Breadcrumbs

Data Collection Campaign - 2018

Applicants

Garbinato Benoît and Huguenin Kévin (from the Information System department).

PhD students: Chapuis Bertil, Kulkarni Vaibhav, Moro Arielle (under the supervision of Benoît Garbinato) and Olteanu Alexandra-Mihaela (under the supervision of Kévin Huguenin).

Motivation

Exploring human mobility while preserving location privacy is a rich field of analysis, which is connected with data mining, distributed systems, security and privacy domains. In the big data era, it is crucial to be aware of all the novel techniques related to these fields in order to investigate new research questions and to devise new geolocated services that ensure location privacy of end users.

In order to properly evaluate artefacts in these domains, we usually rely on mobility traces from public [3, 4, 5] or private [1, 2, 6] datasets (under licence). However, we observed that the quality of existing datasets is not optimal. For example, they contain gaps ranging from several hours to several days, they do not contain a ground truth (an accurate list of points of interest) to check our assumptions and they do not all contain demographic data as well as preferred transportation modes. In addition, users involved in the datasets do not necessarily have social links between each other, which can limit their use in studies related to location privacy. For all these reasons, we want to launch a data collection campaign called Breadcrumbs. This new dataset will address the aforementioned issues and benefit the entire community working with geospatial datasets. Moreover, we also highlight that the ethical committee of HEC already approved this project at the end of August 2017. The research questions we want to explore and address are listed below.

Research questions

1. Is it possible to accurately predict the future location of a user within a short, a medium and a long-term range?
2. How can we classify the users by exploring their movements?
3. How auxiliary information about location data can help in inferring social relationships among users?
4. Can aggregated location traces (consisting of GPS coordinates collected at a high sampling rate) be separated into individual user trajectories?
5. Which features are better suited to train a machine-learning model for mobility modeling?
6. How much does a prediction model trained on aggregated data (multiple users) contribute towards increasing the prediction accuracy of individual users?

Study design and procedure

The goal of this data collection campaign is to collect location data of individuals on a daily basis as well as additional data related to their mobility during a period of six months (from the beginning of March 2018 to the end of June 2018). Location data will be automatically collected via an iOS application, while additional data (ground truth, demographic data, transportation modes) will be collected at the beginning, at intermediate steps during the six months and at the end of the data collection campaign.

<u>Selection of the participants</u>

We will recruit participants via the ORSEE database and select them based on a survey targeted at their mobility behavior. Participants will also have to select a friend and a relative who agree to participate in the data collection campaign. Our goal is to gather participants who demonstrate disparate mobility patterns depending on their location traces and their modes of transportation (students at EPFL, students at UNIL, students only using their cars, students using different means of transportation…).

We will also ask them if they intend to change their smartphones or if they will be abroad for a long time or simply in vacation during the six months. These two aspects could have a strong impact on the quality of the dataset.
In order to eventually obtain a dataset consisting of at least 80 high-quality users, we will recruit 100 participants in total.

Collecting procedure

Kickoff phase. Directly after the selection of the 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 also ask the participants if they know other participants involved in the data collection campaign, i.e., friends or relatives for the relationship ground truth.

Collecting phase. We will provide the participants an iOS application that will run on their smartphone during the duration of the data collection campaign. The location data of each participant will be frequently captured and sent to a server located at UNIL to store them in a safe manner with restricted access. In addition to the GPS data, we will also frequently collect WiFi and Bluetooth data. At the end of the collection phase, we will automatically retrieve their contact information and their agenda information for the period of the data collection campaign in order to be able to answer privacy related questions. Finally, once a month, we will ask them to answer a survey regarding the modes of transportation they used.

Closing phase. 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 about location privacy. For example, the participants will have to indicate the locations where they spent a large part of their time (e.g., home place, work place…) or locations that are important for them on a daily basis (e.g., bus stop…). They will also have to label them in order to improve the semantic of the dataset.

Data quality

As highlighted previously, the quality of the collected location data is crucial.
To avoid large gaps in the data, the participants will regularly receive notifications (by e-mail or via the application). These notifications will typically be sent If they have a low battery level or if we detect that they did not recorded their location during a long period of time. These notifications will mitigate the risk of obtaining low-quality traces.

Data privacy

In order to ensure the privacy of each participant, the data will be stored on a secure server located at Unil. Furthermore, proven cryptographic techniques and strict access control policies (allowing only a few privileged researchers at UNIL to access the data) will be enforced. If we decide to share the entire dataset with the research community, it will be distributed according to a UNIL licence with strict anonymity constraints (names will be removed, etc.).

Monetary compensation

All participants, of the ORSEE database or recruited by an ORSEE participant, will receive a monetary compensation of 100.- CHF if and only if participants follow all the rules indicated in the consent form of the Data Collection Campaign.

Papers

[1] N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez, and J. Laurila
Towards Rich Mobile Phone Datasets: Lausanne Data Collection Campaign
Proc. ACM Int. Conf. on Pervasive Services (ICPS), Berlin, Jul. 2010

[2] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T-M-T. Do, O. Dousse, J. Eberle and M. Miettinen

The Mobile Data Challenge: Big Data for Mobile Computing Research
Proc. Mobile Data Challenge Workshop (MDC) in conjunction with Pervasive, Newcastle, 2012

[3] Yu Zheng, Lizhu Zhang, Xing Xie, Wei-Ying Ma
Mining interesting locations and travel sequences from GPS trajectories
In Proceedings of International conference on World Wild Web (WWW 2009), Madrid Spain. ACM Press: 791-800

[4] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, Wei-Ying Ma
Understanding Mobility Based on GPS Data
In Proceedings of ACM conference on Ubiquitous Computing (UbiComp 2008), Seoul, Korea. ACM Press: 312-321

[5] Yu Zheng, Xing Xie, Wei-Ying Ma
GeoLife: A Collaborative Social Networking Service among User, location and trajectory
Invited paper, in IEEE Data Engineering Bulletin. 33, 2,2010, pp. 32-40

[6] Sonia Ben Mokhtar, Antoine Boutet, Louafi Bouzouina, Patrick Bonnel, Olivier Brette, et al.
PRIVA'MOV: Analysing Human Mobility Through Multi-Sensor Datasets
NetMob 2017, April 2017, Milan, Italy

[7] Arielle Moro and Benoît Garbinato
A System-level Architecture for Fine-grained Privacy Control in Location-based Services
Proceedings of the 12th European Dependable Computing Conference (EDCC'16)

[8] Vaibhav Kulkarni*, Arielle Moro*, Benoît Garbinato (*co-first authors)
Poster: A Mobility prediction System Leveraging Real-Time Location Data Streams
Proceedings of the 22nd ACM International Conference on Mobile Computing and Networking (MobiCom'16)

[9] Vaibhav Kulkarni*, Arielle Moro*, Benoît Garbinato (*co-first authors)
MobiDict – A Mobility Prediction System Leveraging Realtime Location Data Streams
Proceedings of the 7th ACM SIGSPATIAL International Workshop on GeoStreaming (IWGS'16), San Francisco, California, USA

[10] Bertil Chapuis, Arielle Moro, Vaibhav Kulkarni, Benoît Garbinato
Capturing Complex Behaviour for Predicting Distant Future Trajectories
Proceedings of the 5th ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems (MobiGIS'16), San Francisco, California, USA

[11] Arielle Moro and Benoît Garbinato
A Location Privacy Estimator based on Spatio-temporal Location Uncertainties
Proceedings of the International Conference on NETworked sYStems (NETYS'17)

[12] Vaibhav Kulkarni, Arielle Moro, Bertil Chapuis, Benoît Garbinato
Extracting Hotspots without A-priori by Enabling Signal Processing over Geospatial Data Proceedings of the 6th ACM SIGSPATIAL International Conference (SIGSPATIAL'17), Los Angeles Area, California, USA

[13] Bertil Chapuis, Benoît Garbinato, Lucas Mourot
A Horizontally Scalable and Reliable Architecture for Location-based Publish-Subscribe Proceedings of the 36th IEEE International Symposium on Reliable Distributed Systems, Hong Kong, China