An Analysis of Location Prediction Models

S. S. Daodu
Department of Computer Science
University of Benin, Benin City

E. Akinola
Department of Computer Science
Federal University of Technology, Akure

ABSTRACT

Location Prediction is an estimate of a location in which a user will be at a particular place at a particular time within a certain probability. Location Prediction has gained prominence over the past decade which is due to improved technology in mobile communication. Classification of mobile users can be regular or random which can be used to ascertain the pattern of the user over a period of time which also helps in planning the movement of the user.

This paper places emphasizes on the relevance of location prediction models in mobile users. A review of various location prediction model is carried out showing the effectiveness of each model, limitations, and also future work of some research works. Although this article does not give an exhaustive survey of all techniques and applications but it gives a description of several types of algorithms and models used for location prediction.

Keywords

Location Prediction, Semantic, Markov Chain

1. INTRODUCTION

Understanding human mobility has been a long standing subject in academic research due to the multitude of potential applications. Mobility prediction is one of the most essential issues that needs to be explored for mobility management systems. Predicting users next location or place also allows user to anticipate their future movement which enables adequate planning for the day and also provides additional time to be ready for that movement and can be used in location prediction models hence this research. Recently, the researchers on individuals' movement behavior not only concerns the mobility regularity but also explores how to use individual's historical context usage to predict the future location. Many wireless applications are deployed and available to customers via their mobile phones. Variety of these applications and services are based on the determination of the current or future location to get the best usage and results (Bahl et al. 2000).

At present, the most detailed information on human mobility across a big city is collected by mobile devices. Mobile devices recorded the closest mobile tower every time the user uses his or her phone. Location Prediction is also used in allocation of resources and advert deal placements. With effective prediction of location, it may also be possible to answer the queries that refer to the future positions of users. Location update should be performed whenever a mobile user arrives at a location. For effectively predicting the user's future requests, user access patterns are mined from the call logs of users previous requests and then these pattern are used for prefetching. Mobility management in Mobile Computing environments covers the methods for storing and updating the location information of mobile users who are served by the system. Mobility prediction can be defined as the prediction of a mobile user's next movement where the mobile user is travelling between the cells

of a Global System for Mobile communication network. The predicted movement can then be used to increase the efficiency of location prediction. By using the predicted movement, the system can effectively allocate resources to the most probable cells instead of blindly allocating excessive resources in the cell neighborhood of a mobile user. Effective Allocation of resources to mobile user improve resource utilization and reduces latency. Broadcast program generation can also benefit from predicted mobility patterns since the data items can be broadcasts to the cell where the users are moving (Saygin and Ulosoy, 2002)

In time past, there have been various methods used by different researchers for location prediction with each method having a different accuracy level which also gives room for more future work in this field of study. This research entails a survey of review on different location prediction models.

Lee et al., (2005) carried out a research on "Developing a User Location for Ubiquitous Computing". The motivation of this research was to address the problem of some location prediction problem which did not take into consideration the topological and geometry information of a place during prediction. The objective of the research was based on the development of a Computer-aided architectural design (CAAD) based data structure called "Semantic Location Model". The methodology involves that the Semantic Location Model embodies geometrical and topological information which describes the relationship between inhabitants and space also making the description of a physical situation exact and rich. The methodology involves data mining been implemented to extract the inhabitant's location pattern generated day by day. As a result, the self-learning system will be able to semantically predict the inhabitant's location in advance. Also a simulator was developed for the model. The limitation in this research was that the model was not used on a real life dataset and also the model was not applied to various applications in details.

Matthew et al. (2012) presented a work on "Predicting Future Locations with Hidden Markov Models (HMM). The objective of this work was to present a hybrid method for predicting human mobility on the basis of Hidden Markov Models (HMMs). This motivation of this work was gotten from the need to address the development and evaluation of generative models that, by capturing the sequential relations between places visited in a given period by particular individuals, it can support the analysis and inference of statistical patterns for predicting future locations to be visited. The methodology used involved clustering of location histories according to their characteristics and each cluster is then used to train different Hidden Markov Model. The initial clustering of sequence is based on the temporal period associated to the last place visited in the sequence in order to group sequences accordingly. A maximum length of 25 different locations visited and a minimum length of 10 locations were kept and this method gives a prediction accuracy of 13.85% when considered over a region of 1280 square meters. This work was limited due to the

fact that it was not experimented with discriminative models which will address related sequence analysis methods, such a as the classification of human location histories according to semantic categories.

Gambs et al. (2012), presented a research work on "Next Place Prediction using mobility Markov Chains. The motivation of this research was to design a novel algorithm which was used for predicting the next location and for incorporating the previous locations visited before reaching the current state. The objective of this research also addresses the issue of predicting the next location of an individual based on the observations of his mobility over some period of time and the recent locations that he has visited. The methodology employed in this research was an extension of a mobility model called Mobility Markov Chain (MMC) in order to incorporate the number (n) of previous visited locations which was called n-MMC. The limitation encountered in this research was the inability to introduce more explicitly the notion of time in the constructed mobility markov chain which reduced the prediction accuracy of the model. The research birthed an algorithm for next place prediction based on a mobility model of an individual called n-MMC that keeps track of the n previous locations visited.

Gomes et al. (2013), presented a research titled "Where will you go? Mobile Data Mining for Next Place Prediction. The motivation of this work was gotten from the need to enable greater personalization and privacy while bringing the whole learning process on board to the mobile device. The objective of this research was to create an algorithm called NextLocation which learns an anytime classification model and incorporates past data to predict the next place in an incremental manner. The methodology used in this research involved collection of raw data for each user location is using Global Positioning System and Wireless Fidelity. This is then transformed into a semantic place which captures most of the mobility/location based information without including the actual geographic coordinates/access points. Moreover, information from social networking services that support location such as Foursquare, Google Latitude already allows the user to 'check in'. These services already include automatic location detection which can be leveraged to create or enrich the temporal series of semantic locations that is required for next place prediction. The limitation in this research was the inability to develop an online algorithm particularly designed for next place prediction the implementation.

Noulas et al.(2012) presented a research "Mining User Mobility Features For Next Place Prediction in Location Based Services". The motivation of this work is gotten from the need to predict the venue a user will visit next by exploring the predictive power offered by different facets of a user's behavior. The objective of this work was to exploit information on transitions between types of places, mobility flows between venues and spatio-temporal characteristics of user check-in patterns. The methodology employed in this research was the use of foursquare mobile application to collect data. This mobile application also allows users to check in venues via their smartphone, geolocate themselves and also broadcast their location to their friends. Having collected data needed, a set of prediction features that exploit different information is defined, next is to combine the predictive power of individual features in a supervised learning framework. By training two supervised regressors; a regularized linear model and M5 model trees, on past user movements, a supervised learning approach using single features in the prediction of future user movement is used. The limitation of this research was that check ins that happen 24hours from the previous ones were not used in the

next place prediction.

Herder et al. (2014), presented a research titled "Predicting User Locations and Trajections. The motivation of this work was gotten from the need of predicting locations that people will revisit as part of their routine patterns as they observed that human mobility patterns contain strong regularities. People typically spend most of their time at and between a small number of locations and most followed trajectories (e.g the daily commute) also serve as starting points for visits to several other locations that form the long tail of a person's whereabouts. The objective of this research was to predict user locations from data collected through GPS. The methodology involved analyzing, visualizing and discussing patterns found in a dataset of GPS trajectories. Furthermore, the authors compared and analyzed the performance of common prediction techniques that exploit the locations' popularity, regularity, distance and connections with other locations. The limitation in this work was that trajectories considered were only those on the user's current location.

Wen et al. (2014), presented a work on "Improving Location Prediction by Exploring Spatial Temporal Social Ties". The motivation of this work was to bring out the difference in movement patterns and social correlation between weekdays and weekends in predicting individual future location. The objective of this research was to make use of a Fallback Social Temporal Hierarchic Markov Model (FSTHM) to predict the future location. The methodology used involved the division of time into two different discrete time slice, week days and weekends. The state of original Markov is assigned as two different discrete states Secondly the distributions of the visit times at each state are captured to describe the individuals movements habit and a Bayes rule is used to predict the next state according to the temporal and Spatial information. On weekdays, an individual's life is relatively regular while on weekends he usually participates in some social activity. The spatial temporal method could be divided into three (3) categories:-

   I.   Temporal Based Location Prediction Methods: - This kind of location prediction method considers individual's movement to be a series of isolated physical position with a given time stamp. The length of time interval affects the predictive performance.

   II.   Spatial Based Location Prediction Method: - This sort of method mainly considers the spatial mobility regularity in individual movements and neglects the specific time of the places they visited.

   III.   Spatial Temporal Location Prediction Method:- Spatial Based Location Prediction methods define user's movement behaviour as sequential patterns or association rules that frequently appear among location sequences. The temporal information could help to identify the best matches from the patterns that have the highest similarity. The FSTHM model integrates the historical and social ties to obtain the final prediction result.

The limitation of this work was that the influence of the real time social ties was not put into consideration.

Keles *et al. (*2014), presented a research titled "Location Prediction of Mobile Phone Users using Apriori-Based Sequence Mining with Multiple Support Thresholds". The motivation was gotten from the need to provide better services and recommendations for mobile phone operators and smart city administration by using historical movement patterns for

current location prediction of a person. The objective of this work was to present a new algorithm known as an Apriori – based sequence mining algorithm for next location prediction of mobile phone users. The methodology employed in this work was that the authors proposed a modified version of Apriori based sequence mining algorithm that works with multiple minimum support values instead of a global one. The authors considered a predefined number of previous steps in order to predict the next one. They made use of a recursive hash tree based algorithm namely Apriori-based Sequence Mining Algorithm with Multiple Support Thresholds (ASMAMS). This algorithm constructs level based models i.e hash trees whose nodes contain corresponding base station id and frequency count of the sequence corresponding to the path up to this node. ASMAMS algorithm is divided into two (2) phases: Tree Construction and Pruning. In the construction phase, the data is read sequentially and new level nodes are added to the corresponding tree nodes. In the pruning phase, constructed model and the corresponding minimum support values are taken as parameters. In the prediction Phase, the hash tree constructed by ASMAMS to predict user next location. The prediction algorithm takes a sequence with length of level count – 1 as inputs and returns a list of predicted locations. The algorithm firstly checks if the constructed tree contains children nodes of the corresponding nodes (at the level count-1) constitutes the prediction set. If it does not contain the given sequence, it returns an empty list indicating the model cannot predict a location for the given sequence. The limitation of this work was that the model did not have a region based hierarchy that can increase the prediction accuracy

Han *et al.,* (2014), presented a research titled "Text Based Twitter user Geolocation Prediction". The motivation of this research was to present an integrated geolocation prediction framework and investigate what factors impact on prediction accuracy. The objective of this work was to propose a text based method which uses words for geolocation prediction and intentionally excludes Twitter specific activities such as hash tags and also to reliably predict the location of users for whom the location is not known that is where there is only non geotagged data. In this research, exploits of tweets and profile information of a given user was carried out which was used to infer their primary city level location. The authors were also able to reveal from their experiments that geolocation prediction is much easier for languages that are more diverse in usage e.g English. They stated that a user's location is predicted at the regional level by adopting the location centroid for geotagged tweets from that region. The limitation of this research came from the fact that due to the uneven geographical distribution of languages in tweets, users of geographically-diverse languages (e.g., English and Spanish) are much harder to geolocate than users of geographically focused languages.

Zhang *et al.,* (2009) presented a research titled "Location Prediction Model Based on Bayesian Network Theory". The motivation of this research is to design a location prediction model considering multiple predictive factors to improve the prediction success ratio and improve the efficiency of the model. The objective was to propose a location prediction model based on Bayesian Theory for active mobility management. The methodology takes three (3) kinds of predictive factors into account and this include: - Topology Information, Road Topology Information and Movement Information. This model solves multifactor location prediction problem. The relative predictive factors are first coded in the Bayesian Network node and the location prediction results can be calculated based on cell topology information and coded in

the Bayesian Network Model (BNM). In the nodes description, the movement information and statistical information are coded in the node table. All of the predictive factors are coded in the Bayesian Network. In the BNM, each network stores a probability of predictive factors distribution. Predictive factors of angular velocity and acceleration factors are related to the road environment and node characteristics. Users can change their acceleration state according to the actual condition. These two kinds of predictive factors have no distribution characteristics in the state transition process. The limitation of this work was the inability to implement the model using real life datasets

Qiao *et al.* (2015), carried out a research work on "User location prediction with energy efficiency model in the Long Term-Evolution network". The motivation of the research was to predict users' next location or place using an energy efficient technique which also provides the best tradeoff between prediction accuracy and resource consumption. The research objective was mainly to predict a users' location using two domain-independent algorithms, Markov Model and Lempel Ziv algorithm. The methodology involved a comparative analysis been done between the two algorithms after it has been applied to various categories of users and the algorithm with highest level of accuracy is gotten. For the Markov Family, The order-k Markov predictor is independent of time, and it assumes that the current location depends only on the previous k movements. The algorithms of Lempel Ziv family are often used for text compression; they are able to make real time predictions and do not need many resources. The limitation in this research was the inability to effectively group users based on their mobility pattern and spatio-temporal characteristics.

Zhang *et al. (*2016), presented a research on "Location Prediction Based on Transition Probability Matrices constructing from Sequential rules for spatial Temporal K Anonymity Dataset". The motivation of this research came from the need to develop a mainstream privacy protection method for Location based Services that is simple to use and that can also be used for various applications. The objective was to propose two location prediction methods: rough prediction and accurate prediction. The former achieves the probabilities of arriving at target locations along simple paths which include current locations, target locations and transition queries for multiple users while the latter method calculates transition probabilities for detailed paths with n steps from current locations to target locations. The methodology used in this research involved the use of constructed transition probability matrices from mined single step sequential rules and normalizing the transition probabilities in the transition matrices. They also regard a mobility model for an LBS requester as a stationary stochastic process and compute the n-step transition probability matrices by raising the normalized transition probability matrices to the power n. This work was limited by the methods not been able to help data miners and domain experts ensure that privacy-sensitive knowledge is released or accessible only to trusted parties.

Liu *et al. (*2016), presented a research on Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts. The objective of this research was to extend Recurrent Neural Network (RNN) and propose a novel method called Spatial Temporal Recurrent Neural Networks (ST-RNN) which can model local temporal and spatial contexts in each layer with time-specific transition matrices for different time intervals and distance-specific transition matrices for different geographical distances. The motivation of the research came from the fact that although so many methods have achieved

satisfactory results in some applications, they are unable to handle continuous geographical distances between locations and time intervals between nearby behaviors in modeling sequential data. The methodology used in this research was that geographical distance and time between locations was modeled into the conventional RNN. Considering distance information is an essential factor for location prediction, it was necessary to involve it into the model. Similar to time-specific transition matrices, we incorporate distance-specific transition matrices for different geographical distances between locations. Distance-specific transition matrices capture geographical properties that affect human behavior. The limitation of this research was that the internal representation of space and time features were not extracted automatically, it relied on handcraft representations.

2. CONCLUSION

This research has analyzed various location prediction models, their objectives and also stated their various limitations. This can aid users in choosing the location predicting model that best suits their need their location. In future this research hopes to expand its scope on more location prediction model.

3. REFERENCES

[1] Bahl, P., Padmanabhan, V. N., and Balachandran, A. (2000). Enhancements to the RADAR user location and tracking system. *Microsoft Research*, *2*(MSR-TR-2000-12), 775-784.

[2] Gambs, S., Killijian, M. O., and Prado Cortez, M. N. (2012,). Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility* (p. 3). ACM.

[3] Gomes, J. B., Phua, C., & Krishnaswamy, S. (2013). Where will you go? mobile data mining for next place prediction. In *International Conference on Data Warehousing and Knowledge Discovery* (pp. 146-158). Springer, Berlin, Heidelberg.

[4] Han, B., Cook, P., & Baldwin, T. (2014). Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research*, *49*, 451-500.

[5] Herder, E., Siehndel, P., and Kawase, R. (2014). Predicting user locations and trajectories. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 86-97). Springer, Cham.

[6] Keles, I., Ozer, M., Toroslu, I. H., and Karagoz, P. (2014). Location prediction of mobile phone users using apriori-based sequence mining with multiple support thresholds. In *International Workshop on New Frontiers in Mining Complex Patterns* (pp. 179-193). Springer, Cham.

[7] Liu, Q., Wu, S., Wang, L., and Tan, T. (2016). Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts. In *AAAI* (pp. 194-200).

[8] Mathew, W., Raposo, R., & Martins, B. (2012, September). Predicting future locations with hidden Markov models. In *Proceedings of the 2012 ACM conference on ubiquitous computing* (pp. 911-918).

[9] Noulas, A., Scellato, S., Lathia, N., and Mascolo, C. (2012). Mining user mobility features for next place prediction in location-based services. In *Data mining (ICDM), 2012 IEEE 12th international conference on* (pp. 1038-1043). IEEE.

[10] Lee, Y., Choi, J. W., and Lertlakkhanakul, J. (2005). Developing a user location prediction model for ubiquitous computing. In *Proceedings of CAAD Futures* (pp. 215-224).

[11] Qiao, Y., Yang, J., He, H., Cheng, Y., and Ma, Z. (2015). User location prediction with energy efficiency model in the Long Term-Evolution network. *International Journal of Communication Systems*, *29*(14), 2169-2187.

[12] Wen, L., Shi-Xiong, X., Feng, L., and Lei, Z. (2014). Improving location prediction by exploring spatial-temporal-social ties. *Mathematical Problems in Engineering*, *2014*.

[13] Zhang, Y., Hu, J., Dong, J., Yuan, Y., Zhou, J., and Shi, J. (2009). Location prediction model based on Bayesian network theory. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE* (pp. 1-6). IEEE.

[14] Zhang, H., Chen, Z., Liu, Z., Zhu, Y., and Wu, C. (2016). Location Prediction Based on Transition Probability Matrices Constructing from Sequential Rules for Spatial-Temporal K-Anonymity Dataset. *PloS one*, *11*(8), e0160629

[15] Saygin, Y., & Ulusoy, Ö. (2002). Exploiting data mining techniques for broadcasting data in mobile computing environments. *IEEE Transactions on Knowledge and Data Engineering*, *14*(6), 1387-1399.