Proposal

Mina Mousavifar - sem311 - 11279515

Research Question

Nowadays, people use various tracking devices to help them to accomplish daily tasks, such as using GPS to find the fastest route to work. However, this history might be exploited to endanger users' privacy. People's location history can be analyzed to predict their future location. Marketing strategies can exploit this data, and manipulate people's decisions by providing personalized recommendations based on their predicted location. However, the success of these marketing strategies depends on the ability to predict users' future location.
In this project, the success of predicting user locations with *Machine Learning* techniques will be investigated which mainly focuses on modelling and analysis of human behaviour. This predictability highly relies on finding patterns in data such as daily or weekly routines and modelling them effectively. Consequently, the accuracy of the model will be whether a specific user's next location, is between the set of predicted future locations for this user or not. Further, the success of predicting user location will be explored by comparing it to a naive model that predicts a set of user future locations randomly.

Operationalization

Location trajectory plays a vital role in modelling users' future locations. Because we are interested in the next location that a user will visit in the near future. This can be considered as the destination of a trip. So, we define the destination as the cell that a user dwells, which is staying in the same cell for a number of duty cycles more than a certain threshold. Furthermore, the trip can be extracted by tracking when a user changes a grid cell and remains in several grid cells for less than a certain threshold and then stays in a grid cell more than this threshold. We represent the user movement trajectory by a string notation that shows the user grid sorted based on time.
Consequently, our modelling task would change to predict the next reoccurring grid daily or weekly(for more than a certain threshold) given the previous history, and current trajectory. We can examine several models for this purpose. Firstly, Deep Learning techniques such as *CNN*[1] or *RNN*[2] can be applied to this problem to take advantage of locality in the data, because user dwelling in a certain cell might be less important than travelling to another grid cell, for this specific problem. The string trajectory would be the input of our model with a filter with the same size of our dwelling time threshold and the predicted string would be the output. Secondly, we can use a Markov model to predict user future location by assuming another hypothesis that future locations depend only on the current location, not on history. Thirdly, we can extract top locations based on time of the day without considering specific users and predicting based on time. This analysis needs to extract a universal duty cycle for data so that a specific duty cycle points to a specific time for all of the users. Then we can analyze popular locations, based on time of the day or week as the range of specific duty cycles.
We can supplement our GPS data by adding other features such as battery, light, weather, etc and explore whether this data can improve our predictions or impact them adversely. Finally, if we succeed in predicting users' future locations, we would highlight the importance of introducing methods for preserving users' privacy while using location services.

[1] Convolutional Neural Network
[2] Recurrent Neural Network