Efficiently Predicting Taxi Future Location based on Square Error Estimation

Wubin Ma, Hongbin Huang*, Su Deng
Information system engineering key laboratory, National University of Defense Technology, Changsha, Hunan, China

*Corresponding author e-mail: hb_huang@nudt.edu.cn

Abstract. The future locations of taxis can reflect both human activities and urban dynamics in a city. Existing approaches to predict taxi future location are inefficient due to their high complexity. In this paper, we propose an efficient method the proposed method reduces the average computation time by 5.7%, while retaining almost the same prediction accuracy.

1. Introduction
Location-based services (LBS) such as Google Latitude and Loopt are increasingly popular in recent years. We are able to acquire large volume of trajectories of taxis using LBS equipment, and in return, provide the LBS users with intelligent services leveraging the knowledge extracted from the trajectory data.

Witnessing the proliferation of taxi trajectory data, we study the problem of efficiently predicting taxi future locations. Predicting taxi future location is of importance and usefulness in wide spectrum of applications, and Figure 1 illustrates a typical predicting system for LBS.

In this paper, we propose a prediction method based on the mean square error estimation.

2. Related work
Researchers hope to find knowledge from a large number of taxi trajectory data. Zheng et al. [1] used the collected taxi GPS data to mine the location and travel sequence of certain geospatial areas, and obtained the common travel patterns. Therefore, the user can find the surrounding location of interest. Zhang et al. [2] cloud service system, which analyzes the data to predict the traffic situation in a certain period of time in the future, thereby quickly analyzing the traffic conditions for the user. Provide recommended driving routes. Liang et al. [3] analyzed the data of moving trajectories and found that taxis tend to satisfy the exponential distribution, rather than the power law that we usually consider to be satisfied in human flow data. Black [4] tried to find out the potential defects of traffic through taxi GPS data analysis, and helped the traffic road design department to optimize the traffic road design. First, construct a traffic data population tree to detect abnormal conditions and analyze the causes of these abnormal conditions. The paper [5] also uses similar ideas to discover abnormal traffic behaviors in Beijing, and analyzes which factors may cause these abnormalities, so that traffic conditions can be estimated and improved in advance. [6] Focus on the discovery of traffic operation knowledge, and optimize the alternating time of traffic signals through data analysis. [8] and [9] also use some basic algorithms to predict the future location of the taxi and use real data to improve the predictions.
3. Model
In this section, we define urban taxis’ trajectory model and describe the linear mean square estimation theory.

Firstly, we define a sequence state of time and space, constructing the spatial and temporal model of taxis.

**Definition 1.** (Time series). The time series are several sequential time points in some period of time. It denotes as \( T = \{t_0, t_1, t_2, ..., t_n\} \), wherein \( t_i > t_j \) if \( i > j \).

**Definition 2.** (Limited grid spatial state). The physical world is divided to several limited rectangle grids, and there is no overlap between any two grids. The limited grid spatial state represents a rectangle grid of the physical word. The limited grid spatial states are denoted as \( S = \{s_1, s_2, ..., s_m\} \).

The taxis will locate at an accurate spatial state \( s_i \) in some time \( t_k \). Taxis move from one of the limited grid spatial state to another.

**Definition 3.** (Taxi objects). The taxis objects are denote as \( C = \{c_1, c_2, ..., c_p\} \), \( c_i \) present an object, which moves in the spatial state.

**Definition 4.** (GPS trajectory). An object’s trajectory \( T_{tr} : P_1 \rightarrow P_2 \rightarrow \cdots \rightarrow P_n \) is a sequence of time-stamped points, wherein \( P_i \) denotes location place of the taxi, \( P_i = (px, py, t) \), \( \forall k \in [0, n], (P_{k+1} - P_k) = \Delta t, \Delta t > 0 \), \( \Delta t \) indicate the distance between two neighbors location points.

**Definition 5.** (Taxi target prediction). Having \( n \) points of taxi’s trajectories \( T_{tr} : P(0) \rightarrow P(1) \rightarrow \cdots \rightarrow P(n) \), we try to predicted position \( p(n+1) \) in future moment \( t_{n+k} \). In general, the taxis’ target prediction is based on the history of the taxi trajectory information to predict position \( P(n+k) \) in the next \( k \) time intervals \( t_{n+k} \).

We use our model to predict the future location of taxi.

4. Experiments
This section provides the experimental results.

**Datasets.** The first data set is U.S. San Francisco taxi dataset which has a total of 424 moving urban taxis’ trajectory points in 25-day. The second is Beijing taxi data sets and Chengdu datasets [1, 7] in MSRA which collected a total of 1,000 taxis within a week of moving urban taxis’ trajectory points. In our experiment, we mainly validate prediction accuracy time complexity in the condition of different parameters’ values and different scale datasets.

For the San Francisco data sets, the grid spatial state is divided into 10 same rectangles. We take the first 22 days of the cabs’ movement points as historical trajectory, trying to predict future location state of cabs in three days. The value of the time interval is one hour, time-related system parameter \( \alpha \) take value of 0.5. And the dataset for Beijing, the grid state is also divided into 10 same rectangles, the first six days of cab movement points are taken as the historical trajectory. Select among the most likely top 20% predicted positions, giving the percentage of the correct next position. We predict location state of the cabs in the last day, the time interval also takes one hour. And then compare the estimation value of location state and the actual position of the cabs. For the day and night in different time periods, the prediction results are shown in Figure 1.
5. Summary and Outlook

In this paper, Location-based services (LBS) play a very important role in the future of smart cities. As a kind of indispensable modern transportation entity, cab is an important element in location-based services. This study was mainly targeted against a single cab objects, our future work will focus on predicting the change character of the number of urban cabs group.

Acknowledgments
This work was financially supported by Hunan Natural Science Foundation 2018JJ3619 fund.

References
[1] Zheng, Y., Zhang, L.Z., Xie, X. Ma, W.Y., Mining interesting locations and travel sequences from GPS trajectories, in WWW, 2009.
[2] Zhang D, He T, Zhang F, et al. Carpooling Service for Large-Scale Taxicab Networks[J]. ACM Transactions on Sensor Networks (TOSN), 2016, 12(3): 18.
[3] Liang, X., Zheng, X., Lv, W.F., Zhu, T. Xie, K. The scaling of human mobility by taxis is exponential, Physica A, 2011.
[4] Black, John. Urban transport planning: Theory and practice. Routledge, 2018.
[5] Xu M, Wang D, Li J. DESTPRE: a data-driven approach to destination prediction for taxi rides[C],Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2016: 729-739.
[6] Yao, Baozhen, et al. "Short-term traffic speed prediction for an urban corridor." Computer-Aided Civil and Infrastructure Engineering 32.2 (2017): 154-169.
[7] Shi, Kai, et al. "Detrended cross-correlation analysis of urban traffic congestion and NO2 concentrations in Chengdu." Transportation Research Part D: Transport and Environment 61 (2018): 165-173.
[8] Maniak, Tomasz, Rahat Iqbal, and Faiyaz Doctor. "Traffic Modelling, Visualisation and Prediction for Urban Mobility Management." Advances in Hybridization of Intelligent Methods. Springer, Cham, 2018. 57-70.
[9] Gunawardena, Janaka MA, et al. "Predicting Stormwater Quality Resulting from Traffic Generated Pollutants." Influence of Traffic and Land Use on Urban Stormwater Quality. Springer, Singapore, 2018. 55-69.