Spatiotemporal Individual Mobile Data Traffic Prediction

Guangshuo Chen

Spatiotemporal Individual Mobile Data Traffic Prediction

Guangshuo Chen

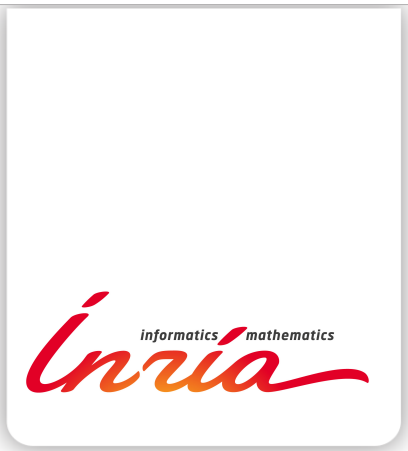
TECHNICAL REPORT

N° 497

Janvier 2018

Project-Teams INFINE

ISSN 0249-0803 ISRN INRIA/RT--497--FR+ENG



Spatiotemporal Individual Mobile Data Traffic Prediction

Guangshuo Chen*†

Project-Teams INFINE

Technical Report n° 497 — Janvier 2018 — 13 pages

Abstract: Understanding the nature of data network traffic is critical in network design, management, control, and optimization. In this report, we leverage two large-scale real-world datasets collected by a major mobile carrier in a Latin American country to investigate the prediction of individual mobile data traffic. Based on our previous analysis on the theoretical predictability, we extend our analysis to the actual prediction and validate the findings, that we have observed in the theoretical analysis, in the actual predicting scenario. We implement the typical algorithms for time series prediction in the literature and test their performance. Then, we propose our algorithms based on state-of-the-art machine learning techniques. Our data-driven test on the performance of these predictors shows that a simple Markov predictor can outperform its legacy counterparts in most of the cases. It achieves a mean accuracy of 62%, but it relies heavily on the historical data and can hardly have an enhancement from knowing individual whereabouts. Our proposed solutions can achieve a typical accuracy of 70%, which outperforms all the legacy ones and have a $1\% - 5\%$ degree of improvement by learning individual whereabouts.

Key-words: Cellular networks; mobile data traffic; prediction

This work was supported by the EU FP7 ERANET program under grant CHIST-ERA-2012 MACACO.

* École Polytechnique, Université Paris Saclay, 91128 Palaiseau, France
† INRIA Saclay-Île-de-France, Université Paris Saclay, 91120 Palaiseau, France

RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE

1 rue Honoré d'Estienne d'Orves
Bâtiment Alan Turing
Campus de l'École Polytechnique
91120 Palaiseau

Prédiction spatio-temporelle du trafic de données mobiles individuelles

Résumé : La compréhension de la nature du trafic du réseau de données est essentielle dans la conception, la gestion, le contrôle et l'optimisation du réseau. Dans ce rapport, nous exploitons deux ensembles de données du monde réel à grande échelle collectés par un opérateur de téléphonie mobile majeur dans un pays d'Amérique latine pour étudier la prédiction du trafic de données mobiles individuelles. Sur la base de notre analyse précédente sur la prévisibilité théorique, nous étendons notre analyse à la prédiction réelle et validons les résultats, que nous avons observés dans l'analyse théorique, dans le scénario de prédiction actuel. Nous implémentons les algorithmes types pour la prédiction de séries chronologiques dans la littérature et testons leur performance. Ensuite, nous proposons nos algorithmes basés sur des techniques d'apprentissage automatique de pointe. Notre test basé sur les données sur les performances de ces prédicteurs montre qu'un simple prédicteur de Markov peut surpasser ses homologues traditionnels dans la plupart des cas. Il atteint une précision moyenne de 62%, mais il repose fortement sur les données historiques et peut difficilement être amélioré à partir de la localisation individuelle. Nos solutions proposées peuvent atteindre une précision typique de 70%, ce qui surpasse tous ceux qui existent déjà et ont un degré d'amélioration de 1% à 5% en apprenant les localisations individuelles.

Mots-clés : Réseaux cellulaires; trafic de données mobiles; prédiction

1 Introduction

The understanding of human behaviors is a central question in many research topics and has contributed to a wide range of applications [1, 2, 3, 4, 5, 6, 7, 8, 9]. In cellular networks, human mobility and mobile data traffic consumption are two significant human habits. The ability to understand them has essential implications in many aspects of cellular networks.

- The high availability of mobility prediction can enable various application scenarios such as location-based recommendation, home automation, and location-related data dissemination and also help improving quality of service [6, 8]. In the literature, a large and growing body of literature has investigated the topic of predicting human mobility [1].
- The better understanding of future mobile data traffic demand can help to improve the design of solutions for network load balancing, aiming at improving the quality of Internet-based mobile services [7, 9]. Compared with the human mobility analysis, a far less group of literature has focus on this topic.

In this report, *we mainly investigate on the topic of understanding and predicting of mobile data traffic*, and put a small attention on the topic of human mobility.

1.1 Studying Data Traffic

Understanding the nature of data network traffic is critical in network design, management, control, and optimization. In general, the better understanding of issues in network traffic leads to more efficient and responsive resource allocation, better QoS, higher efficient ratio channel usage. Hence, since the foundation of the Internet, researchers have put efforts and interests in multiple research disciplines covering local area network (LAN), wide area network (WAN), and more recently, wireless network and cellular network.

1.1.1 Mobile Data Traffic

Due to the wide spreading of mobile phone devices, cellular networks are the most important ones in recent years and the understanding have been carried out at the same time. In general, the studies about cellular networks can be categorized in two typical classes regarding the scale of data traffic, *i.e.*, *aggregated* and *individual* studies.

Aggregated View. In this categories, the studies are to understand mobile data traffic of base stations, applications, and special groups of users, in aggregated views. A summary of the major findings are as follows:

- Characterizations of traffic. The major analyses were carried out in [10, 11, 12] and showed the following main findings. (1) in a base-station level, mobile data traffic have hourly, daily, weekly repetitive patterns [11, 10]. (2) the data usage among adjacent base stations are correlated [11]. Further, spatiotemporal correlation exists in such usage [12, 11].
- Theoretical predictability [13, 14, 15]. This is to analyze theoretical performance that any prediction approaches can achieve in its best performance. In general, this kind of analyses use entropy as measurement of uncertainty of studies phenomena and then use information theory tools to compute a theoretical bounds as predictability. [13, 14] investigated random entropy of base station traffic in three types (voice, text, and data), as well as conditional entropy by temporal, spatial and service related information. They found that knowledge of adjacent cells traffic can enhance the predictability of voice and text more than data,

which supports the spatiotemporal characteristic found in [11]. Besides, [15] studied the traffic in an application level on each base stations.

- Prediction of traffic. (1) Hotspot. [16]performed an empirical study on data hotspots in today's cellular networks using a 9-week cellular dataset with 734K+ users and 5327 cell sites. They show that using standard machine learning methods, future hotspots can be accurately predicted from past observations. (2) base station traffic volumes, Xu:2016eo,Xu:2017ix, showed that in a certain aggregation level, the base station traffic is predictable via a linear combination of four primary components corresponding to human activity behaviors. Shafig *et al.* [17] propose a Zipf-like model to capture the volume distribution of application trafic and a Markov model to capture the volume dynamics of aggregate Internet traffic, a week-long aggregated flow level mobile device traffic data collected from a major cellular operator's core network.

Personal View. The studies in this category aims at understanding data traffic of each individual mobile devices, in other words, personal behaviors when generating mobile data. Compared with the legacy analysis and the aggregated mobile network studies, there are far less studies in this category. A summary is listed as follows:

- `Throughput`. This is is the maximum rate that a mobile device can achieve in data transforming in its network, in other words, the bandwidth. In cellular networks, the throughput depends on the infrastructure (the spec of base stations), the congest of each cell, and the device. It is impacted by both temporal and spatial issues. In individual view, Bui *et al.* [18] proposed a model considering user locations and representing the impact of estimation and prediction errors on the bandwidth availability statistics over a wide range of time scales. In [19], they proposed a refined model aiming at the prediction of short-term thought using Gaussian Random Walks. They tested their models on a theoretical LTE radio model regarding the accuracy of predicting per-user throughput.

- Characterizations of traffic [20, 11, 17, 21]. A group of works have explored this aspect by mining data collected modern 3G networks. They have related important characterizations including (1) heterogeneity in the global usage (a minority of users take accounts for most of the traffic) [20, 11, 21], (2) daily or weekly patterns [11, 21], (3) unclear heterogeneity in age/gender [21]; (4) spatio concentration (most of data traffic is on a small number of locations) [11]. (5) Application concentration (most of a user's data traffic comes from a small number of applications). None of them considered the prediction of per-user mobile data traffic. Besides, Li *et al.* [22] analyzed user behaviors of these three platforms (iphone, andriod, and windows phone) from two aspects: traffic dynamics and user applications, focus on the operating systems.

- Latency [23, 24]. With the wide development of smartphone applications, personal device latencies prediction has been studied for optimizing various service requirement. Liu *et al.* [23] propose new methods for both static latency estimation as well as the dynamic estimation problem given 3D latency matrices sampled over time. We propose a distance-feature decomposition algorithm that can decompose latency matrices into a distance component and a network feature component, and further leverage the structured pattern inherent in the 3D sampled data to increase estimation accuracy. Extensive evaluations driven by real-world traces show that our proposed approaches significantly outperform various state-of-the-art latency prediction techniques. They use matrix completion in prediction. and enhanced their analysis in [24] .

We see that it still lacks the studies of theoretical predictability and actual prediction approaches on the personal view of mobile data network traffic. We have studied the predictability in [25] and focus on the actual prediction in this report.

1.2 Problems and contributions

In our last technical report [25], we have analyzed the theoretical predictability of individual mobile data traffic using tools of information theory. In this report, we push our analysis a step further by proposing the design of practical predictors. In particular, We address the problem of understanding spatiotemporal mobile data traffic demand for individuals, and make the following major contributions:

- We implement the major legacy algorithms for anticipating symbolic time series, and evaluate their performance using extensive tests on large scale mobile phone datasets.
- Our data-driven test on the performance of these predictors show that a simple Markov predictor is able to outperform its legacy counterparts in most of cases. For predicting individual mobile data volumes per hour, the order-2 Markov chain predictor can achieve a mean accuracy of 62%, but it rely heavily on the historical data and can hardly have an enhancement from knowing individual whereabouts, so do the other historical predictors.
- We propose our novel solutions for prediction mobile data traffic via machine learning algorithms
- The data-driven test shows that our proposed solutions can achieve a typical accuracy of 70%, which outperforms all the legacy ones, and have a $1\% - 5\%$ degree of improvement by learning individual whereabouts.
- Interestingly, our analysis show that knowing mobile data traffic of a user can significantly help the prediction of his whereabouts for 50% of the users, leading to an improvement up to 10% regarding accuracy.
- Build upon the results in this report, we confirm the findings about the theoretical predictability.

2 Data preliminary

2.1 Datasets

Our study is based on two real-world datasets describing the cellular network activity of hundreds of thousands of mobile phone subscribers (identically called users) of a major cellular operator in a metropolitan area. All data refers to a consecutive period of 1 year. The first dataset consists of *call detail records (CDRs)* containing timestamped and geo-referenced logs (*i.e.*, of the closest mobile cell tower) of each voice call performed by every user. The second dataset describes the *Internet data sessions* established every time a mobile device needs to exchange IP data traffic through the cellular network.

These two datasets provide different and complementary information: CDR data includes location information that allows reconstructing user mobility, while session data only presents the mobile data traffic volume generated by each subscriber (with no associated geo-referenced log). In both cases, we preprocess the datasets to construct time series of subscriber's locations and data traffic demands that are representative and statistically significant.

CDR dataset: Call detail records are logged every time a mobile device makes or receives a voice call. Each entry contains the hashed identifiers of the caller and callee, the call duration in seconds, the timestamp of the call start time and the location (latitude and longitude) of the cell tower to which the device is connected when initiating the phone call.
Internet data session dataset: Every Internet data session is established upon the allocation of a radio channel for the exchange of IP traffic, and it ends after an idle period over the same channel. Each entry in the dataset contains the hashed device identifier. The same hashing function is used in the CDR and Internet data session datasets, which allows linking users in the two datasets. The volume of upload and download data exchanged in KiloBytes, and the timestamp denoting the starting time of the session. The dataset does not contain spatial information.

2.2 Extracting historical whereabouts and mobile data volumes for individuals

To have a fair population of study, we select from the common users of both datasets given the following conditions:

- selecting weekdays only,
- excluding public holidays,
- having locations in at least 20% of the total hours,
- having volumes in 150 weekdays.

Following the above conditions, we select approximately $6,200$ users as our population of study.

3 Recall the limits of predictability of mobile data traffic

We have already evaluated the theoretical predictability, *i.e.*, the maximum accuracy that any algorithm has potential to achieve in the prediction of individual mobile data volumes [25], on the same group of users as introduced in Sec. 2. As necessary knowledge, we recall the major findings of the predictability analysis as presented in [25]:

- We find that, by just considering temporal correlations in the traffic, 85% of the activity of each user can be anticipated on average.
- We prove the result above to hold across heterogeneous classes of subscribers, based on age, gender, mobility, or mobile service usage.
- We observe a 90% potential predictability of the spatiotemporal data consumption patterns of individual users. This result is due to the strong correlation between mobility and mobile service usage, *i.e.*, to the fact that subscribers tend to generate similar amounts of traffic at each location. This suggests the feasibility of anticipating how much mobile data traffic (as an order of magnitude) will be consumed by a given subscriber and where this will occur in a very effective manner, by knowing the past history of activities of the target individual. Although spatiotemporal information of subscribers can further improve the design of prediction model, we also find that the gain is not dramatic with respect to a technique that only relies on temporal information (*i.e.*, 85%).

In this report, we put the theoretical results above into the practical performance. In particular, we propose actual predictors and evaluate their performance on the same users, which we present in detail in the following sections.

4 Prediction methods

In this thesis, we study the prediction of time series of a set of discrete observations and use the following predictors.

4.1 Traditional time series predictors

4.1.1 Markov model

A *Markov* model is a stochastic model used to model random systems or describe series of observations [26]. In a Markov model describing a time series of observations, each unique observation is modeled as a *state*, the probabilities of transitions from one state to another are learnt from historical observations as transition matrix. The Markov model has different *orders* due to time steps considered to build states. In a k-th order Markov model, the probability of current state only depends on the past k states, *i.e.*, $P(X_t|X_{t-1}, X_{t-2}, \cdots, X_1) = P(X_t|X_{t-1}, \cdots, X_{t-k})$ for any t.

The Markov model is efficient describing time series of discrete observations, *e.g.*, locations, and thus is often used in predicting human locations as in [27]. It can also be used in modeling and predicting mobile data traffic, *e.g.*, time series of symbolized volumes (or magnitudes).

4.1.2 Text Compression Models

Besides the Markov models, the following predictors are also considered: prediction by partial matching (PPM), sampled pattern matching (SPM) and *Active LeZi* (ALZ). They are originally designed for text compression to predict appearance of characters or symbols. They seem promising and utilized in this thesis because good text compressors are mostly good predictors [28].

Prediction by Partial Matching (PPM) PPM is a data compression scheme massively used in lossless text compression [29]. It is an adaptive technique based on combining Markov models of different orders to estimate probabilities of the next character in the input text. As in a Markov model, an PPM predictor also has an order k and builds its model by the historical sequence. It computes the next character probabilities from all k-order to 0-order Markov models through so-called *escape* probabilities. There are multiple PPM implementations particularly on the calculations of escape probabilities. In this thesis, we prefer the implementation of Moffat *et al.* [29], also called "Method C". We use maximum likelihood estimation to compute probabilities of the next symbol.

Sampled Pattern Matching (SPM) SPM is a pattern matching predictor proposed by Jacquet *et al.* [30]. It considers much larger immediately preceding symbols for predicting the next symbol than Markov-based predictors. In an SPM predictor, instead of using an fixed and finite order k (or other others less than k as in PPM), the considered length of immediately preceding context is determined as the fixed fraction (as a parameter α) of the longest context which has previously appeared.

As an example, consider the text prediction problem from the following time sequence:

SLJZGGDLYGSJSLJZKGSSLJZIDSLJZGGZYGSJSLJZ.

The longest context that has been previously seen is "YGSJSLJZ". In the SPM predictor with $\alpha = 0.5$, the considered context is the fractional suffix "SLJZ". Since "SLJZG" has appeared 2 times and the rest ("SLJZ" with "I","K") does only one time, the predictor will predict "G" as the next symbol.

Active LeZi (ALZ) ALZ is a prediction algorithm based on the classical LZ78 data compression scheme, proposed by [31]. Like other predictors in the LZ family, the ALZ algorithm utilizes the power of Markov models to predict the next symbol in the time sequence given the preceding context. It is designed as an *online* algorithm because it is able to incrementally learning the sequence and deliver real time predictions. In an ALZ predictor, a variable window of immediately preceding symbols is maintained, of which the length is the longest phrase previously observed in a classical LZ78 parsing. With this window, the algorithm can compute statistics on all possible preceding contexts, which leads to a better approximation than Markov models in theory. For the pseudo code of this algorithm, we refer the reader to [31, Figure 3].

4.2 Machine learning predictor

We use the multilayer perceptron as our major technique for the design of our predictors. The MLP algorithm is a supervised learning algorithm for both regression and classification problems. A MLP network is an feedforward artificial neural network and is known as the first of its kind. In an MLP predictor, the neuron network is trained by a training set and generates predictions from the input of the testing set. The MLP algorithm accepts different activation functions, layers and neurons.

5 Methodology of prediction analysis

We test all techniques for predicting individual mobile data traffic volumes on our population of study. Recall that, each user has two separate time series of locations and volumes in each hour. Both time series cover the same period of 150 weekdays.

The experiments are performed on each user. Given a user, a target (volumes or locations) to predict, and a prediction technique, we proceed as follows:

1. Initialize the prediction model of the technique using the time series of the earliest 50 days. Set $d = 51$.
2. Predict the target of the 24 hours of d-th day.
3. Update the model using the data in the d-th day. Set $d = d + 1$.
4. Go to 2 if $d <= 150$, otherwise exit.

In other words, we predict the target in the latest 100 days and perform an online updating scheme. For the performance, we use the average accuracy over the predicted data in the 100 days.

6 Temporal mobile data volume prediction

We observe the following in Fig. 1:

- Due to the nature of individual mobile data traffic usage, on average, a user has no traffic in 50% of the users, so even a predictor only predicts 0 can achieve a high accuracy.

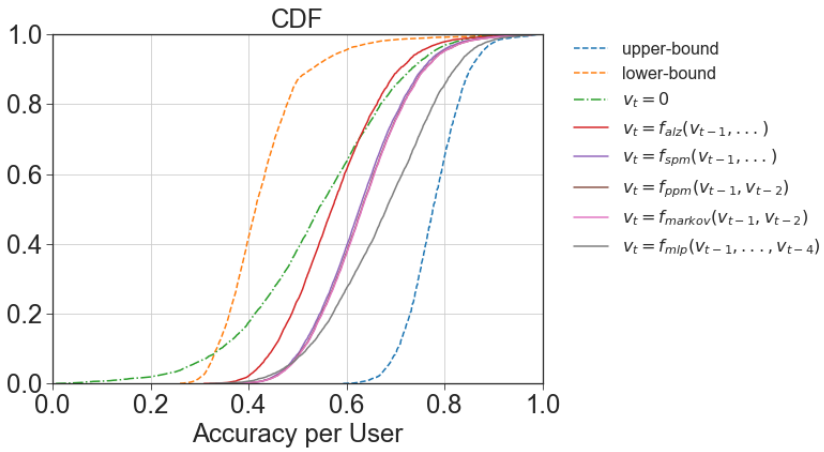


Figure 1: Distribution of the accuracy in the prediction of individual mobile data volume per hour.

- The Markov predictor can achieve a median accuracy of 62% in prediction, while the theoretical upper bound is 75%.
- The Markov enhanced (PPM, SPM, and ALZ) predictors perform worse than the Markov predictor. We guess this is because of the lack of samples. This result is consistent to the result of predicting locations
- The MLP predictor is the best one, which achieve a median accuracy of 70%, closing to its upper bound.

7 Spatiotemporal mobile data volume prediction

As we see in the theoretical analysis that, knowing locations should contribute to the prediction of volume (having a theoretical performance gain up to 5%), hence we test our predictors in this case.

We consider the scenario of predicting the current mobile data volume by knowing the past volumes and locations. We observe the following in Fig. 2 and 3:

- In general view, only the ALZ predictor has a clear enhancement around 3%. In individual view, 75% of the users benefit from knowing locations and have enhancement in the range of $0 - 7\%$ regarding accuracy, for the rest of users, they have worse performance.
- The Markov and Markov-enhanced predictors can not have enhancement from knowing locations, only 10% of the users have enhancement up to less than 5%. We think this is because of the lack of samples.
- In the MLP predictor, 50% of the users can have enhancement up to 10%, but since the rest perform worse, we cannot see a clear enhancement in the CDF plot. We believe that given more samples, this users can also be enhanced (prove this).
- Also in the MLP predictor, we see in Fig. 4 that knowing current location can better enhance the prediction in volume, which show an 2% enhancement.

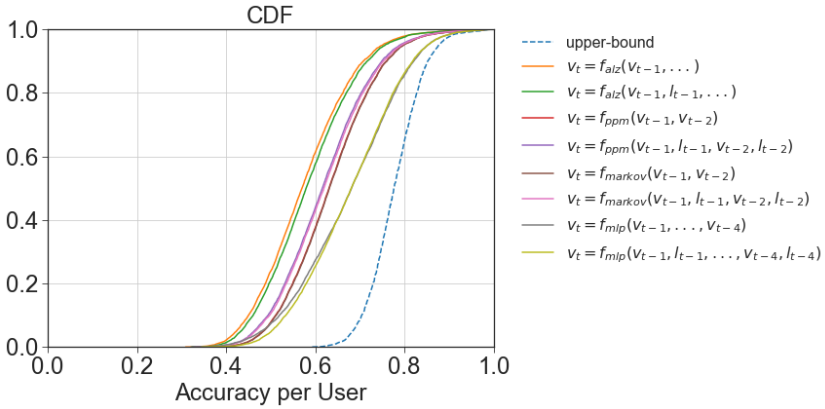


Figure 2: Distribution of the accuracy in the prediction of individual mobile data volume per hour from the history of both volumes and locations

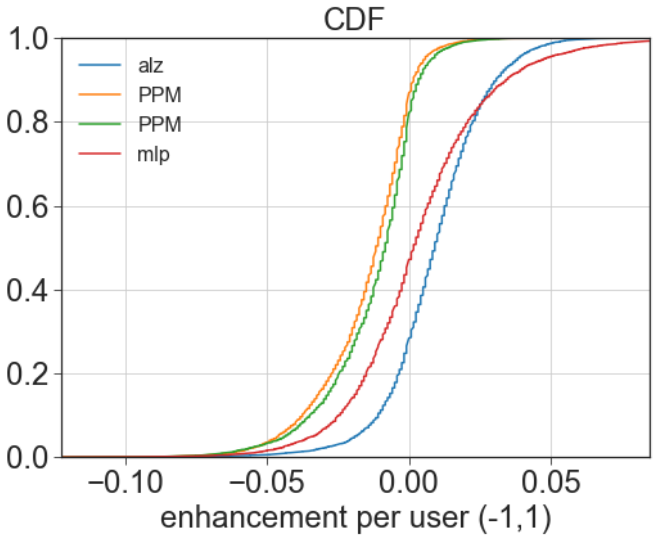


Figure 3: Distribution of the enhancement per user in the prediction of individual mobile data volume per hour from the history of both volumes and locations

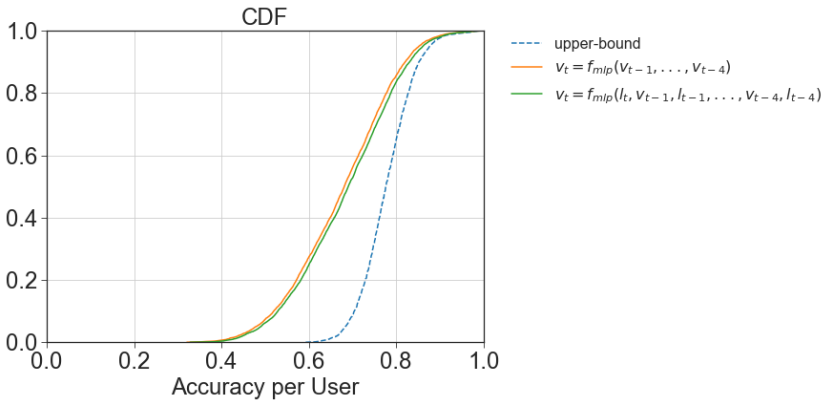


Figure 4: Distribution of the enhancement per user in the prediction of individual mobile data volume per hour from the history of both volumes and locations, and the current location

References

[1] D. Naboulsi, M. Fiore, S. Ribot, and R. Stanica, "Large-scale Mobile Traffic Analysis: a Survey," *IEEE Communications Surveys & Tutorials*, vol. PP, no. 99, pp. 1–1, 2015.

[2] W. Su, S.-J. Lee, and M. Gerla, "Mobility prediction in wireless networks," in *IEEE MILCOM 2000*, vol. 1, pp. 491–495, IEEE, 2000.

[3] P. N. Pathirana, A. V. Savkin, and S. Jha, "Mobility modelling and trajectory prediction for cellular networks with mobile base stations," in *ACM MobiHoc 2003c*, pp. 213–221, ACM, 2003.

[4] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of Predictability in Human Mobility," *Science*, vol. 327, pp. 1018–1021, Feb. 2010.

[5] D. G. Taylor and M. Levin, "Predicting mobile app usage for purchasing and information-sharing," *International Journal of Retail & Distribution Management*, vol. 42, no. 8, pp. 759–774, 2014.

[6] W.-S. Soh and H. S. Kim, "Qos provisioning in cellular networks based on mobility prediction techniques," *IEEE Communications Magazine*, vol. 41, no. 1, pp. 86–92, 2003.

[7] H. Petander, "Energy-aware network selection using traffic estimation," in *ACM MICNET 2009*, pp. 55–60, ACM, 2009.

[8] V. A. Siris and D. Kalyvas, "Enhancing mobile data offloading with mobility prediction and prefetching," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 17, no. 1, pp. 22–29, 2013.

[9] Z. Li, J. Bi, and S. Chen, "Traffic prediction-based fast rerouting algorithm for wireless multimedia sensor networks," *International Journal of Distributed Sensor Networks*, 2013.

[10] F. Xu, Y. Lin, J. Huang, D. Wu, H. Shi, J. Song, and Y. Li, "Big Data Driven Mobile Traffic Understanding and Forecasting: A Time Series Approach," *IEEE Transactions on Services Computing*, vol. 9, pp. 796–805, Aug. 2016.

[11] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks.," *INFOCOM*, pp. 882–890, 2011.

[12] I. Trestian, S. Ranjan, and A. Kuzmanovic, "Measuring serendipity: connecting people, locations and interests in a mobile 3G network," in *Proceedings of the 9th ...*, 2009.

[13] X. Zhou, Z. Zhao, R. Li, Y. Zhou, and H. Zhang, "The predictability of cellular networks traffic," in *2012 International Symposium on Communications and Information Technologies (ISCIT)*, pp. 973–978, IEEE, 2012.

[14] R. Li, Z. Zhao, X. Zhou, J. Palicot, and H. Zhang, "The prediction analysis of cellular radio access network traffic: From entropy theory to networking practice.," *IEEE Communications Magazine ()*, vol. 52, no. 6, pp. 234–240, 2014.

[15] R. Li, Z. Zhao, J. Zheng, Y. Chen, C. Mei, Y. Cai, and H. Zhang, "The Learning and Prediction of Application-level Traffic Data in Cellular Networks," *arXiv.org*, June 2016.

[16] A. Nika, A. Ismail, B. Y. Zhao, S. Gaito, G. P. R. 0001, and H. Zheng, "Understanding and Predicting Data Hotspots in Cellular Networks.," *MONET*, vol. 21, no. 3, pp. 402–413, 2016.

[17] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, "Characterizing and modeling internet traffic dynamics of cellular devices.," *SIGMETRICS*, pp. 305–316, 2011.

[18] N. Bui, F. Michelinakis, and J. Widmer, "A model for throughput prediction for mobile users," *European Wireless 2014; 20th ...*, 2014.

[19] N. Bui and J. Widmer, "Modelling Throughput Prediction Errors as Gaussian Random Walks," Sept. 2014.

[20] C. L. Williamson, E. Halepovic, H. Sun, and Y. Wu, "Characterization of CDMA2000 Cellular Data Network Traffic.," *LCN*, pp. Z000–719, 2005.

[21] E. Mucelli, A. C. Viana, K. P. Naveen, and C. Sarraute, "Mobile Data Traffic Modeling: Revealing Temporal Facets," pp. 1–14, Apr. 2015.

[22] Y. Li, J. Yang, and N. Ansari, "Cellular smartphone traffic and user behavior analysis," in *ICC 2014 - 2014 IEEE International Conference on Communications*, pp. 1326–1331, IEEE, 2014.

[23] B. Liu, D. Niu, Z. Li, and H. V. Zhao, "Network latency prediction for personal devices: Distance-feature decomposition from 3D sampling," in *IEEE INFOCOM 2015 - IEEE Conference on Computer Communications*, pp. 307–315, IEEE, 2015.

[24] R. Zhu, B. Liu, D. Niu, Z. Li, and H. V. Zhao, "Network Latency Estimation for Personal Devices - A Matrix Completion Approach.," *IEEE/ACM Trans. Netw.*, 2017.

[25] G. Chen, S. Hoteit, A. Carneiro Viana, M. Fiore, and C. Sarraute, "Spatio-Temporal Predictability of Cellular Data Traffic," Research Report RT-0483, INRIA Saclay - Ile-de-France, Jan. 2017.

[26] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

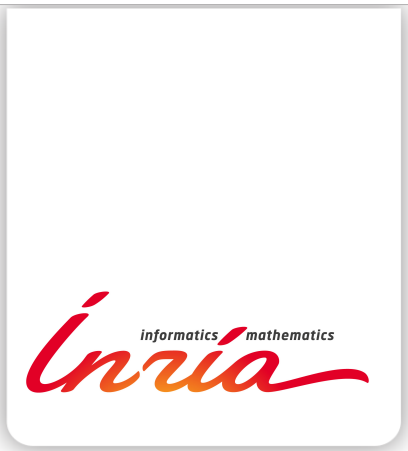
[27] L. Song, D. Kotz, R. Jain, and X. He, "Evaluating next-cell predictors with extensive Wi-Fi mobility data," *IEEE Transactions on Mobile ...*, 2006.

[28] J. S. Vitter and P. Krishnan, "Optimal prefetching via data compression," *Journal of the ACM (JACM)*, vol. 43, no. 5, pp. 771–793, 1996.

[29] A. Moffat, "Implementing the ppm data compression scheme," *IEEE Transactions on communications*, vol. 38, no. 11, pp. 1917–1921, 1990.

[30] P. Jacquet, W. Szpankowski, and I. Apostol, "An universal predictor based on pattern matching, preliminary results," *Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities*, pp. 75–85, 2000.

[31] K. Gopalratnam and D. J. Cook, "Active lezi: An incremental parsing algorithm for sequential prediction," *International Journal on Artificial Intelligence Tools*, vol. 13, no. 04, pp. 917–929, 2004.



RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE

1 rue Honoré d'Estienne d'Orves
Bâtiment Alan Turing
Campus de l'École Polytechnique
91120 Palaiseau

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-0803