Predictability of Human Mobility from Highly Granular Location Data

Rafaela-Ioana Voiculescu

Kongens Lyngby 2014
IMM-MSc-2014-????

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Matematiktorvet, building 303B,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3351
compute@compute.dtu.dk
www.compute.dtu.dk IMM-MSc-2014-????

Summary

TODO - The goals of this thesis is to..

Preface

This thesis was prepared at the department of Informatics and Mathematical Modelling at the Technical University of Denmark in fulfilment of the requirements for acquiring an M.Sc. in Informatics.

The thesis deals with ...

The thesis consists of ...

Lyngby, 01-August-2014

Rafaela-Ioana Voiculescu

Acknowledgements

I would like to thank my....

Contents

Chapter 1

Introduction

TODO

Chapter 2

Related work

There is a high interest and a huge amount of work the scientific community dedicates to understanding the patterns of human mobility. The knowledge we can gain from the results of this work has the potential to benefit a wide variety of industries from the modeling and maintenance of the transportation infrastructe, to the medical industry where we can use this knowledge in trying to prevent the spreading of epidemies. **??**

Various studies have been conducted in order to gain a better understanding of the human mobility patters. These studies give us results that seem to support each other in the idea that people are less spontaneous than they would like to think themselves and that, indeed, our behaviour shows that we are quite rooted into habbits when it comes to the way we travel.

2.1 Mobility patters uncovered by the disipation on bank notes

Brockmann, Hufnagel and Geisel**??** have analyzed the human movement based on the way bank notes were dispersed through the United States (excluding Alaska and Hawaii). Their study shows that a relatively small percentage of

bank notes (23.6%) traveled for more than 800 km, while a fraction of 19.1% did not traveled for more than 50 km even after a year of being observed. The possible explanation the authors have given for these findins are that, in general, people would be less inclined to leave the areas of the large cities or the places they usually conduct their lives.

The problem identified with this approach for tracking individuals is that the bank notes exachange hands and the behaviour which is identified by the way they circulate can't be attributed to a single individual, but rather to different ones that at any moment have had the bank note in their posession. Despite this, the result have a high scientific value as they do identify patterns in human travel behaviours in general.

2.2 Mobility patterns of mobile phone users

A. L. Barabasi, M. C. Gonzalez and C. A. Hidalgo have conducted a study **??** that deals with studying the trajectories of over 100000 mobile phone users with anonymized identities. The study was conducted in order to see if there are any patterns in our mobility habbits. Among the things that have been subjected to testing was the return probability of individuals in the same place as in the past. The study shows there is, in general, a peak in the return probability after 24, 48 or 72 since they have left a particular location. This shows that we humans tend to visit locations periodically. This can be explaineQd by our going to places such as work, school, grocery shops near our home etc.

The authors have also ranked the locations the mobile phone users frequented based on the number of times they have been spotted nearby. The results for this have shown that the probability of finding someone near a location that is ranked for them with a level L can be estimated with $1/L$. Another interesting finding that is mentioned in the paper is that, in general, people seem to be spending the majority of their time in just a few locations, while diving the remaining time just between a limited number of locations that varies for the subjects from as low as 5 to around 50.

There are some note worthy plots that the authors present in the paper. They can be seen in figure and they show that most people travel over short distances, yet there is a small number of people that regularly travel over big distances.

The results of this study are a major indicator that individuals display a high level of regularity and that we have a tendency to spend most of our times in places that are familiar to us, or that require us to visit them regularly (e.g.

home, work).

2.3 Mobility patterns in massive multiplayer online games

R. Sinatra and M. Szell have studied the way in which users of a massive multiplayer online game behave inside the virtuale universe provided by the mentioned game ??. It has been established that the massive multiplayer games provide people with a virtual reality where they can interact with others through their characters and can, in fact, form groups and, as such, display both individual as well as collective behaviour actions that can translate to the non-virtual world ??.

This study gives an interesting insight into the habbits and actions of the characters which are controlled by the players. Among the things the authors have analyzed are the predictability of the characters, the entropy generated by the mobility of the characters in the virtual universe and general strategies or patterns that could be observed.

The game the authors have been using for the study is called Pardus ??. This game is quite complexe, as it allows the manifestation of normal real-life activities such as the creation of alliances or friendships, communication between the players, economic related action, or even actions which have a negative conotation such as attack of another user, removal of a friendship link etc. The universe of the game consists in hundrets of nodes which represent cities or sectors in the game. These virtual cities are tied to each other through links which mark the posibility for the users to move their characters from one place to another.

By analyzing the why in which characters have interacted through the years, the authors have observed that the mobility of the characters through the univers is highly predictable, as users in general will seem to be chosing a random location to visit next in just about 10% of the cases.

2.4 Eigenbehaviours

N. Eagle and A. S. Pentland analyze data of individulas and communities with the purpose of trying to predict and cluster the daily habbits and behaviour of people ??. The consider that the behaviour of one person throughout a day can

be close to a sum of their primary eigenbahaviours throughout that day. The results of the study have shown that when having a weighted sum calculated for the first half of a day, the behaviour of the same person throughout the remaining of the day can actully be approximated with 79% accuracy.

The results have applicability in more fields, as they allow us to consider the possibility of clustering people into various communities based on the similarity of their behaviours. It goes even further, as the findings show that this enables the possibility of calculating similarity for groups as well and thus permiting the a classification that, according to the experiment, can be 96% accurate for determining affiliations in the social network of a particular population.

As a last observation in the paper by N. Eagle and A. S. Pentland it is stated that eigenbehaviours can be used in order to identify the possible friendship ties between people. The observations in this paper have been done based on the Reality Mining dataset that tracked the behavior for 100 individuals at MIT for the duration of one year.

2.5 Human movement recorded through real traces

Studies as the ones with the travel of bank notes or the recorded location of mobile users through telephone is not very exact and does not reflect the real traces for the people. They do provide a very useful estimation, however with the technology that we have access to nowadays, we are able to record mobile phone users' real traces either through GPS or WiFi. The data that can be acquired through these means allows us to conduct studies that can take into consideration a very good approximation of the real location of individuals.

In the paper by M. Kim, D. Kotz and S. Kim **??**, the authors present us with a method in which the locations of users can be estimated based on the WiFi signals that their devices register. The experiment is conducted cosndiering the data for a duration of 13 months. The user traces that have been used consist of the trace data from the Darmouth College. The mobility traces are defined as the lists of access points that are associated to a user's devices at a given timestamp.

The mobility traces allowed the authors to extract the tracks (locations) of the users. They have explored three methods in which the location can be extracted from the data. The first approach presumed the calculation of the center (intersection of medians) of the traingled defined by the past three access point associations of the mobile device of the user. This approach has a downside

since the devices do not necessarly change the associations in a periodic manner. This lead to the second approach which consisted in considering a time window after which the associations needed to be updated in case new associations have appeared during that time. The thrid and last approach explored the use of Kalman filters **??**.

The validation the path extarctors the authors have compared the results with GPS data. This validation has proven that the type of the used device has at the moment a significant importance in how acqurate the results can be as it seems that some devices can be more aggressive in updating the associations with access points while others try to stay associated with the same access points as long as possible before switching to new ones. This leads to problems as different distances between users and access points considered by different devices and as such it affects the estimated paths. The best estimations have been given in this experiment by the approach that used the Kalman filters, however both the other two appraoches have provided fairly good estimations as well.

Another paper which explores the travel patterns from real data is the one written by T. S. Azevedo, R. L. Bezerra, C. A. V. Campos and L. F. M. de Moraes **??**. The authors propose another approach for analyzing the mobility of people. They take into consideration the following movement components: velocity, acceleration, direction angle change and the pause time and they are using the GPS data in order to estimate the locations of individuals. The experiment takes place in a park in Rio de Janeiro and is done based on the data received from around 120 volunteers. The results have shown that people seem to have in general smooth trajectories without abrut changes.

2.6 Entropy and predictability

One step further from understanding the way we travel from place to place is to predict our future locations based on a previous knowledge our our past patterns. There has been an extensive study done in this area of the scientific playground as well and the results which have emerged up until now are remarcable.

In the paper by C. Song, Z. Qu, N. Blumm and A. L. Barabasi **??**, the authors take up the challenge of studying how predictable people can be. They analyze the mobility patterns of mobile phone users and calculate the entropy of these users. The locations are defined by the telephone towers the users are encountering at hourly intervals and the trajectory of the user is given by the ordered sequence of these towers. The real entropy of each user i is calculated

as $\Sigma_{T_i' \subset T_i} P(T_i')log_2(P(T_i'))$, where $P(T_i')$ represents the probability of encountering a time-ordered subsequence T_i' in the sequence of hourly encountered telephone towers T_i.

The results for this particular study show that, for the considered users, the uncertainty of where they could be at a certain moment, based on the real entropy calculated for them would be very low as they would most probably be in one of two locations.

The authors also take a look into the maximum predictability which can be expected for a user. Their results show that, with the right algorithm, a user's future location can be predicted with between $80 - 93\%$ accuracy. This shows that we are less sponatenous than we might think and that our mobility patterns are, in most cases, rooted into a very well established routine.

There have been numerous other methods or experiments conducted in order to analyze or to forcast human mobility patterns. Some of these methods include the Markov chain models **??** **??**, the neural networks **??** or the Bayesian networks **??** as well as some that work with finite automaton **??**. Most of the studies support the idea that people's actions and travel behavior is indeed far from being random and thus the science world needs to dedicate further effort and time in order to use this knowledge in order to improve our quality of life and the world we live in.

Chapter 3

Data

1. why working with wifi data;
2. data structure and fields;
3. how much data was analized
4. how was the data prepared (noise elimination)

Chapter 4

Locations

1. previous studies on what does a location is considering wifi fingerprints
2. analazying what can determine a location fingerprint
    - Plot access points' presence over time considering their signal strength
    - Plot number of samples of each access point over time
    - Plot the average signal strength for various time windows for each access point identified for a user
    - Plot the running average signal strength for 2,5 and 10 min time windows for each access point identified for a user
    - Plot of access point presence in time bin (5 mins used for time bin) without considering its signal strength
3. identifying locations
    - Networks
    - Hidden Markov Models
    - Further improvements
4. matching locations locations
    - Percentage similarity
    - Keeping track of previous locations

- Creating fingerprints

5.

Chapter 5

Entropy and predictability

1. Calculating entropy for users
2. Calculating predictability for users
3. Observations

Chapter 6

Comparing results with GPS results

Chapter 7

Results and comments

Chapter 8

Conclusions

Appendix A

Appendix

Appendix ...

Bibliography