phone, which are logged each five minutes by the Device Analyzer app. Sample 1 shows an example of data about the last 10 used apps (with non-anonymized app name). The

app|recent data includes: the time when the data was logged since the last bootup of the device (second column in Sample 1), the chronologically order of when the app was launches (where 0 is the most recent) and finally the application name and related Android activity. Moreover, as no application can be launched during the 5 minutes gap between two snapshots (in that case, the app|recent would remain the same), we had to filter the app|recent data to detect new application launches; avoiding the analysis of same data several times. In conclusion, the app|recent data give us the last 10 launch applications (maximum), but it is impossible from these data to know exactly when in the last 5 minutes these applications have been launches.

In addition to the app|recent data, we had also access within the data to snapshots (every 5 min.) of the current applications running in the task manager. We decided to not use this information at this stage of our research, as it does not give us a proper overview of the transition between the application usages, because of the 5 minutes gap between each snapshot.

Methodology
To analyze the data, we developed an app usage analyzer in the Java language, which reads and parses the Device Analyzer logs. This data mining application analyzes the application usage by following the general steps of the data mining: the data preparation, the construction of models and the evaluation of the extracted model. The Figure 1 presents an overview of the data mining process behind the app|recent analysis.
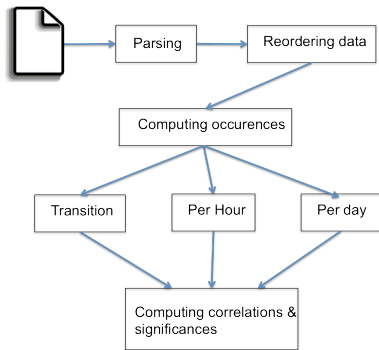
Figure 1 Overview of the Data mining process

Thus, the developed analyzer prepare the app|recent data to the data mining process by reordering the order of appearance of application from the most older to the newest for each group of app|recent data logged. To do so, we simply used a stack data structure and a forward-looking algorithm to put back the app|recent data in order of their launches.

Following the preparation of the data, the application usage analyzer computes the occurrences of transition between the application launches on the devices. At the same time, the system computes also the occurrence of application usage in same hour and same day periods. For instance, with these two last types of information, it is possible to know if there is applications that are more probable to be launch later in a same period of time than an application currently in use. These three kinds of occurrence were represented by three matrixes, which are, de facto, Markov chains [9], i.e. a set of states with a finite set of transitions between states and where the probabilities of transition are memory less (they doesn't depend on the previous states).

Besides using these three Markovian models to recommend application to user, one of our principal goals was to validate the transition probabilities between application uses as usage information that is distinct from the information of the most used application per hour and day. In other words, our goal was to validate if transition probabilities provide information that can predict future application usage, or the probabilities about which applications are most often used in a same period of time provide similar information.

Thus, in the last step of our data mining process, our analyzer correlates the information between the three Markov chains using the test of Fischer (F-Test) and measures their significance using the Z-Test. The F-Test verifies if the variance between the markov chains is similar. The results of these tests are presented in the next section.

Results

We ran our application usage analysis system, on the server of the University of Cambridge, on thirty devices randomly choose, where each device had more than three months of overall data. For privacy purpose, our analysis returned three Markov chains per device where app names and activity names have been replaced by hash strings. However, these hash strings were the same all across the Markov chains, allowing cross-models comparisons.

For the thirty selected devices, we recorded an average of 82 620 application usages per device with a standard deviation of 65 168 usages. Thus, there is a pretty high variation between the applications used between the devices/users.

Regarding the probability of transition between two apps, we collected the information about the highest probability of transition for each app and we compared this information with the occurrences of transition. A first conclusion coming from this analysis shows that more there is a usage occurrence for an application, more the probabilities of transition is distributed between the utilized applications. This information seems trivial, but it signifies that more an application is used; more it is difficult to predict its use. Figure 2 presents an example of application occurrences versus

the maximum probability for one of the thirty analyzed devices. This distribution is roughly the same for each device; the regression matching the best the results is the power regression with an average $R^2 = 0.59$. This distribution is totally different for the probabilities of usage based on the same day/hour information (Figure 3). This information about the distribution of maximum probabilities can help to classify applications in categories where usage transition can either or not been use to predict future usage.

Table 1 Correlation and significance between the three markovian models

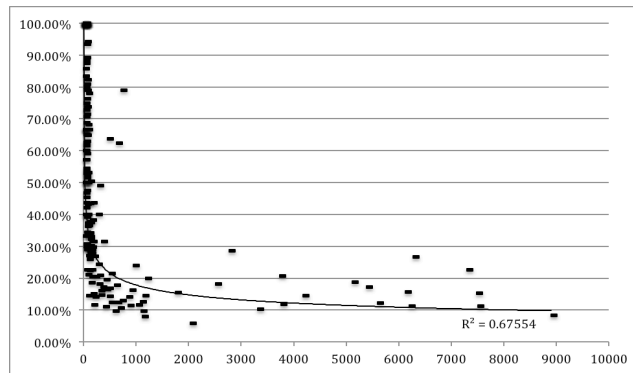|	Average F-Test	Average Z-Test
App transition vs. same hour	0.00117	0.0058
App transition vs. same day	0.00156	0.0054
Same day vs. same hour	0.9931	0.0488

Figure 2 Example of maximum probability of transition between apps versus number of analyzed occurrences for one of the analyzed devices

About the comparison between the application transition versus the probability of usage per hour and day (Table 1), we found that there is a low correlation between the Markov chains representing the transition between the application and the usage per hour and day, with a high significance of this result. This low F-Test score shows that, for the 30 devices we analyzed, the probabilities of transition between application

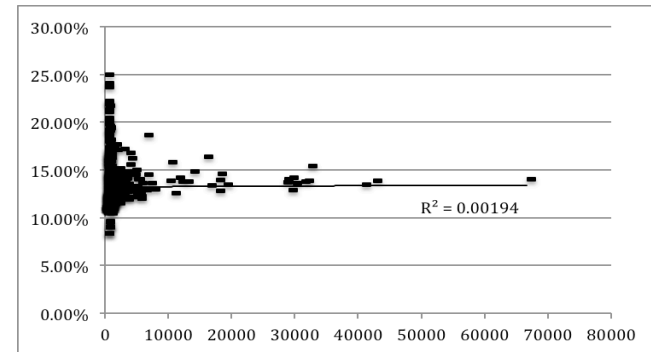usages is a distinct information from probabilities of usage for same periods of time.

Figure 3 Maximum probability of use in a same hour versus the number of analyzed occurrences for one of the analyzed device

Thus, both information as its specificity and provide data than can be compared/used, for instance, to dynamically recommend apps based on the user context. Moreover, our analysis doesn't demonstrate that one of the three Markov chains provide more valuable information. Depending of the type of application, the device's context and the profile of the users, both informations can provide interesting information such as in [7].

On the other hand, there is a strong correlation between the same day and same hour occurrence Markov chains, which mean that a macroscopic analysis of the usage for a periods of time lead to similar results than a microscopic analysis. However, this statement must be corroborated in future works, with other results from different period of time: days of the week, months, periods of the day, etc.

Conclusion

In conclusion, through our analysis on the application usage of the Device Analyzer project's data, we demonstrate that the probabilities of transition between application usages are distinct usage information compared to the probability of use for same periods of time. Our analysis shows also that the probabilities of transition become less obvious when the number of occurrences is high.

Today, recommendation systems are based mainly on social web data and user's service rating. By integrating contextual information and a more in depth description of the users, recommendation systems for mobile devices will be improved. In our next work, we are planning to utilize the presented methodology to recommend mobile apps to the users based on their interaction history. We are also planning to extend our data mining by discrediting the application usage based on location, current activities and other contextual information.

Acknowledgements

References

[1]   Nielsen Agency. How smartphones are changing consumers' daily routines around the globe. 2 Feb. 2014, (2014).

[2]   Shin, C., Hong. J-H and Dey, A.K. Understanding and prediction of mobile application usage for smart phones. Proceeding of UbiComp 2012, ACM Press (2012), 173-182

[3]   Böhmer M. et al. Contextualizing Mobile Applications for Context-aware Recommendation. Adjunct Proceedings of Pervasive 2010, (2012)

[4]   Wagner, D., Rice, A., and Beresford, A. Device Analyzer: Understanding smartphone usage, 10th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, (2013)

[5]   Gouin-Vallerand, C. And Montero De la Cruz, J. A. Analysis of a context-aware recommender system model for smart urban environment, Conference on Advances in Mobile Computing and Multimedia, MoMM2013, (2013)

[6]   Verkasalo, H., et al.: Contextual patterns in mobile service usage, Personal and Ubiquitous Computing 13(5), 331-342 (2009)

[7]   Ke Huang et al. Predicting mobile application usage using contextual information. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12), (2012).

[8]   Yan, B. and Chen, G. Appjoy: personalized mobile application discovery. In Proceedings of the 9th international conference on Mobile systems, applications, and services, MobiSys '11, (2011).

[9]   Norris, J.R. Markov Chains. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, (1998).