International Core Journal of Engineering Volume 7 Issue 6, 2021
ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0035

Compression of Trajectory Data based on LSTM and Smoothed Analysis

Xuesong Chen[1,a], Zhiying Yang[1,b]

[1]College of Information Engineering, Shanghai Maritime University, Shanghai 200000, China.

[a]2652800047@qq.com, [b]zyyang@shmtu.edu.cn

Abstract

With the application of a large number of GPS,RFID,wireless sensors and smart devices, huge volume of moving object trajectory data have been produced, it brings great challenges in research and storage of the data. How to design more efficient compression algorithms has been a hot research area. In this paper, a trajectory data compression algorithm LSTC based on the idea of LSTM prediction and smoothing analysis is proposed. Prediction error is used as the compression threshold in algorithm LSTC. Algorithm LSTC first computes the prediction, and then calculates the distance difference between the predicted trajectory and the original trajectory, determines the compression threshold baised on smoothed analysis, and finally obtains the compressed trajectory by use of the threshold. Experiments demonstrated that algorithm LSTC has better compression efficiency.

Keywords

LSTM Prediction; Smoothed Analysis; Moving Object; Trajectory Compression; Automatic Threshold.

1. Introduction

With the development of traffic network and the popularization and application of GPS,RFID and a variety of wireless sensors, great capacity of moving object trajectory data are produced. People always use these trajectory data to perform data mining in order to get more valuable information. However, the huge volume of the data brings great challenge in existing research and storage devices. According to the report in literature[1], the track data volume of every 10 taxi cars is about 2 MB memory a day, while the demand of hundreds of millions of track data in real life keeps quickly growing. While big part of collected track data is of little value. Therefore, in order to save communication bandwidth and high storage costs, trajectory data compression is getting more and more important. People have design a large of compression algorithms in trajectory data compression research area.

There are three main directions in the field of trajectory compression: compression based on line segment, compression based on road network and compression based on semantics. More study works have done based on line segment trajectory compression. The earliest trajectory compression algorithm is the Douglas—Peucker algorithm (DP algorithm) proposed by Douglas and Peucker[2] in 1973. DP algorithm reserved the trajectory data points which have larger interested value than the compression threshold. The top-down time ratio algorithm (Top-Down Time Ratio algorithms,TD-TR) is proposed by Meratnia[3] et al. Compared with the DP algorithm, TD-TR algorithm uses synchronous Euclidean distance with considering time characteristics instead of vertical Euclidean distance PED, which can preserve more original information. Ke Bingqing et al[4] proposed Angular algorithm in 2016, they use cumulative angle deviation to pick and choose trajectory data points, the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0035

time complexity of Angular is O (n). Lin et al[5] proposed ATS algorithm for preserving velocity features. ATS algorithm segments the original trajectory based on velocity characteristic, and calculates the spatial error threshold of sub-trails and call DP algorithm to compress the result trajectory. Most all the previous compression algorithms considers single attribute of the trajectory, all require tedious experiments for selecting a better compression threshold, they can not determine the threshold suitable for the input trajectory data set. Many of them need more time cost and can not preserve effectively the space-time information of the original trajectory.

This paper proposes a trajectory data compression algorithm LSTC based on LSTM[6] prediction and smoothing analysis[7] to calculate the spatial distance difference between the predicted trajectory and the original trajectory. The spatio-temporal relationship between each trajectory point and the surrounding trajectory points is considered at same time. Algotithm LSTC calculates the compression threshold suitably for the given trajectory, it keeps more valuable information of the original trajectory.

2. Formulation of Algorithm LSTC

2.1 Algorithm LSTC framework

This section presents algorithm LSTC. It comprsses the trajectory data based on LSTM prediction and smoothing analysis (LSTC) as shown in Fig.1. LSTC first pretreates the collected original trajectory data set for eliminating noise to make the data being applicable to experiments. The compression threshold is obtained by using of smoothing analysis method. Details of the LSTC algorithm will be described in the next section.

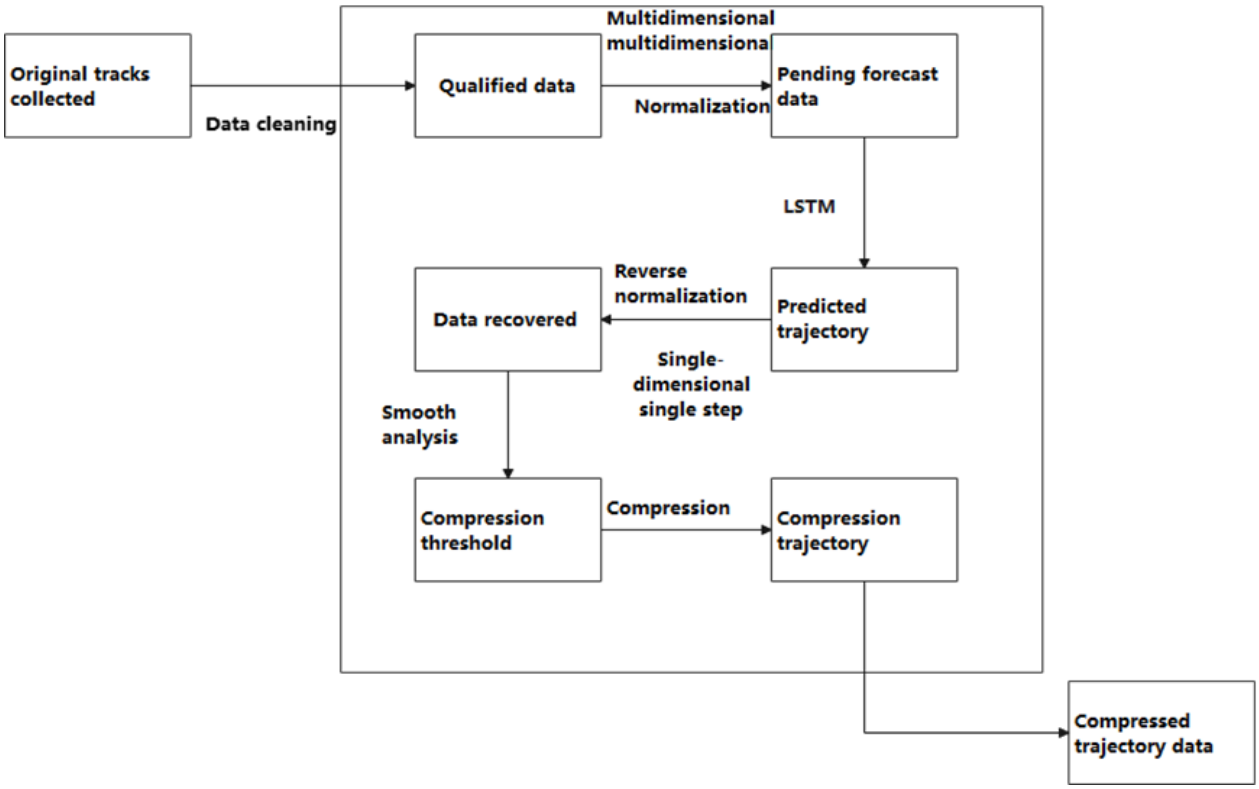

Fig. 1 Framework of LSTC algorithm

2.2 Constructing format data set

First of all, LSTC pretreamtes the original collected track data are as follows: (1)Processing incomplete data records: deleting the data records with more than 50% items missing, other

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0035

incomplete data records will be completed by use of mean value. (2) Deleting abnormal data records: The data record with abnormal higher speed will be removed. (3)Removing repetition point: the track points with longitude, latitude and height repetition.

To make the data set in suitable format, the following steps will be performed: (1) Data normalization: by use of min-max standard, all track data will be transfored to numbers in internal of [0,1] in order to avoid irregularities in the experimental results, which conducive to the processing of subsequent data. (2) Segmentation data set size: According to the ratio of 4 to 1 to divide the trajectory data set into training data set and testing data set. (3) Constructing the multi-demension data set: selecting longitude, latitude, altitude in the trajectory data as features, so the feature dimension of the resulting data records is 3. Using the data of first three moments to predict the data of last three moments to get training set and testing set. So the prediction time step is 3 as shown in Fig.2.

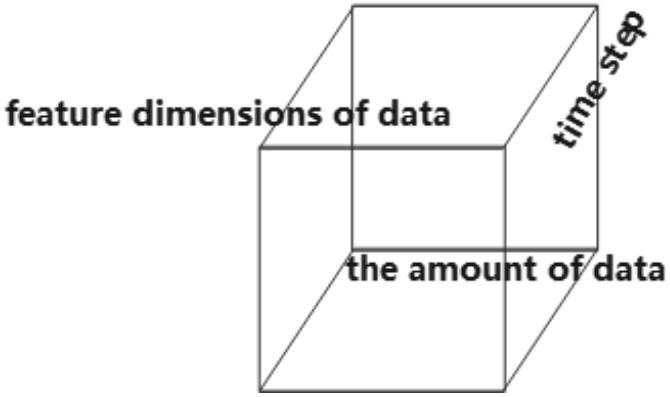

Fig. 2 The format of data set

2.3 Predictive network model LSTM

A long and short memory neural network LSTM is used to predict the trajectory of moving objects. LSTM includes input gates, forget doors and output doors shown in Fig.3. LSTM can avoid gradient disappearance and gradient explosion. C_{t-1} is the unit state of the last moment, h_{t-1} is the output value of the previous LSTM, σ is a sigmoid function, f_t and C_{t-1} calculate the historical information to be discarded, W_f is the weight of the forgotten door, W_i is the weight of the input gate, W_o is the weight of the output gate, X_t is the input value of the current network, i_t is to retain the new information that is the input door, O_t is the output gate of the calculation, C_t is the current state of the unit, the control parameter of the new data formation, h_t is the output value LSTM at the current time, that is, using the new control parameters to produce the output , b_f is forgotten door deviation, b_i is input door deviation, b_o is output door deviation. sigmoid function is used to calculate the memory state of the network as input.

LSTM memory unit will forget some unimportant historical information when predicting the output at a certain time, and remember some important historical information shown in formula (1).

$$f_t = \sigma(W_f[h_{t-1}, X_t] + b_f) \quad (1)$$

The input gate is a gated device used to control how much input is in and out or whether it is allowed in and out, it is shown in formula(2).

$$i_{t=}\sigma(W_i[h_t - 1, X_t] + b_i) \quad (2)$$

The output gate is used to control how much output is in the C_{t-1} of the state value to the output value LSTM at the current time. This is shown in formula (3).

$$O_t = \sigma(W_o[h_{t-1,}X_t] + b_o) \quad (3)$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0035

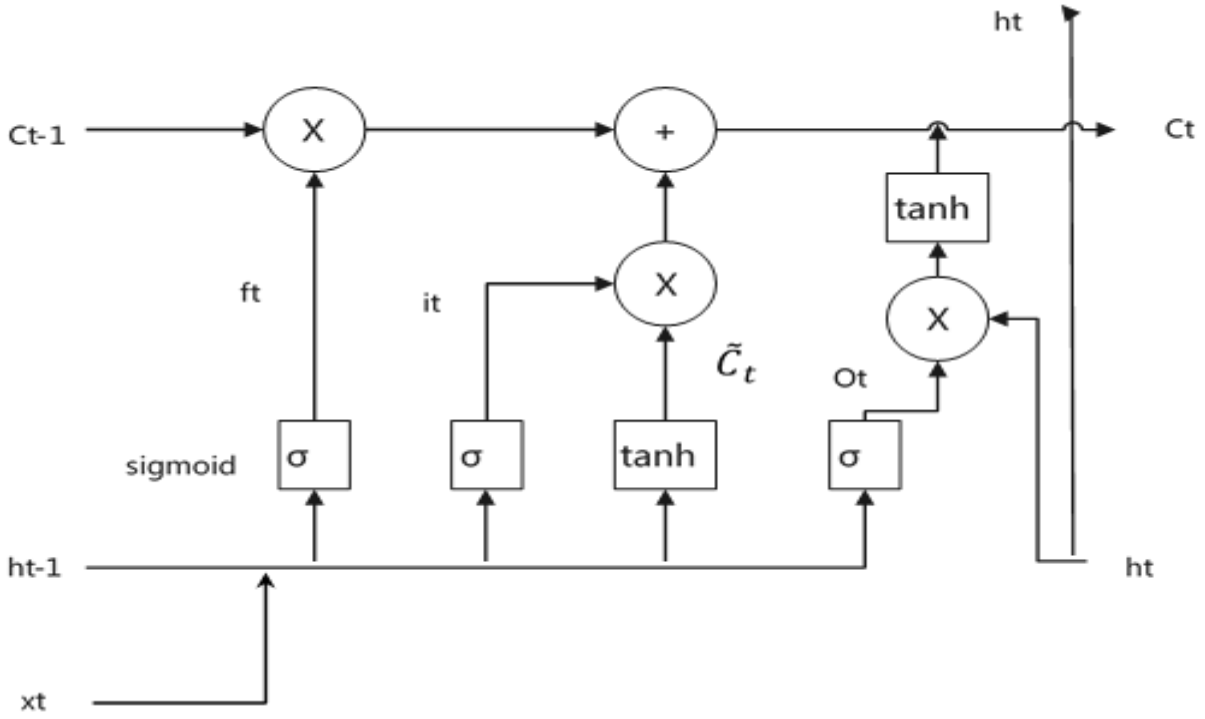

Fig. 3 LSTM

The prediction network model in this paper includes two layers of LSTM network and one layer of full connection layer Dense, shown in Table 1. In order to prevent the model from overfitting, we set the LSTM of each layer Dropout. Dropout to enhance the generalization of the model by stopping the neurons with a certain probability in the process of forward propagation. The error function uses mean square error MSE; optimizer as the Adam. We set the parameters of the network model according to many experiments,we set epochs=50, batch_size=64, verbose=1.

Table 1. The network model

Layer	Parameter
Lstm_1	4
Dropout_1	0.01
Lstm_2	64
Dropout_2	0.01
Dense	3

2.4 Recovery trajectory data format

The prediction data is multi-dimension and multi-step array data, each element in array data is two-dimensional array, the elements in first row of each two-dimensional array being read is extracted, and merged into a new two-dimensional array. This array has the format that we need.

2.5 Compression threshold by smoothed analysis

2.5.1 Distance calculated based on latitude and longitude

There is a certain error between the trajectory data obtained by neural network prediction and the original data. This error needs to be continuously reduced in the field of trajectory prediction, but in the field of trajectory compression, here, we use this error to form the compression threshold. The compressed trajectory is obtained by removing the data less than threshold. The predicted trajectory data can basically reflect the shape of the original trajectory, but it will not be the same as the original

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0035

data. The motion state of the trajectory data is unpredictable, and there are some slow motion trajectory points, which are relatively stable and have little contribution to the whole trajectory data set. Therefore, these trajectory points are the data needed by researchers and should be stored. trajectory prediction uses a 3-dimensional 3-step trajectory dataset. therefore, the data for each prediction is based on the first 3 data. for example: we use the 0,1,2 points of the original data to predict the 3 ,4,5 points, and 1,2,3 points of the original data to predict the 4 ,5,6 points. The more stable data have relatively small prediction error, and the prediction error of the point where the motion state changes dramatically is larger, and the point with large error is the point that the trajectory compression needs to be retained. Calculation of the distance between two latitude and longitude points using Haversine formula[8] , shown in formula (4) and (5), where d is the radius of the earth, lat1 and lat2 denote the latitude of two points and △λdenotes the difference between two points of longitude.

Haversine formula:

$$haver\,sin(d/ra) = haver\,sin(lat2 - lat1) + cos(lat1)\,cos(lat2)haver\,sin(\Delta\lambda) \quad (4)$$

Among them:

$$\text{haver}\,sin(\theta) = sin^2(\theta/2) = (1 - cos(\theta))/2 \quad (5)$$

2.5.2 Determination of compression threshold by smoothed analysis

Haversine Distance_error between the predicted trajectory and the original estimate calculated by the calculation formula reflects to some extent the intensity of the change of the motion state of the original trajectory. How to select the compression threshold is very important to the compression effect of the trajectory.

The concept of smoothing analysis determines the compression threshold: first set a threshold radius R, start at the first point in the Distance_error, calculate the average value within the radius R, and so on, so we get the Smooth_error. Then we get the maximum and minimum Smooth_error, calculate the average of the maximum and minimum, and the average value is the Compress_error we need.

3. Results and analysis of experiments

For the purpose of verifying the LSTC algorithm, We need validation on real data sets.Therefore, This paper selects the Geolife trajectory data set (Geo Life Trjectory Dataset)[9, 10]. Microsoft Research Geo Life the project collected the trajectories of 182 users over the five years from April 2007 to August 2012, to produce the gps trajectory dataset. GPS trajectories of this dataset are represented by a series of time points.Each point contains information such as latitude, longitude and height. Data sets are 1.55 GB, in size Including 17, 621 tracks.Total distance is 1, 292, 951 kilometres.The total time is 50, 176 hours. Among them, 91.5 per cent of the trajectories were sampled more intensively, about 10 meters or 5 seconds per ~,we randomly select three trajectory data to compare the memory size and its profile changes before and after compression.

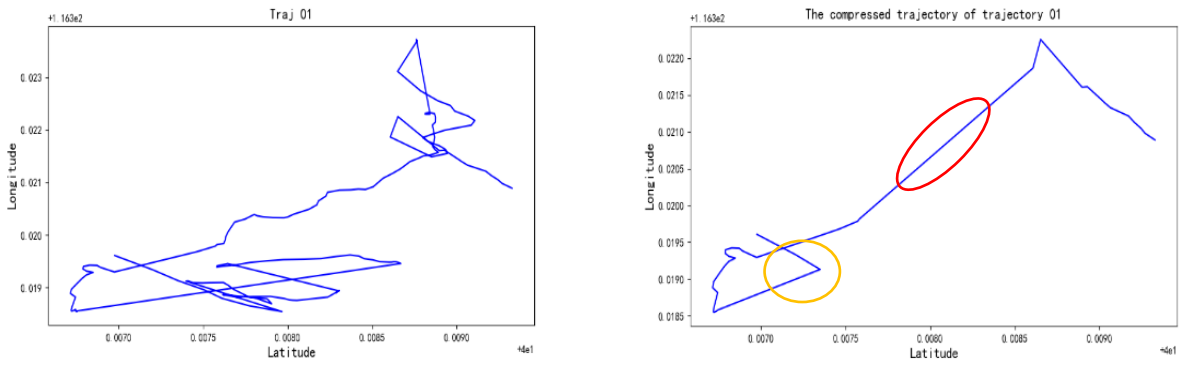

Fig. 4 Traj01

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0035

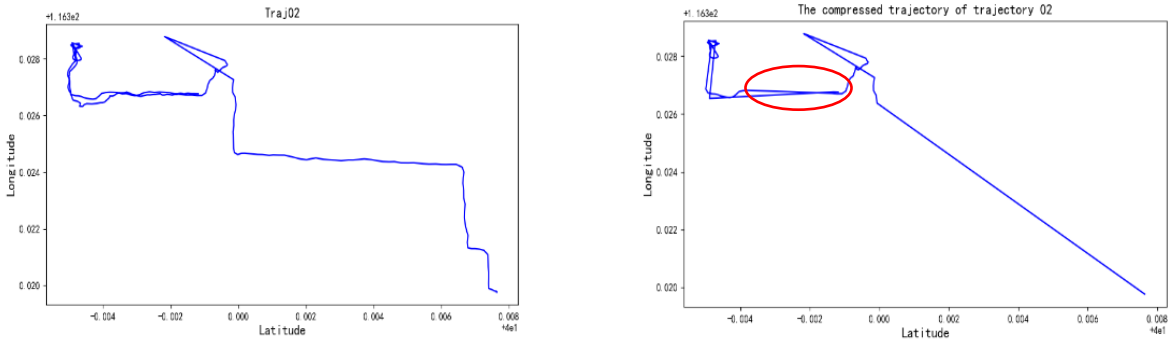

Fig. 5 Traj02

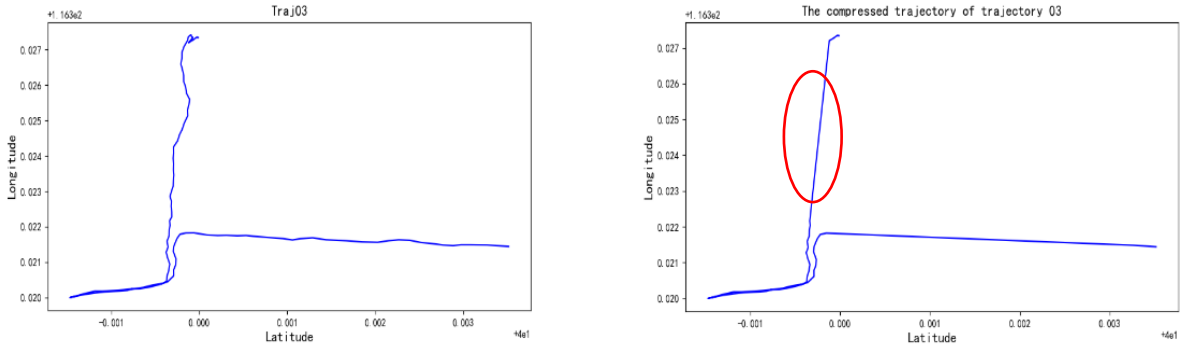

Fig. 6 Traj03

Table 2. The comparing

Trajectory	Compression(%)	R(m)
Traj01	0.7011494253	1/2(Traj01.length)
Traj02	0.6935483871	1/2(Traj01.length)
Traj03	0.45161290323	1/2(Traj01.length)

We set R=1/2 (Traj. length), in which case we obtain a compression threshold for each trajectory, preserving the trajectory points greater than the threshold, which are connected in chronological order to form a compression trajectory. The compressed trajectory is compared with the original trajectory, shown in Fig4, Fig5 and Fig6. We can find that the contours of the three trajectories are basically preserved, Traj02 and Traj03 are relatively complete, and approximately close to half of the original trajectory points shown in Table 2; From the red circle section in the Traj01,Traj02 and Traj03 , we found that the LSTC algorithm can feel the relatively stable trajectory data, which is also an important embodiment of the introduction of smoothed analysis into algorithm LSTC; On the contrary, from the yellow circle in the Traj01, we can observe that the LSTC algorithm is insensitive to the sharp change of direction.

4. Conclusion

Firstly, this paper analyzes the disadvantage that the traditional trajectory compression algorithm can not determine the threshold according to the characteristics of the trajectory itself, and then proposes algorithm LSTC to overcome this shortcoming. Firstly, algorithm LSTC uses the neural network model LSTM to obtain the predicted trajectory. Secondly, the distance error between the predicted trajectory and the original trajectory is calculated. The compression threshold is determined by use of smoothing analysis principle. Finally, the compressed data is obtained by deleting the trajectory

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0035

points less than the threshold. According to the experimental results, algorithm LSTC is more suitable for the stationary trajectory data compression.

Algorithm LSTC introduces neural network and smoothing analysis principle in trajectory data compression. It is more suitable for the compression of stationary trajectory. According to this inspiration, algorithm LSTC can continue to modify and add angle error to improve the original algorithm. The compression algorithm after adding angle error will not only be suitable for stationary trajectory data, but also for larger rotation trajectory data. We will also continue to try the above ideas in the future work.

References

[1] MERATNIAN, ROLFA. Spatiotemporal compression techniques for moving point objects [C]// EDBT 2004: Proceedings of the 9th International Conference on Extending Database Technology. Berlin Springer, 2004: 765-782.

[2] Douglas D H, Peucker T K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature[J]. Cartographica the International Journal for Geographic Informa-lion & Geovisualization, 1973.10(2):112-122.

[3] Meratnia N, By R A D. Spatiotemporal compression techniques for moving point objects [C]// Procof International Conference on Extending Database Technology.Berlin:Springer, 2004:765-782.

[4] B. Ke, J. Shao, Y. Zhang, et al. An online approach for direction-based trajectory compression with error bound guarantee[C]. Asia-Paci?c Web Conference, 2016, 79–91.

[5] Lin c H, Hung C C, Lei P R.A velocity—preserving trajectory simplifi—cation approach [C]// Proc of Conference on Technologies and Applications of Artificial Intelligence.Piscataway, NJ:IEEE Press, 2017:58—65.

[6] ALAHIA, GOEL K, RAMANATHAN V, et al.Social LSTM: Human Trajectory Prediction in Crowded Spaces[C]. IEEE Conference on Computer Vision and PatternRecognition. IEEE Computer Society, 2016: 961-971.

[7] Erkorkmaz K, Altinatas Y.High speed CNC systemdesign. Part 1. Jerk limited trajectory generation andquintic spline interpolation[J]. International Journalof Machine Tools and Manufacture Tools and Manufacture, 2001, 41:1323-1345.

[8] Web Site:https://blog.csdn.net/xiejm2333/article/details/73297004

[9] Zheng Y, Li Q, Chen Y, Xie X, Ma W-Y (2008) Understanding mobility based on GPS data. In: Proceedings of the 10th international conference on ubiquitous computing (UbiComp), pp 312–321.

[10] Zheng Y, Zhang L, Xie X, Ma W-Y (2009) Mining interesting locations and travel sequences from GPS trajectories. In: Proceedings of the 18th international conference on world wide web (WWW),pp791–800.