On mining of prediction for discovery Smart phones applications in Daily Life




S.V.Subramanyam
Dept of Information Technology, Guru Nanak Institution of Technical Campus JNTUH, Hydreabad.

Satish Manchana
Assistant Vice President, Credit Suisse Ag, Singapore.

Nune Sreenivas
School of Electrical & Computer Engineering Aait, Addis Ababa University, Addis Ababa, Ethiopia.

P.Vamshi Krishna
Dept of Computer Science & Engineering, Anuradha Engineering College Sant Gadge Baba Amravati University, Maharashtra.

Abstract:

Smartphone technology have enabled the prevalence of mobile applications Predicting Apps usage has become an important task due to the proliferation of Apps, and the complex of Apps. Prediction of two aspects of human behavior using Smartphone's as sensing devices. the increasing number of mobile Apps developed, they are now closely integrated into daily life. However, the previous research works utilized a considerable number of different sensors as training data to infer Apps usage. To save the energy consumption for the task of predicting Apps usages, only the temporal information is considered.

Android is a modern mobile platform that is designed to be truly open source. Android applications can use advanced level of hardware and software, as well as local and server data, exposed through the platform to bring innovation and value to consumers. Android platform must have security mechanism to ensure security of user data, information, application and network However, the limited re-sources of current Smartphone requires both researchers and companies paying more attention to the way of electively managing mobile applications.

Such an Apps usage prediction framework is a crucial prerequisite for fast App launching, intelligent user experience, and power management of Smartphone's. A challenging task is how to predict mobile user's application usage patterns for improving Smartphone performance. Propose a Temporal-based Apps Predictor (TAP) to dynamically predict the Apps which are most likely to be used. To evaluate the performance of TAP, use two real mobile Apps usage traces and assess the accuracy and efficiency. The experimental results show that the proposed TAP with the MinEntropy selection algorithm could have shorter response time of Apps prediction.

Keywords: Smart phones, MaxProb, Android security, smartphone security, mobile application,, app usage prediction

INTRODUCTION:

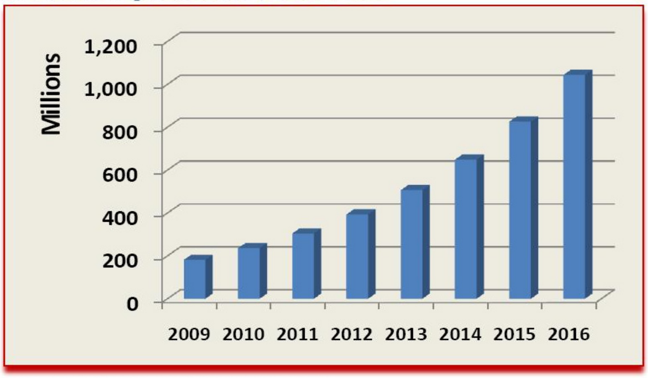
Mobile applications get rapid development due to the advances of the Smartphone technology. Now the Smartphone's usually support various applications and services beyond the traditional speech-centric service, such as music, videos, web browsing, gaming and camera shooting, just to name a few. Consequently, a large number of mobile applications are downloaded and installed, and more importantly, it is common for a Smartphone user to open the ability to foresee what mobile users want to do has many applications such as improving user interfaces or providing relevant recommendations.

Smartphone is the devices that enhance vision of universal computing: their small size, connectivity capabilities, storage capacity, mobility and their multi-purpose use are some of the reasons for their vast pervasiveness the malware has also appeared in the Smartphone platform. Apart from the increasing Smartphone sale, the annual downloads of Smartphone application are also on rise. Furthermore the use of Smartphone perimeter of an organization has increased besides increasing the sale of Smartphone and annual download of application from official as well as free source. Smartphone contains a vast amount of the user data, thus giving a serious privacy threat to the sector.

In additional Smartphone consist of popular web applications like e-mail, YouTube, social media, social networks facebook and twitter etc are being accessed through native applications instead of their useful web browser interface.

In this context Smartphone often manage a vast amount of user data, causing a serious threat to privacy of data. Mobile phones are convenient options for tracking and mining user behavior in daily life as they are usually placed in close proximity to the users . Smart-phones contain many sensors that can record contextual and activity cues, including location, application usage, and calling behavior.

Global Smartphone Sales, 2009-2016



Source: Telecom Trends International, Inc.

This information can be considered as both input and output of a prediction framework, in which the future values of some variables (e.g., next place) are predicted based on the current context (current place, time, etc.). Or run multiple applications simultaneously.

However, the limited capacity of both battery and memory becomes the bottleneck of the Smartphone performance, because several applications running simultaneously will consume a lot of resource, and some of these consumptions are even unnecessarily. Furthermore, too many running applications will also increase reaction time and affect the mobile user experience.

Thus, it is critical for effectively managing the mobile ap-plications so as to improve the Smartphone's performance. Meanwhile, it is impossible to let user do this kind of job by themselves, since this will require a lot of user interventions. Alternatively, if we profile users based on their usage patterns, we can leverage these profiles for developing automatical application managing techniques.

Therefore, in this paper we focus on automatically deter-mining and predicting each user's frequently-used applications in a fixed time slot (e.g. one hour), i.e., finding usage patterns. More importantly, to protect user privacy and save the data flow, it is desirable to directly perform usage patterns mining on users' Smartphone's instead of transiting the raw context data to the back end server and then per-forming mining.

In this way, the Smartphone will efficiently know which application the user may use and respond even before user requests. For example, when a user wants to send a message, his Smartphone guessed the motivation and showed the message application on screen, then user can write the message directly without finding and opening the application manually. After sending the message, the Smartphone estimates he will not use the application again for a long time, and it can shut down the application automatically, both saving the energy cost and freeing memory for upcoming applications.

From the mobile application usage logs, the usage time series can be generated expediently. Forecasting these usage time series is a possible way to judge whether or not an application shall be used . However, the problem of predicting mobile application usage patterns cannot be addressed by traditional time series models since these models are usually too complex and time consuming to be adopted by Smartphone for making real-time prediction. Thus, novel prediction methods that are both simple and effective are desperately required.

To this end, we explore the unique characters of the mobile application usage logs. Specifically, we note that there are some periodic changes in mobile user behavior. It is common that most people are working and resting by the rule of 24 hours per day and 7 days per week. Along this line, we propose a simple but efficient prediction method named Prediction Algorithm with Fixed Cycle Length (PAFCL), which considers both the periodic changes of behaviors and the influence of recent actions of mobile users. Experimental results on real-world data which is collected by Nokia show that our methods can predict usage patterns more precisely.

Constructing predictive models of human behavior has been a topic of interest in the area of recommendation systems , context-aware services , and personalized and adaptive interfaces . While these studies sup-port the claim that human behavior can be inferred through mobile phones, none of these studies, to our knowledge, exploit the information made avail-able through multiple sensors such as location, call logs, and proximity to others in order to build or enhance predictive models capable of determining aspects of user behavior in the future.

Hence, in this paper, we consider the task of predicting human behavior based on multiple Smartphone sensors using statistical methods. Our framework is inspired from prediction algorithms commonly used in signal processing for forecasting time series.

More precisely, we predict the next location of a user and which application he/she will use based on the current context consisting of location, time, app usage, Bluetooth proximity, and communication logs. This approach allows modeling the interplay between the predicted variables to study relationships between the place where a user stays and the possibility that he would make a phone call, use the cameras and so on.

Furthermore, other sources of information, such as the list of nearby Bluetooth devices or system information, are also exploited in order to enrich the user context and improve the predictive models. In summary, our paper makes the following four contributions. To our knowledge, this is the rst work attempt to study the application usage prediction task jointly with location. Second, we study the impact of each data type to the predictive performance of the two tasks.

Third, we investigate the use of generic behavior patterns for improving prediction performance of personalized models. And nally, we conduct our analysis on a longitudinal dataset involving 71 volunteer users carrying a Smartphone over 17 months, from the Lausanne Data Collection Campaign. This dataset allows us to extensively study the predictive performance conditioned on various aspects such as the evolution of predictive performance over a long period of time.

II.RELATED WORKS:

To the best of our knowledge, the prediction problem of Apps usage in this paper is quite different from the conventional works. We focus on not only analyzing usage history to model users' behavior, but on personalizing varied types of features including hardware and software sensors attached to Smartphone's. The proposed algorithm selects different features for different users to satisfy their usage behavior.

Although there have been many research works solving the prediction problem in different domains, such as music items or playlist prediction , dynamic preference prediction , location prediction, social links prediction , and so on, the prediction methods are only based on analyzing the usage history. In, the author selected features from multiple data streams, but the goal is to solve the communication problem in a distributed system.

Analyzing human behavior with mobile phones has received considerable interest in the recent past. For example, the Reality Mining dataset has been extensively used for this purpose.

This dataset was collected over a course of 9 months from 94 students and stat MIT using Nokia 6600 phones for recording call logs, Bluetooth devices in proximity, cell tower IDs, application usage, and phone status. Location that is inferred from cell tower data can be recorded by mobile phone operators for millions of people, oering the possibility of large-scale analysis of human mobility such as predictability of human mobility or identifying human daily activity patterns . Note that the resolution of location data inferred by cell tower IDs is relatively low compared to GPS data, which is used in this paper, resulting in uncertainty on the user position. Besides cell tower and GPS, short distance wireless network data (e.g., Bluetooth and WiFi) can also be used for positioning, with some advantages such as high spatial resolution and the ability to work in indoor environments.

A comparative evaluation of Smartphone platform is performed by using a set of evaluation criteria that are elaborated. The proposed criteria concern the development platform and the developer. The platform based criteria are objective, relying solely on the platform characteristics. Contrarily the developer related criteria are subjective, giving details about the development effort and as a result depends on the developer's skills and backgroundCurrently, only a few studies discuss mobile Apps usage prediction.

Although the authors in adopted location and time information to improve the accuracy of Apps usage prediction, the total number of Apps is only 15. Concurrently, in the authors stated that the prediction accuracy could achieve 98.9%, but they still only focus on predicting 9 Apps from a set of 15. In , Therefore, for different users, they always use the same three features to predict their Apps usage. In the authors investigated all possible sensors attached to a Smartphone and adopted a Naive Bayes classification to predict the Apps usage. However, collecting all possible sensors is inefficient and impractical. Moreover, the usefu l sensors for different users could vary according to users' usage behavior. We claim that for different users, we need to use different sets of features to predict their usage. In this paper, we collect only the subset of all features which are personalized for different users.

This paper is the first research work which discusses how to select suitable sensors and features for different users to predict their Apps usage. Through the personalized feature selection, we could perform more accurate predictions for varied types of usage behavior, , reduce the dimensionality of the feature space, and further save energy and storage consumption.

In addition, the proposed KAP framework derives the implicit feature by modeling the usage transition among Apps.

2. PROBLEM STATEMENT:

To ease presenting the problem of predicting mobile application usage patterns, we first define several notations.

Definition 1. The mobile application usage pattern is the mobile application usage habits of a specific user in special time slot.

By analyzing the past mobile application usage pattern-s hiding in the usage logs, we can predict the future usage patterns with some methods. Since we focus on the methods that are suitable for running on each smartphone, the follow-ing discussion just depends on the usage logs of the specific user.

As have said, the application usage logs mainly contain when and which the application is used by the specific mo-bile user. Segmenting the logs into the equal time slot, the time series of application usage numbers can be generated.

Definition 2. The time series NA = {NA1; : : : ; NAt} presents the usage number NAi of mobile application A by speci c user in each time slot Ti, i = 1; : : : ; t.

Given specific user and the length of the time slot, we can generate time series for each mobile application that the user has used. In other words, each time series stands for the way of the user using a mobile application in a specific time slot. It is obvious that, given a time series NA from T1 to Tt, the next value NA;t+1 can be predicted. Correspondingly, the prediction result can be considered as the possible usage number for each application in the next period.

Similarly, by predicting the value NAj ;t+1 for each mobile application Aj , the usage pattern of the specific user in time slot Tt+1 can be generated. In this way, we can achieve the target that predicting mobile application usage pattern of each mobile user. Based on the prediction result, it is easy to determine that the application with higher predicted number should be used in the next time slot.

Using the prediction results, the Smartphone can effec-tively manage the mobile applications to improve its performance and enhance user experience, and meanwhile avoid the wastage of both energy and memory.

III.EXPLICIT AND IMPLICIT FEATURES:

separate the features into two main categories: the explicit feature and the implicit feature. The explicit feature represents the sensor readings which are explicitly readable and observable. The implicit feature is the Apps usage relations.

A. Explicit Feature Collection:

The hardware sensors we use for the explicit feature. As different models of smartphones could have different sets of hardware sensors, we only list the most common ones whose readings are easy to record. It is totally free to add or remove any hardware sensors here.

To show the prediction ability of different types of mobile sensors, we randomly select two users from the collected dataset and perform kNN classification via the three types

Analyzing human behavior with mobile phones has received considerable interest in the recent past. For example, the Reality Mining dataset has been extensively used for this purpose. This dataset was collected over a course of 9 months from 94 students and sta at MIT using Nokia 6600 phones for recording call logs, Bluetooth devices in proximity, cell tower IDs, application usage, and phone status.

Location that is inferred from cell tower data can be recorded by mobile phone operators for millions of people, o ering the possibility of large-scale analysis of human mobility such as predictability of human mobility or identifying human daily activity patterns . Note that the resolution of location data inferred by cell tower IDs is relatively low compared to GPS data, which is used in this paper, resulting in uncertainty on the user position.

In the context of Human Computer Interaction, utilizing information from a user's behav-ior to modify mobile interfaces has been investigated in the past. Vetek et al. for example, proposed a method to dynamically improve the recommendation of shortcuts on the home-screen of a smartphone using user context. These authors used an unsupervised clustering based method that uses location data, in the form of GSM Cell IDs to learn user contexts to improve recommendations of these shortcuts.

Human behavior is to some degree predictable by a model that captures some emergent patterns from the behavioral data. For example, Eagle et al.

Characterized the emergence of mobility behaviors using factor analysis. Similar analysis was also done using topic models, which were in the past used in text mining and natural language processing. While not directly considering the predic-tion task, some work focused on the understanding of how people use their phone , which might be used as prior knowledge for prediction meth-ods. In a preliminary work, we presented an analysis on how people use their phone conditioned on various contexts, where the context is represented by semantic location (i.e., place categories) and Bluetooth density.

The main challenge is how to use e ciently the data from a large number of sensors in a uni ed framework. As simple method is use discrete representation of context then build a model for each state separately. This is an e ective method if the number of data types is limited, such as for the combination of two data types: time and location. More sophisticated methods require redesigning the predictive model to integrate multiple types of observation, such as the spatio-temporal decision tree or the probabilistic latent variable models which combine spatial, temporal, and social relation information for predicting human movements.

Compared to our recent work considering long range predictions of human mobility such as how long a user will stay in the current place, this paper focuses on very short future predictions. We also develop a more complete representation of context which allows us to learn useful generic models for predicting several dimensions of human mobile behavior.

Location data preprocessing. From the raw location coordinates extracted by the sensing software through GPS and known WiFi-access points, se-mantic labels were assigned to the most common places using a two-step procedure described in .

3.2. Data representation:

We use data streams to encode the human behavior over time and use statistical methods to learn a predictive human behavior model. We rst group together location labels, applications usage, and other contextual information into 10 minute time slots (the length of the time slots is a hyper-parameter and can be changed depending on the application). This enables us to view Smartphone data in the form of a set of temporally arranged data streams ( ), with each stream corresponding to one modality such as GPS or BT. this is conceptualized in Figure 3. Several of these streams can now be grouped into categories. In our work, of the many possibilities, we concern ourselves with six categories that are explained below.

4. Mobility prediction: results and discussion:

IV.PERSONALIZED FEATURE SELECTION:

The goal of the personalized feature selection is to use as fewer features as possible to guarantee an acceptable accuracy. Due to the energy and storage consumption of collecting sensors readings and Apps transition relations, we should select useful features for different users in advance.

4. EXPERIMENTS:

Our analysis starts with a basic analysis of generic behavior patterns and initial results of the prediction task. Then we present experimental results for personalized prediction, and investigate the combination of generic and personalized prediction.

4.1. Generic human mobility patterns:

Correlation Analysis. We use correlation analysis to characterize the dependency between contextual variables and the next location. A strong correlation indicates a high predictability degree of the output using a lin-ear model on the considered contextual variable. Table 1 reports the top 8 variables for each location data stream. The strong negative correlation between a destination and the contextual data stream of the same place cat-egory (e.g. Home vs. loc(Home)) re ects the fact that we consider the sequence of location changes in this analysis. That is, if the current location is Home, then the next location cannot be Home by construction. Note that the correlation is not perfect (j j < 1) since if the current location is not Home (loct(Home) = 0) then the next location can be Home or another label (i.e., the value of loct+1(Home) is not determinable). Among the 6 categories of data streams, we observe that loc and time dominate the set of top variables. For example, Transition is positively cor-related with the three dominant locations Home, Other and Work. Work is highly correlated with morning time slots and negatively correlated with weekends.

Restaurant is positively correlated with the time slots around noon. (We note that the correlation between Home and the time slot 3:30AM-4:00AM comes a daily reset of the phone client software which may generate unknown location states.) Interestingly, the inferred state data streams state provided by the recording software also appear frequently in the table and become important for predicting infrequent place categories such as Transport-Station, Shopping, Entertainment, or Other. Finally BT and Call-log data streams appear to have a weaker correlation to the next destinations but all correlation values are statically significant.

Generic human mobility prediction model. The above analysis suggests that generic mobility patterns exist, so that we can learn generic models to predict human mobility. Our preliminary experiments show that Random Forest is slightly better than Linear Regression models, but the Linear Regression model is much faster. To this end, we use Linear Regression to train our generic mobility prediction model.

The baseline performance corresponds to the static model that always predicts the most frequent destination as output, which is Transition in our data. Looking at the accuracies obtained with single data type, we found similar endings which are highlighted in the correlation analysis

4.2. Personalized mobility prediction:

Accuracy while generic human mobility can be captured, people have their \own" routines in real life. Therefore, personalized predictive models are expected to be more accurate than generic model. We believe that complex patterns exist in both generic and personalized data, but generic complex patterns cannot be extracted e ciently from a data collected by a rather low number of users (N=71). A more advanced generic human mobility model calls for data from a large number of users and the ability to exploit group people with similar patterns.

By combining single personalized predictors from the family of Markov models, Linear Regression, and the Random Forest, we obtain the best personalized model (Blend Personalized) which reaches an accuracy of 0.641 using only personal behavior data. Finally, combining the generic model with the personalized model increases the overall accuracy to 0.657, which shows the usefulness of generic model on the prediction problem. We will revisit this aspect later when discussing the analysis of prediction accuracies over time.

Which are important data types for personalized location pre-diction? Again, we study the impact of each data types on the prediction accuracy. Location remains the most important contextual variable for predicting mobility, yielding a prediction accuracy of 0.481. Surprisingly, BT becomes key contextual information for predicting next location. Furthermore, adding BT data streams to the model significantly improves the accuracy from 0.588 to 0.635. As BT data streams are proxies of face-to-face interactions, these results suggest that information about whom the user is with has a significant impact on where will he go.

Gathering Android smartphone data made easy

V.App usage prediction: results and discussion:

5.1. Generic app usage patterns:

Correlation between contextual variables and apps that will be opened. Table 5 reports the top 8 contextual variables for each of the 9 common apps that are used by at least 10 users. One common observation is that the use of an app at a given time is highly correlated with the use of that app in the next 10 minutes. For example, in the case of Camera (= 0:38), if the user is using camera then there is a high probability that the user continues to take more photos/videos in the next ten minutes.

Similar to the case of mobility correlation analysis, most of correlation values in Table 5 are small but statically significant (p-value<0.01). We can end many meaningful results such as Text message is likely to be opened in the next 10 minutes (to send or to receive messages) if the user recently sent/received SMS. Another interesting example is the case of Pro les app which is used for changing ringing types of the phone.

Generic app usage prediction model. We use Linear Regression to train a generic app usage prediction model which outputs activation scores

5.2. Personalized app usage prediction:

Which data types are important for personalized app usage pre-diction? Table 7 reports evaluation results of the best single model (Random Forest) using different feature sets. Compared to results of generic model, we and that the importance order of data types has changed significantly. While App remains the most important contextual information, Location and Bluetooth are now found to be useful for predicting app usage. State and Call log remain relatively useless for the prediction.

The detail results of app usage prediction with different models are shown in Table 8, which show that context-aware models outperform significantly the baseline predictive model using only app usage frequency. Furthermore, we observe again that the prediction performance of personalized model can be improved by exploiting generic model. In particular, the improvement is significant in the rest few weeks of data collection

The effect of the number of apps. Finally, we study the effect of number of apps on the performance and on the improvement of context-aware model over the baseline (Most Frequent). To track the tendency of performance with respect to the number of apps, we used a second order polynomial to t the data. The drop in recall values as the number of applications used increases is not linear, as seen in these graphs. Furthermore, we found that the improvement in performance of our models.

VI.Conclusion:

Propose a simple but efficient frame-work named Prediction Algorithm with Fixed Cycle Length (PAFCL), which can be used on the Smartphone's for real-time predicting each user's mobile application usage pattern-s.

Experimental results on the real-world data which is collected by Nokia show the effectiveness of PAFCL, and the proposed methods under this framework usually predict usage patterns more precisely.

Smartphone devices are multi-purpose portable devices enclosing a vast amount of third party applications that augment the device's functionality. It also showed the easiness of malware application development by average programmers that have access to the official tools and programming libraries provided by Smartphone platforms.

A silver bullet solution against similar attack scenarios is not available. Some of the possible steps that can reduce the possibility of being attacked is "prevention is better than cure". However, the following steps can be a possible way-out to reduce/avoid a potential malware outbreak in smartphones:

a) User awareness, i.e. informing user about security and
privacy risks in Smartphone platforms; and

b) Providing secure application distribution in Smartphone platform

VII.Acknowledgement

We thank the partial support of Nokia Research (LS-CONTEXT project) and SNSF (SONVB project). Mobile Phones:

A Tool for Social & Behavioural Change," is a collaborative work of UNICEF India and Digital Empowerment Foundation. At Digital Empowerment Foundation, we thank the significant role played by Mr. Osama Manzar, Founder Director in successfully bringing out the white paper and the conduction of the consultation.

Satish Manchana is Working as Assistant Vice President at Credit Suisse AG, USA in Information Technology department. Pursued Bachelor of Technology in Computer Science from Jawaharlal Nehru Technological University,Hyderabad. He is Microsoft Certified Professional,Vmware Certified Professional and ITIL Foundation certified.

He works on providing architecture,design and development for Microsoft platform solutions. His areas of interest include Cloud Computing,Virtualization and Windows Server Automation.

REFERENCES:

[1]N. Eagle, A. (Sandy) Pentland, Reality mining: sensing complex social systems, Personal and Ubiquitous Computing 10 (4) (2006) 255{268.

[2]S. N. Patel, J. A. Kientz, G. R. Hayes, S. Bhat, G. D. Abowd, Farther than you may think: An empirical investigation of the proximity of users to their mobile phones, Ubiquitous Computing 4206 (2006) 123{140.

[3]V. Bellotti, B. Begole, E. H. Chi, N. Ducheneaut, J. Fang, E. Isaacs, T. King, M. W. Newman, K. Partridge, B. Price, P. Rasmussen, M. Roberts, D. J. Schiano, A. Walendowski, Activity-based serendip-itous recommendations with the Magitti mobile leisure guide, in: Proc. CHI, 2008, pp. 1157{1166.

[4]B. D. Ziebart, A. L. Maas, A. K. Dey, J. A. Bagnell, Navigate like a cabbie: probabilistic reasoning from observed context-aware behavior, in: Proc. UbiComp, 2008, pp. 322{331.

[5]R. Bridle, E. McCreath, Inducing shortcuts on a mobile phone interface, in: Proc. IUI, 2006, pp. 327{329.

[6]N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez, J. Laurila, To-wards rich mobile phone datasets: Lausanne data collection campaign, in: Proc. ICPS, 2010.

[7]A. Oulasvirta, R. Petit, M. Raento, S. Tiitta, Interpreting and acting on mobile awareness cues, Human{Computer Interaction 22 (1-2) (2007).

[8]H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, D.Estrin, Diversity in smartphone usage, in: Proc. MobiSys, 2010.

[9]A. Vetek, J. Flanagan, A. Colley, T. Ker•anen, Smartactions: Context-aware mobile phone shortcuts, Human-Computer Interaction (2009).

[10]A. Clauset, N. Eagle, Persistence and periodicity in a dynamic proximity network, in: Proc. DIMACS, 2007.

[11]M. C. Gonzalez, C. A. Hidalgo, A.-L. Barabasi, Understanding individ-ual human mobility patterns, Nature 453 (7196) (2008) 779{782.

[12]T. M. T. Do, D. Gatica-Perez, By their apps you shall understand them: mining large-scale patterns of mobile phone usage, in: Proc. MUM, 2010.