Trajectory Prediction Based on the Notion of Time and the Influence of Location of Historical Time Step

Fangxin Liu[1,4], Ming He[1,2,*], Yong Liu[1], Huan Zhou[3] and Qiuli Chen[1,2,4]
[1]College of Command Information System, PLA Science and Technology University, Nanjing, China
[2]The 61th Research Institute of PLA, Beijing, China
[3]Institute of Vocational Education, Tongji University, Shanghai, China
[4]Nanjing University of Information Science and Technology, Nanjing, China
[*]Corresponding author

***Abstract*—The development of wireless communication technology, sensor technology and so on, the spatial-temporal data record objects' movement that provide massive information about the activity regularity, due to the close relation between the mobile terminal and human. In this paper, we present a model of predicting the next location of an object that moves on the ground based on Markov chains that we coined as K time steps trajectory prediction algorithm (K-TSTP). We consider not only the spatial historical data but also consider the notion of time and the influence of location of historical time step in the prediction model. In order to evaluate the efficiency of our proposed prediction model, we use the data set that provided by Unicom. Experimental results show that our K-TSTP algorithm has increased the accuracy and reduced the execution time of prediction than the original Markov chain.**

Keywords-trajectory prediction; markov chain; spatial-temporal data

I. INTRODUCTION

The technology about get location, spatial orientation, sensor networks, wireless communication rapid development and popularization produce a huge variety of mobile objects used to represent: the track (such as animals, people, cars, airplanes, etc.) movement of data, temporal trajectory data now It has become one of the hot countries. An individual carrying the mobile phone unintentionally generates many spatial-temporal trajectories. When the individual does not know that his movements are recorded, these trajectory information can reflect the individual's actual activity regularity. Therefore, how to mine the spatial and temporal patterns of behavior patterns and related properties, to construct a reasonable trajectory prediction model, to achieve the optimal trajectory prediction is an urgent problem to be solved in the practical application. A Markov model is proposed for predicting the future individual's location of moving objects, which uses the K-means algorithm to extract the Points Of Interest (POIs) before building the Markov model in [3]. Ref. [5] is through the Markov model to predict the future location, but it extracted POIs through DJ-cluster algorithm. Based on the above research and data observation, we can find different frequent access points have their own functions, we discover frequent access points on the basis of the moving objects' activity regularity and frequent access points function. In [6], the authors put forward Markov chain model to predict the position of the vehicle, which Markov chain model is multiple order Markov chains. And the transition probabilities matrix used in the prediction model is calculated for each vehicle's basis, due to different moving object has different activity regularities and the transition matrixes are based on the historical data in the monitoring range. So the computation of the algorithm in [6] is larger. Trajectory prediction of moving objects is an important part of the research of mobility of human and social network. The research on the trajectory consisting of the spatial and temporal data is helpful to the research of the social network and all kinds of recommendation. Based on this purpose Spyropoulos et al. in [4] in order to well simulate the movement of real moving objects, a mobile model is proposed, called "community based mobility model". In their research, the model set the two states: the "roaming state" and the "local state", where the "roaming state" and the "local state" respectively represents the random motion in the direction outside the local community and the random motion in the direction within the local community. In terms of theoretical research, Ref. [10] through the study on flow pattern of anonymous mobile phone users to analyze the predictability of human activities. Their core idea is to construct a graph where each node is associated with the percentage of residence time in the cell.

Now most of the trajectory prediction is based on the geographical position, and ignores the time, the history and the interaction between them. So the performance of trajectory prediction based on Markov chain is not very good, or the applicability is not widely. In this paper, we present a new model of predicting the next location of an object that moves on the ground, we consider not only the spatial historical data but also consider the notion of time and the influence of location of historical time step in the prediction model.

The remainder of this paper is organized as follows. First, we describe the trajectory prediction model and the Trajectory

Prediction Algorithm (K-TSTP) for the trajectory prediction in Sect. 2. Afterwards, we present the experiment analysis and evaluation results with the data of campus students in Sect. 3. Finally, this work is concluded in Sect. 4.

II. TRAJECTORY PREDICTION

A. Trajectory Prediction Model

First, the trace set MT, which is obtained from the device having the function of GPS, is converted to the trajectory, and the trace is expressed as:

$$\mathrm{MT} = \{mt_1, mt_2, mt_3, \cdots, mt_n\} \tag{1}$$

where mt_n represents the trace of n-th moving objects. Each trace mt_k consists of trajectories of different time, and can be expressed as:

$$mt_k = \{T_1, T_2, T_3, \cdots, T_m\} \tag{2}$$

where m represents the different date / day, the trajectory T_j can be expressed as:

$$T_j = \{(loc_1, t_1), (loc_2, t_2), (loc_3, t_3), \cdots, (loc_n, t_n)\} \tag{3}$$

where loc_n and t_n in the tuple respectively represents the location and the timestamp.

After obtaining the above data, we preprocess the spatial-temporal trajectory, and make the trajectory into the following form:

$$PT_j = \{(FAP_1, t\_start_1, t\_end_1), (FAP_2, t\_start_2, t\_end_2), \cdots, (FAP_n, t\_start_n, t\_end_n)\} \tag{4}$$

where each of the elements in three tuple represents respectively: the location FAP_i of frequent moving objects activities, the arrival time t_start_i at stay point FAP_i, the living time t_end_i at stay point FAP_i.

In the Markov chain, the probability of one state (i.e., the location, the activity) moving to others depends on the historical probability and the current state, the historical probability indicates probability matrix of the transitions among states. More specifically, it is composed as follow:

1) Each position/activity/state, such as $FAP_1, FAP_2, \cdots, FAP_n$ corresponds to a frequent access points (FAP). According to the property, the future position depends on the current position and is independent of the past position.

2) The transition probability matrix, represents the probability of moving from position FAP_i to position FAP_j.

B. Trajectory Prediction Algorithm(K-TSTP Algorithm)

Therefore, we design the trajectory prediction algorithm to predict the next position of the objects. The algorithm is composed of four steps as follow:

Step 1. Extracting frequent access points for moving objects based on their activity regularity in the monitoring range. Trajectory is composed of a set of time series of trajectory points, frequent access points should satisfy the following conditions:

(1) $\mathrm{FAP}_k = \mathrm{Trajectory}_{i,k} \in \mathrm{Trajectory}_i = \langle \mathrm{Trajectory}_{i,i}, \mathrm{Trajectory}_{i,i}, \cdots, \mathrm{Trajectory}_{i,k}, \cdots, \mathrm{Trajectory}_{i,n} \rangle$

(2) $\mathrm{Trajectory}_{i,k} = (t_1, x_1, y_1), (t_2, x_2, y_2), \cdots, (t_i, x_i, y_i), \cdots, (t_m, x_m, y_m)$

where $\forall$ i, $\exists \varepsilon$, satisfy $d = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \le \varepsilon$;

(3) $|\mathrm{Trajectory}_{i,k}| = m \ge MinPoin$

where $\mathrm{Trajectory}_{i,k}$ represents the k-th trajectory of i-th trace, ε and $MinPoint$ respectively represents the distance threshold and the length threshold.

(4) $\mathrm{FAP} = MBB == \langle (MBB.t_{min}, MBB.t_{max}), (MBB.x_{min}, MBB.x_{max}), (MBB.y_{min}, MBB.y_{max}), (\forall p_j \in \mathrm{FAP}) \rangle$

where $MBB.t_{min}$ and $MBB.t_{max}$, $MBB.x_{min}$ and $MBB.x_{max}$, $MBB.y_{min}$ and $MBB.y_{max}$ respectively represents frequent access points The maximum and minimum values of the trajectory points contained in the frequent access points in each coordinate axis(including x axis, y axis and $time$ axis).

The trajectories after extracting the frequent access points can be expressed as follows:

$$\mathrm{Trajectory}_i = \mathrm{FAP}_{i,1}, \mathrm{FAP}_{i,2}, \cdots, \mathrm{FAP}_{i,k}, \cdots, \mathrm{FAP}_{i,n} \tag{5}$$

where $\mathrm{Trajectory}_i$ represents i-th trajectory, $\mathrm{FAP}_{i,n}$ represents the n-th frequent access point in i-th trajectory, n represents the number of frequent access points in i-th trajectory.

Step 2. Dividing the date, time and group of activity based on the activity regularity. The daily activity transition of the population is closely related to the time of transition, due to the special social attributes of human activities. Division of the crowd, the moving object is divided into work, students and other; division of activities time period, from 7:00 to 24:00; division of activities date, which is divided into working days(including Monday, Tuesday, Wednesday, Thursday, Friday), weekends and holidays.

Step 3. Computing the transition probability matrix of each date domain and each crowd. According to the **Step 2**, we can get 3 (3 kinds of crowd) *3 (3 kinds of activity date) = 9 transition probability matrix. As we know, the position which was passed a long time ago has little influence on the trajectory prediction. Thus we calculate the transition probability matrix for the $\mathbf{K}$ time step ago (as shown in Figure II). Finally, we can use the equation (6) to calculate the transition probability matrix for the K time step ago to get the final the transition probability matrix for Step 4.

$$ATPM_{1\,time\,step} = \begin{array}{c|ccc} & A & B & C \\ \hline A & 0.25 & 0.4 & 0.35 \\ B & 0.1 & 0.33 & 0.57 \\ C & 0.31 & 0.5 & 0.19 \end{array}$$

(a)The transition probability matrix for the one time step ago

$$ATPM_{2\,time\,step} = \begin{array}{c|ccc} & A & B & C \\ \hline AB & 0.25 & 0.4 & 0.35 \\ AC & 0.1 & 0.33 & 0.57 \\ BB & 0.33 & 0.48 & 0.19 \\ CA & 0.81 & 0.07 & 0.12 \\ CB & 0.62 & 0.1 & 0.28 \\ CC & 0.33 & 0 & 0.67 \end{array}$$

(b)The transition probability matrix for the tow time steps ago

FIGURE I. THE TRANSITION PROBABILITY MATRIX FOR THE ONE/TOW TIME STEPS AGO.

$$\begin{cases} w_1 = random(1/k?,1) \\ w_2 = random(0,1-w_1) \\ w_k = random(0,1-w_{k-1}) \\ ATPM_{final} = w_1 * ATPM_{1\,time\,step} + w_2 * ATPM_{2\,time\,step} + \cdots + w_k * ATPM_{k\,time\,step} \end{cases} \quad (6)$$

Step 4. According to the current state, the nearest access point as well as the transfer matrix to predicting the next position.

Algorithm: K-TSTP algorithm

Input： The original training trajectory data OT
Influence of k time steps on predictive value;
Output： the Prediction result

```
PT = Preprocessing the original training trajectory data by removing redundant trajectories;
Sub_PTs = Dividing the trajectory data after preprocessing PT into 9 sub-trajectories based on date and crowd;
for each Sub_PT in Sub_PTs:
for each trajectory point p in Sub_PT:
Put the trajectory point p into FAP set;
/* FAP set represents the data set of frequent access points */
end for;
ATPMs Computing the transition probability matrix for Sub_PT;
/* ATPMs represents 9 transition probability matrixes of each activity date domain and each activity crowd */
end for;
for each transition probability matrix ATPM in ATPMs:
    w_1 = random(1/k,1); w_k = random(0,1-w_{k-1});
 /*Computing the weight w_K for the k time step ago*/
 ATPM_final = w_1 * ATPM_1time step + w_2 * ATPM_2time step + ... + w_k * ATPM_k time step;
    /* ATPM_k time step represents the transition probability matrixes for the K time steps ago */
 for each row in ATPM_final:
 Find the column C that corresponds to the maximum value of the row;
 next_position = C;
end for;
end for;
return next_position;
/* next_position represents the result of prediction*/
```

III. EXPERIMENT EVALUATION

The experiment data in this paper are collected in district of shanghai, the acquisition device of the data set is the smart phone, the acquisition time interval is half an hour, data set provided by Unicom, samples of original data records as shown in Table I.

TABLE I. SAMPLES OF ORIGINAL DATA

id	date	Lat0	Lon0	…	Lat23	Lon23
F137XX21	20151201	121.2839	31.34085	…	121.4033	31.21361
20feXX51	20151201	121.3898	31.21461	…	121.2903	31.21461
…	…	…	…	…	…	…

In the experiments, we split the trajectory data into two sets: the training set, which is used to build prediction model, and the testing set, which is used to evaluate the performance of the prediction model. We verify the performance of the algorithm by using the running time, the accuracy and predictability of the algorithm.

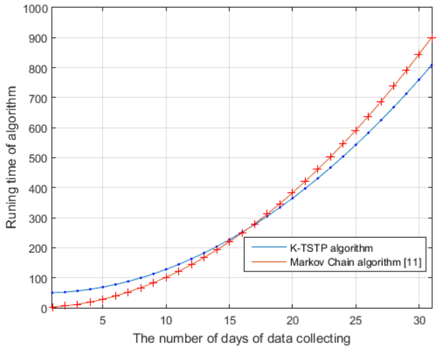

FIGURE II. THE FIGURE ABOUT RUNNING TIME OF ALGORITHM

We can see from Figure II, at the beginning of the implementation phase of the algorithm, the running time of K-TSTP algorithm is more than Markov Chain algorithm, with the increase of the amount of data, the running time of K-TSTP algorithm is less than Markov Chain algorithm, the K-TSTP algorithm has a clear advantage over Markov Chain algorithm when the amount of data is large. The time-consuming of K-TSTP algorithm at the beginning is larger, due to calculating the transition probability matrix for the time step ago to get the final the transition probability matrix for predicting.

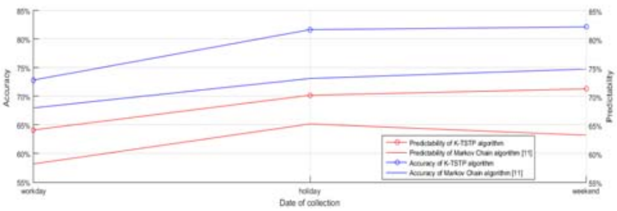

FIGURE III. EVALUATION RESULTS ON ACCURACY AND PREDICTABILITY

We can see from Figure III, the accuracy of K-TSTP ranges from 72.87 % to 82.15 % and the predictability ranges from 67.99 % to 74.18 %. While the accuracy of original Markov chain only ranges from 64.14 % to 71.31 % and the predictability is less than 67 %. The experimental results show that the TPA gives us a satisfying prediction result.

IV. CONCLUSION

In this paper, we have presented an algorithm for next location prediction called K time steps trajectory prediction algorithm (K-TSTP) that considers the notion of time and the influence of location of historical time step in the process of prediction. We dividing the date, time and group of activity based on the activity regularity and computing the transition probability matrix of each date domain and each crowd. Experiment results show that the accuracy of K-TSTP ranges from 72.87 % to 82.15 %. We concluded that the K-TSTP did much better in forecasting the trajectory than Markov chain [11].

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant No. 61301159, 61303267; Natural Science Foundation of Jiangsu Province under Grant No. BK20150721, BK20161469; China Postdoctoral Science Foundation under Grant No. 2015M582786, 2016T91017; Engineering Research Center of Jiangsu Province under Grant No. BM2014391. Primary Research & Development Plan of Jiangsu Province under Grant BE2015728.

REFERENCES

[1] Lerin P M, Yamamoto D, Takahashi N. Encoding network–constrained travel trajectories using routing algorithms[J]. International Journal of Knowledge and Web Intelligence, 2013, 4(1): 34-49.

[2] Kellaris G, Pelekis N, Theodoridis Y. Trajectory compression under network constraints[M]//Advances in Spatial and Temporal Databases. Springer Berlin Heidelberg, 2009: 392-398.

[3] Popa I S, Zeitouni K, Oria V, et al. Spatio-temporal compression of trajectories in road networks[J]. GeoInformatica, 2015, 19(1): 117-145.

[4] Spyropoulos, T., Psounis, K., Raghavendra, C.S.: Performance analysis of mobility-assisted routing. In: Proceedings of the 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), pp. 49–60 (2006)

[5] Giannotti F, Nanni M, Pedreschi D, et al. Unveiling the complexity of human mobility by querying and mining massive trajectory data[J]. Vldb Journal, 2011, 20(5):695-719.

[6] Parent C, Spaccapietra S, Renso C, et al. Semantic Trajectories Modeling and Analysis[J]. Acm Computing Surveys, 2013, 45(4):115-123.

[7] Stefanakis. SELF: Semantically Enriched Line simpliFication[J]. International Journal of Geographical Information Science, 2015, 29(10):1-19.

[8] Liu J, Zhao K, Sommer P, et al. Bounded Quadrant System: Error-bounded trajectory compression on the go[C]//Data Engineering (ICDE), 2015 IEEE 31st International Conference on. IEEE, 2015: 987-998.

[9] Lv C, Chen F, Xu Y, et al. A trajectory compression algorithm based on non-uniform quantization[C]//Fuzzy Systems and Knowledge Discovery (FSKD), 2015 12th International Conference on. IEEE, 2015: 2469-2474.

[10] Chen M, Xu M, Franti P. Compression of GPS trajectories using optimized approximation[C]//Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012: 3180-3183.

[11] Li, H., Shou, G., Hu, Y., Guo, Z.: Mobile edge computing : progress and challenges. In: 2016 4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, Oxford, United Kingdom, pp. 83–84 (2016)

[12] Gambs, S., Killijian, M.O., et al.: Next place prediction using mobility Markov chains. In: Proceedings of EuroSys 2012 Workshop on Measurement, Privacy, and Mobility (MPM), pp. 1–6 (2012)