Novelty seeking?!!

Abstract:

The enormous use of mobile applications and the wide spread of use location based services, as Gowalla, Foursquare, google maps, make it a must to discover these extracted data, which is the check-ins of the users, in order to maintain a better service and offer new services, as predicting the new places to be visited, predicting the traffic jammed areas, so we conducted this experiment to classify the check-ins to be visited either Novel or regular points of interest (POI), by extracting the main features of the current check-ins, using the features used by Defu et al.[1] and others, we can predict the novelty or the next Point of Interest (POI) up to 82% overall accuracy, thanks to the added features and the using of gradient boosting framework (XGBoost).

Introduction:

The emerging use of mobile technology with location features, has been widely occurred and location services have attracted millions of users, such as Gowalla, Foursquare, and adding it to other common social platforms like Facebook and twitter, these companies needed to enhance their services with applied analysis over these enriched big data, the researchers used such datasets to solve punch of problems, like analyzing human mobility patterns, travel patterns, social ties and its influence to POI, traffic predictions, and next check-in prediction and its enhancement to advertising and POI recommendations, as we go through the analysis of we found the novelty prediction is one more type of analysis, and it can play an important role in advertising and POI recommendations, it can be applied if a user is expected to visit a novel POI the system should be able to list a number of recommendations based on the users friends or the user’s historical interests, but this is the limit of our analysis, and in future work adding the POI semantic predictions would enhance the results and the application more importantly, but this feature is out of scope as the used dataset (Gowalla) has no semantics of the check-ins, for these experiments we focused on extracting features from the data to be able to predict the novelty of the next POI, using the temporal, spatial, and historical features, then we applied a classifier to classify novelty or regularity of the next POI based on these features, the problem turned to be a binary classification problem, this algorithm is a gradient boosting framework, which has a good fame at Kaggle’s competitions, called extreme gradient boosting (XGBoost), it uses decision trees, to solve these problems with high accuracy and speed. The next sections are ordered as related work, model features, proposed model, results and conclusion.

Related work:

Location based predictions is becoming at the focus of many research projects, as it has great impact on advertising, traffic congestion control and much more fields, these analyses consist mainly of three types. First, predictions using GPS trajectories for the places where the user spent more time, figured out by checking the time GPS points are close to each other for longer times. Then applying sequencing models to predict the next place to be visited, Ashbrook [2002], applied Markov model to get the next location, based on the frequency of the locations found at the users history, Gambs et al. (2012) used markov chains to figure out the next POI using Geolife dataset, also Mathew et al. (2012) proposed a hidden Markov model for prediction based on the observed time intervals to predict the next locations, these models at the level of single user, can’t predict new POI, furthermore, at the level of a group of users, the model can predict other private POI like home or work of one user to another user with the frequency of the first user home. Second, predictions using location-based social networks (Foursquare, Gowalla), check-ins to predict the next check-in the user expected to check-in which the user can check-in at the place where he/she visited, these datasets are sparse, as the user not always using the application with all the visited places, commonly the user check-in at the important POI, so the predictions can have much more error margin, these predictions as well as the GPS based datasets, researches often use Markov models to tackle these type of analysis, as the data can be processed to be sequences of states the user moving between them, Y.-J. Kim (2013), [Sung-Bae Cho](http://www.sciencedirect.com/science/article/pii/S092523121500569X) (2016), these techniques are used for user specific predictions, otherwise when researches need to solve the problem for all users often then use collaborative recommendation models, Zheng et al. (2010), Ye et al. (2011), these models can recommend novel places for the users based on the social connections between the users and also the spatial and temporal features, but these models failed to personalize the analysis, so they can fail to the problem of predicting private POI of one user to another (home, work), another way to solve the problem can be called a hybrid analysis, which separates the predictions for the novel places to be solved distinguished from predicting the user’s regular places, Yingzi Wang (2015)

#----------------------------------------------------------------

Location-based social networks has been recently becoming hot topics. It has encountered new opportunities and challenges because users’ mobility data continues to accumulate over time and more information comes in companion. One of them is the introduction of social networks so that users’ mobility may be directly or indirectly influenced by their friends. Therefore, there is much work on mobility modeling with the presence of social relationship. For example, Cho et al. [2011] proposed a periodical and social based model to predict the next location and concluded a small but significant effect of social relationship. Gao et al. [2012a; 2013] first built a Hierarchical Pitman-Yor (HPY) process model for each user to capture the long-range dependence among locations at the same time of producing the power law distribution of check-in frequency, which resembles the observations in the checkin behavior [Cheng et al. 2011; Gao et al. 2012a]. They also took social relationship into account to build a hybrid model so as to integrate the prediction of users’ self with friends, and also discovered a small impact of social relationship on location prediction. Sadilek et al. [2012] proposed a Dynamic Bayesian Network model to predict users’ future locations based on their friends with the presence of temporal information. Noulas et al. [2012a] and Chang and Sun [2011] built a prediction model using feature engineering and took into consideration plenty of features, including location popularity, users’ self and friends’ preference, topics and categories of locations and so on. Nevertheless, the major discovery from their experiments is still the importance of users’ preference. This is because users often return to previously visited locations, in particular after a certain period of the services usage [Cheng et al. 2011]. Although location prediction on LBSNs has exploited the influence of friends, which can come from human collective behaviours, e.g., having dinner with friends, or the word-of-mouth recommendation from them, it doesn’t fully explore the collaborative social knowledge. For example, at least it should include the knowledge from users sharing similar mobility patterns. Therefore, the proposed CEPR model is different from these existing approaches. It not only tries to fully capture collaborative social knowledge based on recommendation techniques but also makes better use of the individual power of regularity and recommendation based on Exploration Prediction so that it prevents regularity (individual preference) from always playing a dominating part in location prediction. To fully exploit the collaborative social knowledge, many researchers take out location exploration history and concentrate on the prediction on them. Since this task cannot be finished by traditional prediction algorithms, it usually resorts to recommendation techniques. In [Ye et al. 2011; Gao et al. 2012b; Noulas et al. 2012b], the authors employed collaborative filtering models, leveraging the similarity between users on mobility patterns and social relationship, for POI recommendations. Yang et al. [2013] and Liu and Xiong [2013] enhanced POI recommendation with textual information, such as tips and categories, of POIs. In addition to leveraging the extra information for alleviating the data scarcity problem, Zhang et al. [2013] proposed localized matrix factorization for recommendation based on matrix block diagonal forms. And of note is [Ye et al. 2011], which exploited geographical influence for location recommendation by assuming the power law distribution of distance between pairs of locations. Hence, the POIs nearer to users’ previous check-ins are ranked higher. However, according.

Proposed Features:

Algorithm:

Future work:

Acknowledge:

Conclusion:

References:

Mathew, W., Raposo, R., and Martins, B., 2012. Predicting future locations with hidden markov models. In: Proceedings of the 2012 ACM conference on ubiquitous computing, 5–8 September, Pittsburgh. New York: ACM, 911–918.

S. Gambs, M.O. Killijian, M.N. Prado Cortez. Next Place Prediction using Mobility Markov Chains. Eurosys' [MPM](http://www.cambridgeplus.net/MPM12/) 2012.

Y.-J. Kim and S.-B. Cho. A HMM-based location prediction framework with location recognizer combining k-nearest neighbor and multiple decision trees. In Hybrid Artificial Intelligent Systems, pages 618–628. Springer, 2013.

Defu Lian, Xing Xie, Vincent W. Zheng, Nicholas Jing Yuan, Fuzheng Zhang, Enhong Chen, CEPR: A Collaborative Exploration and Periodically Returning Model for Location Prediction, ACM Transactions on Intelligent Systems and Technology, 2015.

Yingzi Wang, Nicholas Jing Yuan. Regularity and Conformity: Location Prediction Using Heterogeneous Mobility Data, 2015.