Analyzing Stochasticity in Mobility Profiles

January 15, 2018

Theoretical significance and importance of the study

The pervasiveness of mobile devices equipped with internet connectivity and global positioning functionality (GPS) has resulted in the generation of large volumes of human **mobility data**. This data is used to construct **mobility models** of individuals, which are essential for applications such as urban planning, traffic management, consumer profiling and targeted advertisements [4]. These models are formulated by capturing vital information regarding user mobility such as movement velocity, changes in direction, number of visited places, periodicity of visitation and others alike [8]. Our research work mainly revolves around constructing the mobility models in an efficient and privacy preserving manner for applications such as mobility prediction useful for improving location-based services. Location-based services provide contextual information to users depending upon their location, typical examples include Uber and Google Maps.

In the literature, it has been extensively shown that human mobility is repetitive and highly predictable with regards to the temporal and spatial regularities [1]. Such regularities result in a problem known as **overfitting**, where a model only learns the dominant trend observed in a population moving in a given area [2]. Overfitting leads to poor performance in applications such as mobility prediction [3], location recommendation [9] and synthetic mobility traffic generation [5]. This occurs when regular users change their behavior on account of personality trait of neophilia, where people show propensities of novelty seeking [6]. In order to avoid this problem of overfitting, some amount of stochasticity has to be added to the dataset under consideration [7] . Adding mobility profiles having such stochastic behavior increases the entropy of the general population and hence is known to eliminate this problem to a certain degree. According to our observation, tourists who are on short term visits and students on an exchange program show this required level of stochasticity in their mobility behaviors. As a result, we plan to collect mobility traces of 50 students who are on an exchange program. This dataset will help us to answer the following questions:

- Do mobility profiles of exchange students have a higher level of stochasticity in comparison to regular students? If yes, then to what extent? (quantified in terms of the number of visited places, transportation modes used, movement periodicity and level of entropy).
- Does adding such profiles to a dataset of regular user pool eliminate the problem of generating homogenous synthetic mobility traffic? The synthetic traffic is generated by applying Deep Learning on the trajectories of subjects. Such synthetic datasets address the lack of mobility datasets available in certain cities.
- Do such profiles improve point of interest recommendation schemes formulated by using techniques such as collaborative filtering? Collaborative filtering is a method of making au-

tomatic location predictions about the interests of a user by collecting location preference information from many users (collaborating).

In addition, this dataset will be used to correlate another traffic related dataset to improve the movement of public transport controllers in the city of Lausanne.

Description of the empirical set-up and study design

After the selection of all participants, we will ask them to answer a preliminary survey to collect demographic data and information regarding their mobility behavior (gender, age, working type, means of transportation, etc.) as well as their perception of location privacy. We will provide the participants an iOS application that will run on their smartphones during the duration of the data collection campaign. The location data of each participant will be frequently captured and stored at a server located at UNIL. At the end of this data collection campaign, we will ask the participants to answer a last survey in order to collect a ground truth composed of their points of interest as well as their new perception of location privacy. Complete anonymity will be maintained during the entire data collection process. **This study is accepted by the ethics committee.**

The collected data will then be analyzed to formulate the mobility models based on our previous work. These models will be added to the dataset pool we already have to study the stochasticity impact on the following four projects to answer the aforementioned questions.

1. On the difference between the mobility profiles of regular students and students of an exchange program.
2. Generating synthetic mobility traffic by applying deep learning on user trajectories.
3. Privacy aware collaborative filtering for improving mobility prediction.
4. An efficient model for public transport controller movement by randomization in space and time.

In addition to the above projects, the insights gained from the analysis can be useful to universities to improve the exchange programs by catering to the students requirements.

References

[1] Song, Chaoming, Zehui Qu, Nicholas Blumm and Albert-Laszlo Barabasi. Limits of predictability in human mobility. Science 327 5968 (2010): 1018-21.

[2] Schmidhuber, Jurgen. Deep Learning in Neural Networks: An Overview. Neural networks : the official journal of the International Neural Network Society 61 (2015): 85-117.

[3] Wang, Yingzi, Nicholas Jing Yuan, Defu Lian, Linli Xu, Xing Xie, Enhong Chen and Yong Rui. Regularity and Conformity: Location Prediction Using Heterogeneous Mobility Data. KDD (2015).

[4] Zheng, Yu, Quannan Li, Yukun Chen, Xing Xie and Wei-Ying Ma. Understanding mobility based on GPS data. UbiComp (2008).

[5] Uppoor, Sandesh, Oscar Trullols-Cruces, Marco Fiore and Jose M. Barcelo-Ordinas. Generation and Analysis of a Large-Scale Urban Vehicular Mobility Dataset. IEEE Transactions on Mobile Computing 13 (2014): 1061-1075.

[6] Noulas, Anastasios, Salvatore Scellato, Neal Lathia and Cecilia Mascolo. Mining User Mobility Features for Next Place Prediction in Location-Based Services. 2012 IEEE 12th International Conference on Data Mining (2012): 1038-1043.

[7] Gal, Yarin, Yutian Chen, Roger Frigola, Shane Gu, Alex Kendall, Yingzhen Li, Rowan McAllister, Carl Rasmussen, Ilya Sutskever, Gabriel Synnaeve, Nilesh Tripuraneni, Richard Turner, Oriol Vinyals, Adrian Weller, Mark van der Wilk and Yan Wu. Uncertainty in Deep Learning. (2016).

[8] Andrea Hess, Karin Anna Hummel, Wilfried N. Gansterer, and Gunter Haring. 2015. Data-driven human mobility modeling: A survey and engineering guidance for mobile networking. ACM Comput. Surv. 48, 3, Article 38 (December 2015), 39 pages.

[9] Lian, Defu, Yong Ge, Fuzheng Zhang, Nicholas Jing Yuan, Xing Xie, Tao Zhou and Yong Rui. Content-Aware Collaborative Filtering for Location Recommendation Based on Human Mobility Data. 2015 IEEE International Conference on Data Mining (2015): 261-270.