Recommend Driving Area-Route Based on Taxi Trajectory Experience

Lei Du, Hongyuan Yang
School of Software Engineering, Beijing Jiaotong University, Beijing, 100089, China
E-mail: 18121723@bjtu.edu.cn, 17221294@bjtu.edu.cn

Abstract—As the era of big data has arrived, it is no longer a problem to obtain sufficient GPS data for research and practical applications. The taxi trajectories contain the driving experience of the taxi driver. In this paper, we will use the trajectory experience of taxis to recommend the driving area-routes. We propose a grid method that can extract the trajectory information of a specific road segment efficiently. Then, we use the processed information to form a probabilistic statistical model of the recommended route. We compare the recommended routes at different times, and compare the recommended routes with the routes in Baidu navigation. It is recommended to select the optimal route according to different time periods, and the recommended routes are basically consistent with the results in the navigation.

Keywords—taxi trajectory, grid method, route recommendation

I. INTRODUCTION

Given the increasingly complex traffic network and the increased traffic flow, how to develop an effective driving route recommendation strategy has become an inevitable topic in the transportation field. An efficient recommendation strategy not only allows travelers to reach their destinations faster, but also eases traffic jams. Taxi drivers generally have a wealth of driving experience and will choose the most efficient route to reach their destination to get the most profit. When selecting a route, they will consider distance, potential traffic congestion, traffic lights and other traffic information. These traffic factor considerations are learned by experienced drivers, which imply that the intelligence of experienced drivers may serve as a valuable resource to learn practically fast driving route data instead of using map match algorithm and digital vector map

A. Dataset Exploration

The dataset includes 500 thousand trajectories each day during a week in Beijing. Each trajectory records longitude, latitude coordinates and acquisition timestamps. The time interval of acquisition varies from 10 seconds to 2 minutes.

B. Dataset Processing

Coordinates in GPS are represented by longitude and altitude, which cannot be applied in plane calculations directly. As UTM is a projected rectangular coordinate system represented by X and Y, we will convert GPS coordinates into UTM coordinates.

Data acquisition frequency plays an important role in the selection of grid size. The selection of the grid size should conform to the principle that the road structure in the grid should be as simple as possible. This makes the grid as small as possible to some extent. However, the smaller the grid, the
The grid method extracts the taxi trajectory and reduces the amount of data describing the trajectory while maintaining the trajectory’s characteristics. It converts the original series of coordinates into the index of the grid through which the trajectory passes, and makes the calculation of the massive trajectory data more efficient. In addition, the grid method can effectively map the trajectory to the actual coordinate area instead of the complex and time-consuming map matching algorithm.

The grid method divides the two-dimensional space into \( n^2 \) square grids, each of which has four equal sides (meters) for in-out directions. As shown in Fig. 1, the area is divided into 3*3 square grids. \( G_n \) is the grid index and each grid has a side length of 1000 m. The track entrance-exit directions are sequentially labeled as 1, 2, 3, and 4 in a clockwise direction.

As shown in Fig. 2, we have gridded the trajectories on the three-ring area of Beijing. The red outline structure of the road network in the figure is formed by a large number of taxi track points.

After the trajectory with a series of coordinates is processed by the grid, the trajectory information \( G = \{g_0(d_1, d_2), g_1(d_1, d_2), g_2(d_1, d_2), ... \} \) can be extracted, \( g_n \) is recorded as the passing grid index, \( d_1 \) is the direction into the grid, and \( d_2 \) is the direction out of the grid. As shown in Fig. 3, the blue points are the recorded track points, labeled \( T_n(x_n, y_n, t_n) \) containing horizontal and vertical coordinates and time point information. The blue line is the real trajectory. The red line uses the line between the two track points to simulate the real trajectory, and the boundary with the grid will form an entry-exit point, denoted as \( M_n(\bar{x}_n, \bar{y}_n, \bar{t}_n) \).

### III. PRELIMINARIES

#### A. Grid Method

In daily life, the driver who is familiar with the road structure, but because of the complexity of the road conditions, does not know which way to go. Area-Route Recommend is a good solution to this problem. It takes experience from the trajectory of a large number of taxi drivers and recommended general direction and route area. This is very in line with the way people think. When recommend the route, usually don't need to express the specific road to go. The driver only needs to remember the key areas and directions to go through. The recommended Area-Route is denoted as Path = \{\( G_1, G_2, G_3, G_4 \) ... \} and \( G_n \) is the passed grid number.

#### IV. AREA-ROUTE RECOMMEND

##### A. Progressive Region Approach

We propose a progressive region approach to recommend route regions by moving closer to the region where the target grid is located until the target grid is reached. The area around the grid consists of some annular areas surrounding the grid and spreads out. The center grid is the destination grid, and each circle that expands outward is a region \( R_m \). Fig. 4 shows the correlation of the ring region (0 ring refers to the target grid itself) to the target grid, recorded as \( \alpha_m \).

\[
\alpha_m = \frac{N(R_m, G_e)}{N(R_m)} \tag{1}
\]

\( N() \) denotes the number of tracks, and \( N(R_m, G_e) \) is the number of track that passes through the annular range \( R_m \) and reaches the destination \( G_e \). \( N(R_m) \) is the number of tracks pass through \( R_m \).

The grid method divides the two-dimensional space into a plurality of square areas, recommending the area to be passed by the user. The recommended path area is denoted as \( P = \{G_1, G_2, G_3, G_4 \ldots \} \) and \( G_n \) is the passed grid number. Record that the current grid is \( G_e \), there are four optional directions \( D_1, D_2, D_3, D_4 \) can choose to enter the next grid \( G_{d_n} \), the destination grid is \( G_e \), and we define the annular region \( R_n \) around the \( G_e \), as show in Fig. 5, in the right corner. Then the probability of going out from each direction is

\[
P_n = \alpha_0 N(D_n, R_0) + \alpha_1 N(D_n, R_1) + \cdots + \alpha_m N(D_n, R_m) \tag{2}
\]
\(N(D_n, R_m)\) is the number of track that passes through the direction \(D_n\) and reaches the destination annular region \(R_m\), as the left picture of Fig. 5.

**B. Trajectory Probability Analysis Over Time**

Since the traffic flow is dynamic, the selected path will change over time. Fig. 6 shows the statistics of the number of trajectories from two different directions of one grid to another as time passes.

The distribution indicates that most of the time the taxi will more prone to take the 2 directions to reach the destination grid, but the trajectory in the 1 direction will be slightly more than 2 directions from 16:30 to 15:30. Since we collect data for one week, we can eliminate the accidental situation. The possible reason is that the 2 directions are prone to traffic jams and the number of vehicles in the 1 direction is increased. According to this characteristic, we divide the time period. In the calculation of \(N(D_n, R_2)\), we count the number of trajectory within half an hour of the recommended time.

**V. EXPERIMENT**

We tested on the loop in Beijing and experimented six recommended area-routes for the same endpoints at different starting points at 8:00 am. We compared the recommended results with the recommended results of Baidu map, and the results are shown above (Fig. 7). The continuous color squares are the recommended regional routes of the model, and the solid color lines are the specific routes recommended by Baidu map. The comparison shows that except for the red, green, and pink paths, there are some inconsistencies, and the other routes are consistent.

In addition, we also conducted recommended tests at different time periods, as shown in Fig. 8 (A), during the peak period (approximately 8:00 am to 9:00 am, 16:00 pm to 18:00 pm), the red path is recommended. The black path is recommended for the rest of the time. Although the black path is longer than red, it is a loop and there are fewer red lights, which results in shorter time. Drivers prefer to follow this path. In Fig. 8 (B), the test corresponds to Fig. 6 in which driver is more inclined to take the direction 2, so the blue line is recommended most of the time, but during the peak hours, the driver will gradually turn black because the
blue line is getting congested. And the black path will be recommended at this time.

VI. CONCLUSION AND FUTURE WORKS

A. Conclusion

In this paper, we propose a model for fuzzy recommendation of driving routes based on a large number of taxi trajectories. We dig the driver's driving experience from the taxi trajectory data to recommend an efficient driving area in a certain time. Because the acquisition of road network structure information and the credibility of map matching algorithm are quite difficult, this paper uses the grid method to process a large amount of trajectory data, and it is faster and more efficient.

B. Future Works

The size of the grid can be compared. The time interval of coordinate acquisition plays a key role in the selection of the grid. How much grid can extract the trajectory features corresponding to the time interval can be further tested. Since some roads are separated by a corner of the square and are difficult to group on one side. The shape of the grid can be further tested, like a hexagon or a shape with more sides. The hexagon can handle such problems better.

ACKNOWLEDGMENT

This work was supported by the National College Student Innovation Training Program. (No. 190185013)

REFERENCES