Poster: A Mobility Prediction System Leveraging Realtime Location Data Streams

Vaibhav Kulkarni*
vaibhav.kulkarni@unil.ch

Arielle Moro*
arielle.moro@unil.ch

Benoît Garbinato
benoit.garbinato@unil.ch

Distributed Object Programming Laboratory
University of Lausanne
CH-1015 Lausanne, Switzerland

ABSTRACT

Location-based services today, exceedingly depend on user mobility prediction, in order to push context aware services ahead of time. Existing location forecasting techniques are driven by large volumes of data to train the prediction models in a centralised server. This amounts to considerably long waiting times before the model kicks in. Disclosing highly sensitive location information to third party entities also exposes the user to several privacy risks. To address these issues, we put forth a mobility prediction system, able to provide swift realtime predictions, evading the strenuous training procedure. We enable this by constantly adapting the model to substantive user mobility behaviours that facilitate accurate predictions even on marginal time bounded movements. In comparison to existing frameworks, we utilise less volumes of data to produce satisfactory prediction accuracies. This in turn lowers the computational complexity making implementation on mobile devices feasible and a step towards privacy preservation. Here, only the predicted location can be sent to such services to maintain the utility/privacy tradeoff. Our preliminary evaluations based on real world mobility traces corroborate our hypothesis.

Categories and Subject Descriptors

H.4.2 [**INFORMATION SYSTEMS APPLICATIONS**]: Spatial-temporal systems

Keywords

Realtime Mobility Prediction; Mobility Behaviour; Location based Services

1. INTRODUCTION

The rapid proliferation in the number of applications offering Location-based Services (LBS), such as Google Now and Google Maps equipped with user location forecasting, makes it evident that mobility prediction is becoming a key paradigm of such services. However, numerous data breaches and malicious third party entities has casted a shadow over LBS. As can be seen, it is quite evident that their success depends on how well the user privacy is taken into account.

Existing mobility prediction techniques, utilise about 70% of the data, exclusively for model training. This results in substantial waiting times until the model is able to produce usable predictions in real deployment scenarios. We find that, a major downside with learning on a large dataset is the shadowing effect on marginal user movements, which appear insignificant as a whole. This has a direct impact on the granularity of predictions. The present works, which link user behaviours with forecasting models, are only able to produce statistical motion patterns that do not truly grasp the inherent nature of human movements. On the other hand, when this data is sent to third party servers, a malicious entity can easily infer sensitive user information such as significant places using simple heuristics. Furthermore, the algorithmic cost of making predictions on a mobile device is relatively high due to the complex ensemble techniques used, making it necessary to have a powerful server.

*Co-Primary Student Authors

MobiCom'16 October 03-07, 2016, New York City, NY, USA

DOI: http://dx.doi.org/10.1145/2973750.2985263

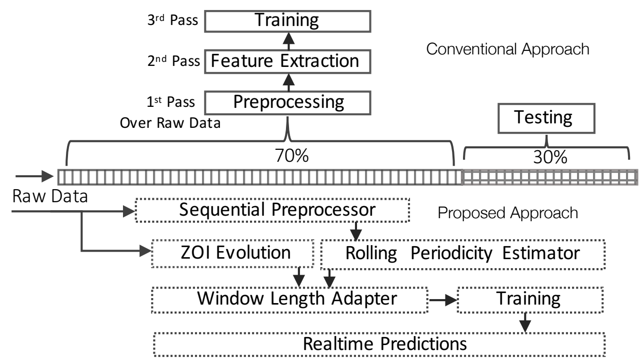

Figure 1: Traditional Prediction Systems *vs.* Our System. The process on the top, depicts the traditional mobility prediction approach. The process chain shown at the bottom gives an overview of our technique.

2. PROBLEM STATEMENT AND CONTRIBUTIONS

The main goal of our approach is to reduce the amount of data required to produce predictions with a satisfactory accuracy to small time windows. We analyse the realtime user mobility behavioural changes to adapt this window length accordingly. We quantify user behaviours in terms of the

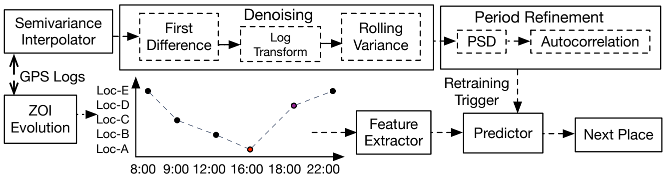

Figure 2: Movement periodicity estimation coupled with realtime prediction.

evolution of frequently visited places with time and the periodicities of movements amongst those places. The major contributions are listed hereafter.

- We propose a mobility prediction system on realtime sequential location logs, constantly adapting to substantial user mobility behaviours.
- The lower computational complexity, due to lesser data involved, makes implementation on hand held devices feasible. This avoids the usual long waiting periods and obtains quicker predictions.
- Our reactive frequently visited place computation scheme, models the behaviours restricted to small time bounds, achieving predictions with fine granularity as compared to conventional approaches.

3. RELATED WORK

Predicting future movements based on the mobility behaviours has been studied widely in the literature. Such a scheme to predict users next move based on the current position and involving a large part of the dataset for training is presented in [2, 6]. [4] first extracts all the frequently visited places by a user and uses the transitions within them to create a pattern tree. [1] applies clustering technique to extract such places with mean travel time to formulate the mobility models. Such approaches operate on the assumption that the frequently visited places will be static throughout, which does not necessarily hold in practice. According to our observation, such places evolve over time and have to be monitored continually to adapt to user behaviours. [5] computes the periodicity, by analysing the users visit frequency within certain places which is used to predict the next visit. However, the analysis is performed on the entire dataset, as opposed to our work in real time processing and retrieving non stationary periodic patterns lasting only for short time intervals.

4. SCOPE AND DESIGN

In this section, we discuss the quantification of mobility behaviours with respect to the evolving frequently visited places and the periodicity of movement. We describe the different families of prediction techniques used to evaluate the prediction accuracy. Futher, we present the system as a whole and illustrate, how the mobility behaviours are coupled with the predictors to produce next place predictions in realtime.

4.1 Mobility Behaviours

Frequently visited place evolution. We present the sudo code (see Algorithm 1) to obtain the user's frequently visited places in realtime. Here, a cluster represents a unique visit in a delimited area and a cluster group represents a zone consisting of several clusters intersecting each other. Two intersecting clusters are merged. Finally, we define a

Algorithm 1 ZOI discovery algorithm

Require: cluster, clusters, group, groups, zois
function ANALYZE(loc) ▷ Called when a new loc is detected
  if distance(loc, cluster.centroid) $\leq d_{max}$ **then**
    cluster.add(loc)
  else
    if cluster.stayingTime() $\geq t_{min}$ **then**
      tryToMerge(groups)
      if !tryToPut(cluster, groups) **then**
        create new group
        group.put(cluster)
        groups.add(group)
      clusters.add(cluster)
      if currentVisitNB < visitThreshold **then**
        currentVisitNB = visitNB(clusters)
      updateZOIs()
    create new cluster
    cluster.add(loc)
function UPDATEZOIS
  zois = $\emptyset$
  for group $\in$ groups **do**
    if group.clusterNB $\geq$ currentVisitNB AND time(now, group.lastVisit) $\leq t_{v_{max}}$ **then**
      zois.add(group)

Zone of Interest (ZOI) as a frequently and recently visited group of these clusters. Tracking these bounds over time enables to discover their evolution which in turn captures the user behavioural movement patterns.

Movement periodicity. A major challenge here is to identify periods, which do not repeat precisely at same times in addition to having multiple interlaced patterns in the non-stationary time series. Thus, standard period estimation techniques, such as autocorrelation and Fourier transform, cannot be directly applied. Further, the realtime location logs, cannot be assumed to arrive at a uniform rate. We use semivariance interpolation which conceals the incoming data stream about spatial variance at a specified distance to get a uniformly sampled stream. We further process the stream by taking the first difference and applying the log transform to get a series with constant variance. To calculate the periodicity, we calculate the power spectral density to get the candidate periods and feed them to autocorrelation estimator to rectify any false alarms resulting due to spectral leakage as shown in Figure 2

We evaluate the system performance with Mobility Markov chain (MMC) models, simple classification, artificial neural networks, recurrent neural networks and Fourier extrapolation.

4.2 Overall System

We formulate the problem as a non-stationary time series prediction, where the model needs to be retrained according to variations in the incoming data stream. In our case, these correspond to user movements and the variations link to changing periodicities and number of ZOIs as depicted in Figure 1. We empirically determine that the model accuracies are affected for an autocorrelation index change of 0.2 and greater, which serves as a trigger for retraining. The feature vector consists of, ZOI evolution (representing the movements), starting and stay times which are fed to the predictors. In case of MMC models, their inherent structure leads to a strong dependency on the ZOI evolution pattern. The size of the training window in realtime is determined by computing the mean between each ZOI update, which sets

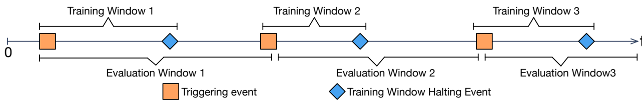

Figure 3: Realtime Evaluation Scheme.

the threshold to limit the window length. In case the threshold is reached without detecting any updates, the training is halted and resumes with the detection of a new update. Thus, the continual tracking of ZOI updates and periodicity ensures the freshness of the predictions.

5. PRELIMINARY RESULTS AND DISCUSSION

We base our experiments on the Nokia dataset [3], consisting of mobility traces of 168 users for a period of at least 30 days. Our evaluation scheme is depicted in Figure 3. The prediction accuracy is considered as the fraction of samples for which the model correctly predicts the next location. We evaluate the entire dataset to determine the parameters of ZOI and set d_{max}, t_{min} and $visitThreshold$ to 60 meters, 900 seconds and 6 visits respectively. To simulate realtime incoming data, we read the data-points sequentially, according to the timestamps.

Prediction technique	Baseline accuracy (%)	Number of users (% days greater than baseline) <=20%	>20% & <=60%	>60%	Number of satisfactory users
1-NN	59.28	168	0	0	101
ANN	60.85	103	65	0	129
RNN	72.79	37	131	0	149
Fourier ext.	63.87	65	103	0	112
1-order MMC	57.19	137	17	14	93
2-order MMC	98.21	158	6	4	142

Table 1: Dataset Analysis.

The preliminary findings of our evaluation are detailed in Table 1. As baselines, we compute the accuracies following the traditional approach of training on 70% data and evaluating on the rest. We observe that, 34% of users reach a satisfactory accuracy (>50% correct predictions for at least 30 days) with less than 100 days of data. We noticed that 2-order MMC and RNN based predictors achieve higher prediction accuracies. In case of 2-order MMC, this is due to accounting for the current state and the previous state to make the prediction. RRN blends the input vector at current state with the previously learnt state vector to yield a new state. Thus taking the entire history into account before making new predictions, effectively combining high level direction with low level modelling. We also saw a clear correlation between the periodicity and the accuracy of classification, neural networks and Fourier extrapolation, as these techniques weigh current state more than the past depicting linearity with the periodicity. In Figure 4, we show the running accuracy difference between our system and for all predictors against the baselines at each training model update for one user. We see this trend across the dataset, where RNN and 2-order MMC achieve better results, sometimes even exceeding the baselines making them ideal predictors for the system. The complexity of a learning model is directly linked to a quadratic equation which involves inverting a matrix having a complexity of the order n^3 and the training time has the order of n^2 where, n is the size of training data. Thus reducing the model complexity as compared to the model formed on 70% of the dataset.

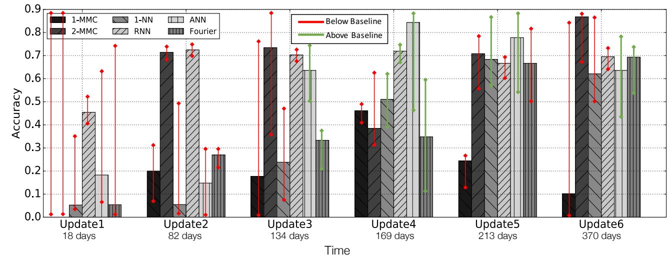

Figure 4: Comparison of our systems' accuracy for individual predictors against the baseline accuracies for one user.

6. CONCLUSION AND FUTURE WORK

With the growing ubiquity of location-aware devices, becoming more powerful everyday, it will become possible to compute mobility predictions locally, without resorting to backend servers. Yet, traditional approaches rely on processing large datasets on powerful servers, which makes the process quite tedious and slow, with the additional concern about location privacy, threatening the widespread adoption of LBS in the coming days. In this poster, we take a step towards this era, forming the prediction model on a handheld device to produce swift realtime predictions. We practically demonstrate that is it possible to achieve satisfactory prediction accuracies, utilising lesser volumes of data, taking the mobility behaviour in to account.

Our future work, will attempt to have an ensemble approach to select suitable predictor in realtime according to behavioural changes, to achieve higher accuracies as we observed that certain family of predictors are better suited for particular mobility behaviours. We will also quantify the computational cost of the approach on an actual mobile device and optimise the process to have fewer model updates, which has a direct impact on the cost.

Acknowledgment. This work is partially supported by the Swiss National Science Foundation grant 146714.

7. REFERENCES

[1] A. Daniel and S. Thad. Using gps to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, pages 275–286, 2003.

[2] V. Etter, M. Kafsi, and E. Kazemi. Been There, Done That: What your Mobility Traces Reveal about your Behavior. *The Proceedings of Mobile Data Challenge by Nokia Workshop at the Tenth International Conference on Pervasive Computing*, June 2012.

[3] J. K. Laurila, D. Gatica-Perez, I. Aad, B. J., O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen. The mobile data challenge: Big data for mobile computing research. In *Pervasive Computing*, 2012.

[4] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti. Wherenext: A location predictor on trajectory pattern mining. KDD '09, pages 637–646. ACM, 2009.

[5] B. Prabhala, J. Wang, B. Deb, T. L. Porta, and J. Han. Leveraging periodicity in human mobility for next place prediction. In *WCNC*, pages 2665–2670, 2014.

[6] L. H. Tran, M. Catasta, L. K. McDowell, and K. Aberer. Next place prediction using mobile data. Mobile Data Challenge by Nokia, 2012.