Personalized Movement Model and Trajectory Prediction Method based on Urban Multi-source Data

Qinghua Chen

College of Artificial Intelligence, Wenzhou Polytechnic, Wenzhou, Zhejiang 325035, China.

47449970@qq.com

Abstract

With the continuous development of mobile sensing technology and the wide application of positioning technology (GPS and Beidou system, GSM positioning, wireless indoor positioning, etc.), the technologies of Internet and Internet of Things have penetrated into all aspects of life. We have gathered more and more trajectory data. The trajectory data provides a basis for understanding the activity characteristics of urban people's life and work, evaluating urban governance measures, and providing decision support for optimizing urban layout and improving traffic facilities. In this paper, we study the personalized movement model and trajectory prediction method based on urban multi-source data. The problem of data sparsity will be solved by using the user's own historical trajectory and group trajectory data in a balanced manner. Besides, we integrate the urban multi-source data to improve the granularity and accuracy of trajectory prediction.

Keywords

Multi Source Data; Movement Modeling; Trajectory Prediction; Long and Short Term Memory Cycle Neural Network (LSTM).

1. Introduction

Location based services (LBS) are imperceptibly changing people's lives, such as road navigation, location query, real-time traffic query and so on. Trajectory prediction is an important modular for location-based services. According to accurate trajectory prediction, it can provide users with many services, such as urban traffic scheduling and congestion avoindance; applications at smart home which adjusts the environment (heating, cooling, ventilation, etc.) according to the specific time when users go home; accurate advertisement for specific user groups; costs reduction; revenue improvement with accurate advertisement, and so on.

According to the granularity of prediction, trajectory prediction can be divided into destination prediction, long-time trajectory prediction, short-time trajectory prediction and so on. Destination prediction is to predict where users will go in the future according to several places with historical data. The granularity of this kind of prediction problem is relatively coarse. Hence, it is relatively easy. Short time trajectory prediction is to predict the position of objects in real time according to the current speed and direction of cars or pedestrians, combined with the width and direction of the roads. This fine-grained trajectory prediction is mainly applied to object anti-collision, but this method is not suitable for long-time trajectory prediction.

In the long-time trajectory prediction, the trajectory patterns are mainly determined by users' historical trajectory records (for example, a user often goes to a supermarket after work, and then goes home). In the process of predicting future trajectory, finding the user's travel pattern, estimating the location of the next few moments, or building a probability model (Markov chain Hidden Markov chain, etc.) is applied to describe the user's travel status and predict the location distribution probability in the next few moments according to the user's location status in the previous moments. The method based on user travel mode needs to mine the mode at first, and then find the matching mode during the process of prediction. Those methods are time-consuming. In the probabilistic Markov chain, the order of Markov chain is also difficult to determine. In addition, the traditional trajectory prediction methods will encounter the problem of data sparsity in personalized trajectory prediction. The main reason is that the active area of a single user is limited. For example, in a city, the number of roads a user often passes through is very small. In the process of training the model according to the user's own trajectory data and making personalized trajectory prediction, if the user appears in a strange environment, the prediction model will fail.

In order to solve the shortcomings of existing trajectory prediction methods, this paper will discuss the personalized movement model based on Long and Short Term Memory cycle neural network (LSTM), and propose a trajectory prediction method based on the LSTM. In the process of personalized movement modeling, we will encounter the problem of data sparsity. In order to solve the sparsity problem, a separate module will be trained by using the trajectories of all users, and the module will be used as an input of each user's separate model. When making trajectory prediction in an unfamiliar area of the user, the module can predict the user's trajectory according to the trajectory mode of the other users. In addition, traditional trajectory prediction methods generally only consider historical trajectory data. However, there are many relevant data that can be obtained in cities, such as POI (point of interest), air quality, weather, working days, etc. These factors will affect the travel mode of the users. Therefore, using urban multi-source data will help to improve the prediction accuracy. The model based on LSTM will be extended to integrate urban multi-source trajectory data so as to establish a more accurate personalized prediction model.

2. Research State and Future Development Tendency

A lot of work has been done for the analysis and application of trajectory data. We have conducted extensive and in-depth research on trajectory data and related technologies. The research status and dynamic analysis are as follows.

2.1 Research on trajectory pattern mining method

Pattern mining is to analyze the motion patterns of one or more moving objects. Possible motion patterns include gathering / group patterns [1-3], sequential patterns [4], periodic patterns [5,6], etc. If each track is regarded as a sequence, a sequential pattern is usually defined as a subsequence, so that at least one track contains this subsequence. Zheng et al. [2] studied the problem of mining sequential patterns of semantic trajectories. This method, called SPLITTER, is divided into two steps to find fine-grained sequential patterns. The SPLITTER first gathers the trajectories near the same geographical location to find the coarse-grained pattern, and then a top-down model is used to split the coarse-grained model into the required fine-grained model. Periodic pattern is another common pattern that is very important to understand the behavior of moving objects. Li et al. [5,6] considered the problem of mining periodic patterns, and used reference location points and probability models to mine periodic motion behavior respectively.

2.2 Research on annotation method of trajectory

How to segment a whole trajectory into multiple related parts and label them respectively is an important research topic. Many work has proposed various trajectory segmentation methods and semantic annotation methods. Literature [7] uses hidden Markov model to divide the equipment trajectory sequence into two states: stationary and moving according to the action type. Literature [8-10] propose and use different analysis methods to divide the trajectory data into migration part and resident part, so as to improve the effect and performance of labeling, searching and querying in each part of the trajectory.

It is proposed that by analyzing the mobile phone radio signals received by different signal towers, the user's moving trajectory can be marked as various states such as stationary, walking, driving and so on. Using the GPS trajectory data of taxis and the actual geographic information such as road network, we can first find the suspended trajectory points through the stop state detection algorithm, so as to divide the trajectory into multiple segments, and then use the probability Classification Model and semi hidden Markov model to model the local change characteristics and long-term driving characteristics of the track respectively, and finally labels them as idle, busy or stop. In some studies, people's daily movement trajectory is randomly divided into multiple segments, and their motion features such as speed, direction change rate and speed change are extracted respectively. The calculated motion features are input into the decision tree for training, and they are marked with driving, cycling, bus taking and walking according to the traffic mode.

2.3 Research on trajectory prediction method

According to the time scale of the prediction, trajectory prediction is divided into short-time prediction and long-time prediction. Short-time prediction is mainly based on the current position and speed of the object to establish a physical model to predict the movement of the object in a short time. Literature [11] establishes a vehicle motion model to predict the vehicle trajectory according to the current speed, direction and road width. Literature [12] uses LSTM neural network to model pedestrians in crowded space and predict personalized pedestrian trajectory. These short-term trajectory prediction are mainly applied in the field of automobile or robot anti-collision, which makes decisions by predicting the movement of other objects around to prevent collision.

With the popularity of mobile phones, there are more and more check-in data with geographical location provided by social networks. Many studies predict the trajectory of users based on these social network check-in data. In literature [13-15], trajectory prediction is carried out according to the mode of users' daily trajectory data with the social status of users in social networks.

The check-in data of social networks often contain some semantic information, which often helps to improve the accuracy of prediction. The accuracy of prediction can be improved by predicting the category of the place (such as cinema, hospital, etc.) that the user will visit next, and then predicting the specific location of the place. It can also be used to capture the landmarks passing through the track, extract the activities corresponding to the trajectory, and then predict the trajectory.

3. Trajectory prediction method based on urban multi-source data

3.1 Urban multi-source data acquisition and preprocessing

First, we have to collect urban multi-source data, including weather, air quality, POI, trajectory, etc. Weather conditions, air pressure, temperature, humidity, etc. Air quality includes PM2.5, PM10, etc. The missing air quality and weather conditions also need to be interpolated and normalized. The trajectory data is divided first. For example, if the time interval between two adjacent trajectory points is long, it needs to be divided into two independent trajectory. Besides, the GPS abnormal points of the trajectory are needs to be deleted. Then, each trajectory is mapped to the road network, and the GPS position is corrected according to the road network to eliminate the impact of GPS offset.

Data sources for trajectory include more and more government open data sets and multi-source data accessible through open channels, such as remote sensing data, check in data, traffic road condition data, POI data and WiFi distribution data. All the data forms urban multi-source data sets in our method. It includes obtaining trajectory data from vehicle GPS and smart phones, air quality data from air monitoring stations, weather data from the National Meteorological Administration, and POI data from map websites such as sky map. These data are stored in the corresponding system, and will be preprocessed. Besides, he trajectory data are semantically annotated, so that these data can be applied in trajectory prediction.

3.2 Trajectory prediction model based on urban multi-source data

Traditional trajectory prediction methods will encounter the problem of data sparsity in personalized trajectory prediction, because the area of a single user is limited. For example, in a city, the number of roads a user often passes through is very small. Hence, we explore the complex correlation mechanism between urban multi-source data and trajectory data, propose the fusion model of urban multi-source data and trajectory data, and establish the balance mechanism between personalization and data sparsity. The trajectory prediction model mainly includes two parts: General LSTM and individual LSTM. First, use the trajectory data of all users to train the general LSTM model, and take the trained model as a module of the individual LSTM. Then, for each user, we use their own trajectory data to train the individual LSTM model. It should be noted that the previously trained general LSTM does not need a gradient descent learning process.

3.3 Trajectory prediction method based on urban multi-source data

This paper studies the problem of trajectory prediction, design and implementation of a trajectory prediction method based on LSTM neural network, and provides trajectory prediction services upward. First, a general LSTM model can be trained using the trajectory data of all users. The input of the model includes two types of data. One is the data that changes rapidly with time, including trajectory points, the number of POIs around trajectory points, semantically labeled labels, etc; The other is the data that does not change rapidly with the increasement of trajectory points (weather, air quality, etc.).

Since the weather and air quality data are generally coarse-grained and will not change rapidly with the increase of trajectory points, they can be considered as static data. In the model, the influence of static data on all trajectory points is the same. For example, in fog and the other weather, the driving speed of cars is generally slower than usual. Therefore, in the model, these static data only act on each gate of LSTM.

In the model, we only input the value of one position in the sequence at each time. The general LSTM model can set the input number of trajectory points at one time. The more the number, the better the prediction effect of the model. Finally, the results of trajectory prediction can be used in a variety of applications, such as accurate advertisement, smart home, traffic scheduling and so on.

4. Conclusion

Trajectory data records the position sequence with time mark left by the objects (such as cars, people, etc.) in the process of moving. Trajectory big data has become an important kind of big data with stable sources and easy access. This paper focuses on long-time trajectory prediction. Traditional trajectory prediction methods encounter the problem of data sparsity in personalized trajectory prediction. In order to solve the sparsity problem, POIs, air quality, weather, working day and other relevant data that can be obtained in cities are selected. These multi-source data is helpful to improve the prediction accuracy. This paper extends the model based on LSTM, integrates urban multi-source trajectory data, and establishes a more accurate personalized trajectory prediction model.

Acknowledgements

This work was supported in part by the Wenzhou Polytechnic project under grant No. WZY2021010 and WZYBSZD202101, and the project of Wenzhou Science & Technology Bureau (G2020017).

References

[1] K. Zheng, Y. Zheng, N. J. Yuan, and S. Shang, “On discovery of gathering patterns from trajectories,” in IEEE ICDE, 2013, pp. 242-253.

[2] K. Zheng, Y. Zheng, N. J. Yuan, S. Shang, and X. Zhou, “Online Discovery of Gathering Patterns over Trajectories,” IEEE TKDE, vol. 26, no. 8, pp. 1974-1988, 2014.

[3] X. Li, V. Ceikute, C. S. Jensen, and K.-. Tan, Lee, “Effective Online Group Discovery in Trajectory Databases,” IEEE TKDE, vol. 25, no. 12, pp. 2752-2766, 2013.

[4] C. Zhang, J. Han, L. Shou, J. Lu, and T. F. L. Porta, “Splitter: Mining Fine-Grained Sequential Patterns in Semantic Trajectories,” PVLDB, vol. 7, no. 9, pp. 769-780, 2014.

[5] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye, “Mining periodic behaviors for moving objects,” in ACM KDD, 2010, pp. 1099-1108.

[6] Z. Li, J. Han, B. Ding, and R. Kays, “Mining periodic behaviors of object movements for animal and biological sustainability studies,” Data Mining and Knowledge Discovery, vol. 24, no. 2, pp. 355-386, 2012.

[7] J. Krumm, and E. Horvitz, “LOCADIO: Inferring Motion and Location from Wi-Fi Signal Strengths,” in MobiQuitous, 2004, pp. 4-13.

[8] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vangenot, “A conceptual view on trajectories,” Data & knowledge Engineering, vol.65, no.1, pp.126-146, 2008.

[9] L. O. Alvares, V. Bogorny, B. Kuijpers, J. A. F. de Macedo, B. Moelans, and A. Vaisman, “A model for enriching trajectories with semantic geographical information,” in ACM GIS, 2007, pp. 22.

[10] C. Du Mouza, and P. Rigaux, “Mobility patterns,” GeoInformatica, vol.9, no.4, pp.297-319, 2005.

[11] A. Houenou, P. Bonnifait, V. Cherfaoui, and W. Yao, “Vehicle trajectory prediction based on motion model and maneuver recognition,” in 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013, pp. 4363-4369.

[12] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social LSTM: Human Trajectory Prediction in Crowded Spaces,” in CVPR, 2016, pp. 961--971.

[13] R. Assam, and T. Seidl, “Check-in Location Prediction Using Wavelets and Conditional Random Fields,” in IEEE ICDM, 2014, pp. 713-718.

[14] J. McGee, J. Caverlee, and Z. Cheng, “Location prediction in social media based on tie strength,” in ACM CIKM, 2013, pp. 459-468.

[15] N. Yang, X. Kong, F. Wang, and P. S. Yu, “When and Where: Predicting Human Movements Based on Social Spatial-Temporal Events,” in SDM, 2014, pp. 515-523.