Next location prediction

Introduction

As a result of using smartphones GPS sensors, the location based services has raised (google now, foursquare), these services support the end-user with better recommendations depending on the context (location alarm, recommend nearby places, weather updates related to the location), with the use of new features like predicting the next location of the user, which he/she expected to visit, or next place recommendations, a new level of applications can be offered in advertising new places or even taking a decision based on the expected activity the user is going to take (traffic jams avoidance, offering new restaurants and make hotel reservations).

For the research of human mobility behavior there are different types of datasets are available:

* first type, is the raw anonymous GPS datasets (i.e. longitude, latitude, timestamp) with a small sampling rates of 5 to 10 seconds which can be considered as continuous data (Geolife),
* Second type, is chick-ins data, which are the places a user may checked, with difference hours between two check-ins, so it is considered discrete data.
* Third type, is the CDR files using the mobile service providers, this type also considered sparse data, because the mobile user can get disconnected from the mobile network due to low coverage,
* Fourth type is rich GPS dataset with location coordinates and the support of other types of data such as SMS, call duration, Wi-Fi, and the running applications.

Location data can be used in a large segment of applications like context aware applications (google now) which can recommend services based on the user location (location alarms, recommend places, location temperature), traffic recognition applications that allow users avoid congestions and recommend better routes, also recommender systems for advertising (ex. A restaurant recommender can promote a list of restaurants based on spatial and temporal features of the user, it can then make a reservation while a restaurant still has spaces, also location advertising application (ex. Yellow Pages) which can promote local businesses based on user current and next location.

Main research points: location related applications using raw dataset (Geolife) are mainly depend on:

1. Identifying the points of interest,
2. Identifying place and route popularity (popular places, travel sequence),
3. Identifying home and work location, 4-Identifying semantic places,
4. Identifying user demographics (gender, age, education) (Zhong 2015),
5. Predicting next location,
6. Recommender system to propose new locations,
7. Decision taking applications based location.

The methods for analyzing location histories can be classiﬁed according to the manner by which data are modeled, into three general distinct approaches, namely:

1. State-space models,

State-space models attempt to capture the variation in spatial sequences through sequence models such as generative Hidden Markov Models (HMMs) [17], discriminative Conditional Random Fields (CRFs) [21, 19], or extensions of these two well-known approaches

1. Data mining techniques,

Explore frequent patterns and association rules, by deﬁning a trajectory as an ordered sequence of time-stamped locations, and using sequence analysis methods such as modiﬁed versions of the Apriori algorithm

1. Template matching techniques.

Compare extracted features to pre-stored patterns or templates, using similarity metrics speciﬁc for sequential or time-series data.

Papers

1. (Wei Huang, 2015), using Geolife dataset they solved the problem of identifying stay points using DBSCAN with distance threshold of 200 meters, they considered the time threshold of 20 minutes, overcoming the problem of overlapping path by exclude the points within the same spatial region if they exceed the temporal threshold.

They enhanced the human movement prediction accuracy by solving the activity change problem, which is the problem of changing home, work or a routine activity, they solved this problem by calculating the similarity between the changed activities and if a similarity measure passes a threshold the activities considered as one.

They used Markov chain to predict the user’s next location, they solved the problem of changing activity to improve the prediction coefficient from 0.295 to 0.762 for high changing activity users, and from 0.965 to 0.971 for low changing activity users.

1. (Jie Yang, 2014), using Geolife they solved the problem of identifying stay points problem by using a DBSCAN variant algorithm which they proposed using the distance and speed as score for relation between points, if this score exceeded a threshold it would means the points are related to the same cluster.

They showed that the use of their clustering technique is more powerful than fixed time and space algorithms and extract more stay points, also the use of Variable Order Markov Model enhanced the prediction precision and solved two issues: zero frequency problem and high space complexity.

1. Ashbrook and Starner (2002), Using GPS logger of one user in a period of four months, they exploit these data to predict the next location of their user, they used k-means to determine the points of interest in the user’s trajectories.

Then they used Markov model to predict the next location, as they claimed, the prediction accuracy gets better with higher order Markov models, they used second order model. They didn’t show their results but they compared between using first and second order Markov model.

1. (Wesley Mathew, 2012)

Compute the set of sequences corresponding to all possible next places to be visited in order to group temporal sequences according to three clusters with one Hidden Markov model for every cluster. use the forward algorithm to compute the probability of all such sequences, and Return the next place corresponding to the sequence with the highest probability.

They cluster the historical locations based on temporal characteristics into three clusters for weekdays and weekends, and they attached a hidden Markov model for every cluster to find the best state transitions and probabilities, and finally they used Baum-Welch algorithm in order to find the most probable next location

The results of their work shows a prediction accuracy of 13:85%, when considering regions of 1280 squared meters.

1. (Josh Jia-Ching Ying, 2011), Used two modules and using the feature of location semantic , the two modules are: offline module to extract the locations of the and it’s semantics of the user trajectories and store it in a tree structure , this tree considered as the user’s trajectory with only stop points and their semantics, and the online module apply a similarity measure for the input trajectory and match it with the tree structured history and get the highest match, with the matched trajectory considered the one with next location.

References

1. Ashbrook, D. T. (2002). Learning significant locations and predicting user movement with GPS. *Wearable Computers* (pp. 101 - 108). IEEE.
2. Jie Yang, J. X. (2014). Predicting next location using a variable order Markov model. *SIGSPATIAL International Workshop on GeoStreaming.* ACM .
3. Josh Jia-Ching Ying, W.-C. L.-C. (2011). Semantic Trajectory Mining for Location Prediction . *SIGSPATIAL.* Chicago: ACM .
4. Wei Huang, S. L. ( 2015). Predicting human mobility with activity changes. *International Journal of Geographical Information Science*, 1569-1587.
5. Wesley Mathew, R. R. (2012). Predicting Future Locations with Hidden Markov Models. *UbiComp .* ACM.