THE TAXIS' EXPERIENCE KNOWLEDGE MODELING AND ROUTE PLANNING

Tang Luliang* Li Qingquan Chang Xiaomeng

State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, 430079, China (tll, qqli)@whu.edu.cn; xiaomeng_chang@sina.com.cn

Commission VI, WG VI/4

KEY WORDS: floating car data (FCD); TEKM; route planning; GIS-T

ABSTRACT:

It is believed that the routes chosen by taxi drivers are more reasonable than regular drivers, and the dynamic navigation based on real-time traffic information is not mature both in the software and hardware fields, and the real-time data acquisition, data processing and communication need to be improved, and short-term traffic forecast in large area is still not accurate enough to route planning. So, it is important to plan the route using taxi drivers' experience and knowledge. This paper establishes a taxis' experience knowledge model (TEKM) based on analyzing historical floating car data (FCD) and taxi drivers' planning route rules. The road network is classified into different experience levels network, and a new route planning algorithm is proposed based on TEKM. In the end, we conducted 350 experiments with the road network and historical FCD of Wuhan, and compared the TEKM with the traditional shortest path algorithm. The results show us that the routes planed by TEKM are more coherent and consistent, and that the travelling time can be reduced.

1. INTRODUCTION

In recent years, with the social and economic development, the amount of the motor vehicles increases rapidly. Transportation problems are becoming increasing seriously especially such as traffic congestions, traffic accidents and the conflicts between the roads and the vehicles (Mcqueen, 2002).

The existing route planning methods generally tend to find shortest distance (Fisher, 1990; Lu, 2001; Cherkassky, 1996) or the shortest time (Lu, et al,1999), and some scholars conducted some research to find solutions when considering the turn penalties and prohibitions conditions (Han, et al., 2002; Ren, et al., 2004; Zheng, et al.,2004). To meet people's demand, some navigation systems propose some routing strategies such as highway priority, freeway priority, and minimum travel expenses. Hector Gonzalez presents a self-adaptive fastest path algorithm capable of efficiently accounting for important driving and speed patterns using historical traffic data (Gonzalez, Han & Li, 2007).

All existing researches few care about people's cognition, while people's subjective consciousness of choosing road plays an important role in route planning which is generated from long-term travel activities and experience. It is believed that the routes chosen by taxi drivers are more reasonable and feasible than the public. At present, dynamic navigation based on real-time traffic information is not mature both in the software and hardware fields, and the real-time data acquisition, data processing and communication need to be improved, and short-term traffic forecast in large area is still not accurate enough to route planning. So it is important and potential for people to daily travel using the taxi drivers' experience for assisting route planning. This paper establishes a taxis' experience knowledge model (TEKM) based on analyzing the historical FCD and the taxi drivers' route planning experience and cognition. Based on this model, the road network is classified into different experience levels network, and a new route planning algorithm is proposed which is based on TEKM.

2. TAXIS' EXPERIENCE KNOWLEDGE MODEL FOR ROUTE PLANNING

2.1 The taxis' cognition for road choosing

When travelling, people generally choose the shortest path, whereas the taxi drivers do not. They will choose the ideal route according to their own routing experience and cognition generated in the long term driving activities. The routing rules are based on the following observations:

① Taxi drivers are familiar with the traffic conditions, which make them that they can easily avoid the congestion sections during traffic peak hours, though the route planned by taxi driver is not the shortest one, the travelling time can be reduced significantly.

② Taxi drivers usually choose higher level roads in the road network, except that the traffic volume of these roads is high or there are low-level roads with significant advantages over the high ones, which make that the route planned by taxi driver is coherent and consistent.

③ Taxi drivers have less probability to choose the out-of-the-way and remote road sections with poor availability, while the route planning with the traditional shortest path algorithms is not well controlled, and sometimes the route sections are too narrow or poor traffic environment, which make that the route planned by taxi driver can avoid these problems.

2.2 The framework of Taxis' experience and knowledge modeling

This paper establishes a taxis' experience knowledge model (TEKM) for public route planning, which is not only taking into consideration the route distance, but also the travelling time and the taxi drivers' experience and cognition, The framework is presented in Fig 1.

* Corresponding author, tll@whu.edu.cn or tangluliang@hotmail.com

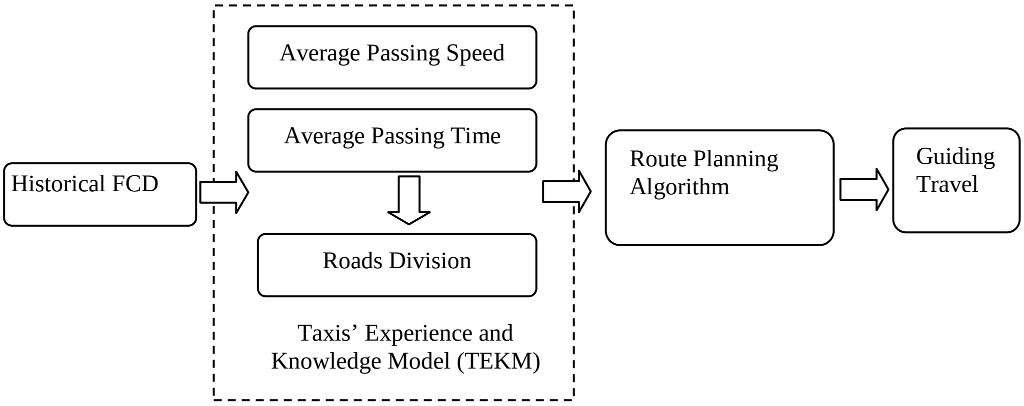

Fig.1. Steps of taxis' experience and knowledge modelling

① TEKM Acquires the road network average speed and the time when the floating car passing by the road segment, and obtains the historical traffic state in specific time of day by analyzing and processing historical FCD.
② TEKM Obtains the floating car passing times by the road section by statistic and analyzing the historical FCD, and establish the road experience grades classification.
③ Based on the historical traffic state and the road experience grades classification, TEKM establishes the taxi drivers' experience and knowledge routing model.
④ According to the taxi drivers' experience knowledge model, this paper proposes a new route planning algorithm to guide public route planning.

2.3 Taxis' experience knowledge model

2.3.1 The floating car passing time and average passing speed in the each road section: As the floating cars usually run the whole day and have a high coverage of the city road network, it is believed that it would be accurate to estimate the average passing speed and passing time on each road segment in specific time section by analyzing the historical FCD. TEKM acquires the average passing speed by summarizing the amount of the vehicles and summarizing each floating car speed on the road section. It means that the road average passing speed of the whole vehicles on the road segment i in specific time period t, which can be showed by following equation 1.

$$\overline{V_i}(t) = \frac{\sum_{j=1}^{m}\sum_{k=1}^{n_{(i,j)}} V_{(i,j)}^{k}}{n_i(t)} \cdot f(t) \qquad (1)$$

Where k is the numbers of the floating cars, $V_j^{i,k}$ is the k^th sampling point instantaneous speed of the floating car j on road segment *i*, n_j^i is the sampling point numbers of the floating car j on road segment *i*, $n_i(t)$ is the whole sampling point numbers of all the floating cars in time section t and $n_i(t)=\sum_{j=1}^{m} n_j^i$. $f(t)$ is the parameter that controls the availability of the sampling FCD, when the sampling point data is in time period *t*, $f(t)$ is 1, otherwise $f(t)$ is 0.

With the average passing speed $\overline{V_i}(t)$, it is easy to acquire the experience average passing time on road *i* in time period *t*.

$$\overline{T_i}(t) = \frac{L_i}{\overline{V_i}(t)} \qquad (2)$$

2.3.2 Road network division: Urban roads is generally divided into highways, main roads, sub-trunk roads, slip roads, but such road network classification is not practically useful for guiding route planning. Some scholars have conducted some research on floating car data to reveal time space distribution characteristics of the road network (Xin, Chen & Lin, 2008). Inspired by these studies, this paper analyzes the historical FCD and summarizes the amount of average float car passing times on each road segment. Based on the frequency, this paper divides the road network into several high or low experience grades.

Usually there are two sampling modes of collecting FCD, one collecting FCD mode is by interval time, and the other is by interval distance. As the sampling result is affected by the length of road segments and the road traffic conditions, the distribution of the amount of sample FCD on each road segment is not balanced. It is accurate to summarize the float car passing times on some road segment after recovering each floating car**'s** route, whereas the complexity of the algorithm is highly increased. In this paper, in the condition of that there are large sample FCD, we use statistical methods to analyze the historical FCD which is up to the probability distribution in statistics. When the sampling mode is time dependent, the average floating car passing time on road segment *i* in time period *t* can be computed by the equation 3:

$$\overline{n_i(t)} = \frac{\Delta T \cdot n_i(t)}{T_i} \qquad (3)$$

Where ΔT is the sampling time interval, T_i is the experience passing time on the road segment i, $n_i(t)$ is the amount of the sample FCD, $\overline{n_i(t)}$ stands for the average float car passing times on the road segment *i* in time period *t*.

Though there are great differences in road network division between different regions and countries, the road network is generally classified into four to six grades in various standards and specifications. In this paper, by taking the following logarithmic model, the roads are classified into high or low levels based on the passing frequency of the floating car in each road section.

$$C_i(t) = \log_{\alpha}^{\overline{n_i(t)}} \tag{4}$$

Where α is a parameter that controls the grades of the road network division and should be appropriately set according to real conditions Generally the road network is divided into four to six levels, different kinds of roads will not be notably classified and the model accuracy will be reduced if the grades are too small or too large. Suppose the largest average float car passing times of all the road segments in some time period is M, the road grades wanted to be *L*, *α* is supposed to be set with $\alpha = M^{\frac{1}{L+2}}$. It is because in logarithmic model, when the base number is small, increasing with the amount of the sampling FCD, the model value increases fast, and sometimes some road segments with little difference can be divided into different levels. One solution is that first the roads is classified into L+2 grades according to the original model, and then done with some merging operation. In this paper, the roads of level 1 and level 2 are merged into the first grade and level 3 and level 4 ones are merged into the second grade, and then the roads network is reclassified into L grades. This method can better divide the road network.

Due to the traffic volume in each road segment varies over the time period, and the road congestion degree during day is also different from the night, the taxi drivers' choice will be adjusted accordingly. The road network hierarchy that built by the driving experience and knowledge will also be changed from day to evening.

As the sample amount of the FCD in each road segment is closely related with the time, it is important to take the time factor into consideration. According to the people's traveling rules and the features of traffic condition, this paper uses 2 hours as a time interval and divides all day into 12 time sections. The specific division is as Tab 1:

Tab. 1 Traffic period division of the whole day

Time section(t)	1	2	3	4	5	……
Time interval	1:00—3:00	3:00—5:00	5:00—7:00	7:00—9:00	9:0—11:00	……

When we perform experience classification of city road network according to eq.4, we can find that there are few samples (Fig.2 (A)) collected by floating car in shorter road sections or at some intersections due to shorter passing time on these roads and adjacent road sections' road level is not continuous. So we adopt a smoothing method to balance floating car sample data quantity of adjacent road sections. The processed floating car data can reflect the actual passing times of car in every road section more accurately. The problem that there are usually short of sample data in some shorter roads or at some intersections can be solved. This solution can also restore road section's real experience level so that it maintains the consistency and coherence of adjacent road sections.

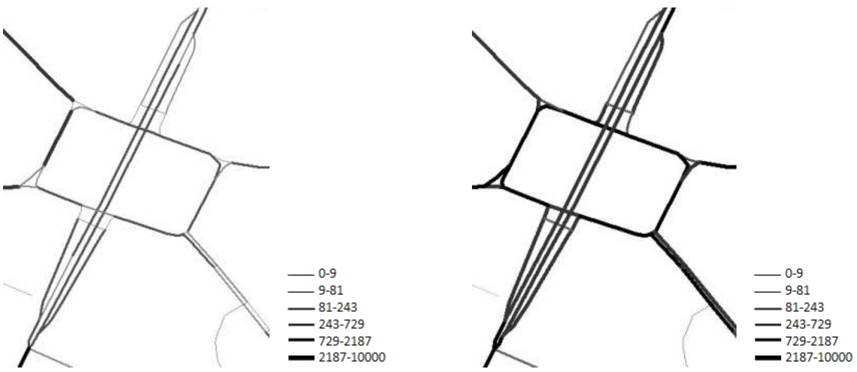

Fig. 2 Classification of the experienced levels at some intersection. (A) the classification result using the direct method, (B) the classification result using the smoothing method

We can obtain Wuhan City's traffic road experience level network according to eq.4 by taking advantage of historical average passing times of road section when we set α equal to 3 and t equal to 4. As is shown in Fig.3, road experience level network which is classified on the basis of floating car's occurrence frequency on every road section can better maintain the figure of Wuhan City's traffic artery and also reflects accessibility situation of road to a certain extent.

2.3.3 Taxis' experience knowledge modeling

When routing, common people often pay more attention to some factors such as the route distance, travelling time, road grades, traffic conditions, travel dispense, environment surrounding with the roads and other conditions, whereas the taxi drivers will take into consideration all these factors overall and form their own subjective routing cognition. Comprehensively considering the travel time and trip distance as well as the most taxi drivers' routing cognition and rules, this paper establishes an experience and knowledge model.

$$RoutePlane(V,E) = E[T(t), S, C(t)] \tag{5}$$

Where S is the trip distance, *T(t)* is the travel time in specific time period, *C(t)* is the average experience grades of all the road segments during the trip. It is described as finding the minimum value of the routing function $E[T(t), S, C(t)]$ and $RoutePlane(V,E)$ is the best route.

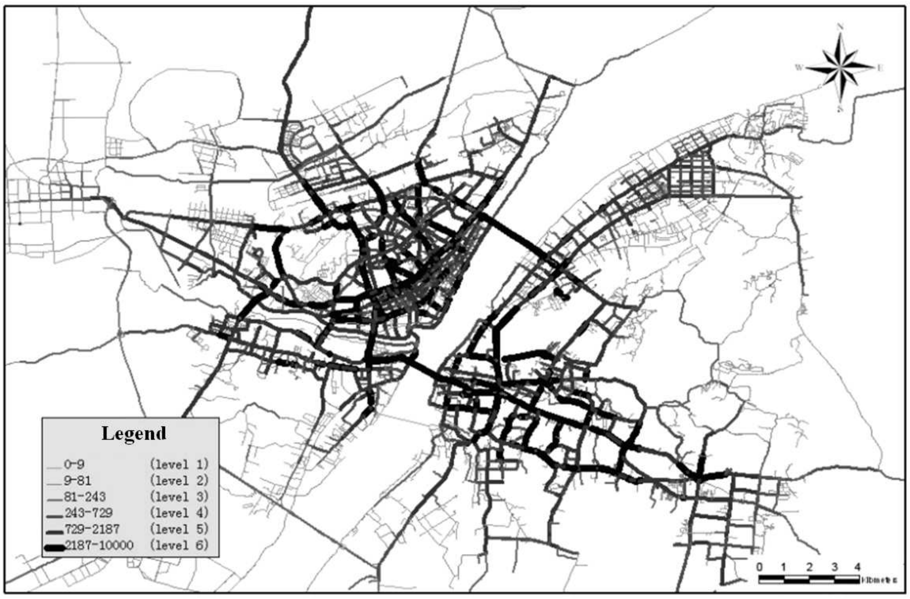

Fig. 3 Experience levels of Wuhan road network for the period 7:00—9:00 A.M.

3. ROUTE PLANNING BASED ON TEKM

3.1 Experience knowledge value of road section

The experience knowledge model based on taxi drivers' experience comprehensively takes into account three factors: path length which is most considered about, average passing time and the road experience level. This paper presents a unified description of these factors by using experience knowledge value of road sections. The experience knowledge value of road section in city road network is defined as the following equation 6:

$$W_i(t) = \frac{\mu_1 \cdot \tilde{T}_i(t) + \mu_2 \cdot \tilde{S}_i}{\mu_3 \cdot \tilde{C}_i(t)} \tag{6}$$

In the equation 6, $\tilde{T}$ is the experience average passing time, $\tilde{S}$ is the length of the road section, and $\tilde{C}$ is the road experience level. And average passing time, path length, road experience level should be non-dimensionally normalized. μ_1, μ_2, μ_3 respectively affects the importance of these three factors to the experience knowledge value of each road section.

For the road segment length and the experience average passing time, we adopt the same non-dimensional normalization approach. Road segment length can be non-dimensional normalized by the following equation:

$$\tilde{S}_i = \frac{L_i}{Min(S)} \tag{7}$$

Where Min(S) is the total length of the shortest path. Average passing time of road section i can be non-dimensional normalized by eq.8:

$$\tilde{T}_i(t) = \frac{\overline{T}_i(t)}{Min(T)} \tag{8}$$

$$\overline{T}_i(t) = \frac{L_i}{\overline{V}_i(t)} \tag{9}$$

Where Min(T) is the passing time of the fastest path that computed based on the experience average passing time of each road section.

Road experience level can be non-dimensional normalized by the following equation:

$$\tilde{E}_i(t) = \frac{\log_{\alpha}^{\tilde{n}_i(t)}}{\log_{\alpha}^{Max(n)}} \tag{10}$$

Where Max(n) is the maximum average floating car passing times of all the road sections in a given time interval. For $W_i(t)$, μ_1, μ_2, μ_3 are the coefficients of each factor, however, the actual coefficients that influence the experience model are $\frac{\mu_1}{\mu_3}$ and $\frac{\mu_2}{\mu_3}$. $\frac{\mu_1}{\mu_3}$ is defined as the index ratio B_s of distance to experience level, and $\frac{\mu_2}{\mu_3}$ is defined as the index ratio B_t of passing time to experience level. In order to take into consideration of the influence degree of various factors to route planning and route choice, the sum of B_s and B_t should be 1. As a result, the number of factors which influences the weight of each road section can be reduced, which is the index ratio of distance to experience level or the index ratio of passing time to experience level. So the weight formula of road section can be simplified as equation 11 or 12:

$$W_i(t) = B_t \frac{\tilde{T}_i(t)}{\tilde{C}_i(t)} + (1 - B_t) \frac{\tilde{S}_i}{\tilde{C}_i(t)} \tag{11}$$

$$W_i(t) = (1 - B_s)\frac{T_i(t)}{C_i(t)} + B_s \frac{S_i}{C_i(t)} \tag{12}$$

If $\frac{\mu_1}{\mu_3} > \frac{\mu_2}{\mu_3}$, it is a routing strategy of distance priority, otherwise, it is a time priority strategy.

3.2 A route planning algorithm based on TEKM

After setting the corresponding weights for road sections of the urban city network according to the weight formula, we propose a route planning algorithm based on experience knowledge model and establish an objective equation of route planning algorithm. This algorithm based on experience knowledge of taxi drivers' experience and cognition is contributed to route planning for other vehicles. The route planning arithmetic operators based on experience knowledge can be described as equation 13:

$$\overline{Dest(t)} = \sum_{i=1}^{N} W_i(t) \Big/ N$$

$$= \sum_{i=1}^{N}\left[B_t \frac{T_i(t)}{C_i(t)} + (1 - B_t)\frac{S_i}{C_i(t)}\right] \Big/ N$$

$$= \sum_{i=1}^{N}\left[(1 - B_s)\frac{T_i(t)}{C_i(t)} + B_s \frac{S_i}{C_i(t)}\right] \Big/ N \tag{10}$$

The process of route planning based on experience knowledge of taxi drivers is a process of seeking the minimum route planning arithmetic operator. $\overline{Dest(t)}$ is defined as the average road weight of all the trip, it can be obtained by calculate the average weight of all the road sections. At the same time, the process of route planning based on experience knowledge of taxi drivers is a process of seeking a route which owns the minimum average road weight of the found trip between origin and destination. When coding, we can establish the descriptive structure of the road point (RoadTopoPoint), the road topology (RoadTopology), and the road experience knowledge value (RoadExperience) as following:

```
Struct  RoadPoint
{
      long    RoadPointID;    /* the id of the road point */
      double   PointX;        /* the X of the road point */
      double   PointY;        /* the Y of the road point */
}
Struct  RoadTopology
{
      long      RoadID;             /* the id of the road */
      RoadTopoPoint  StarPoin       /* the star point of
                                       the road topology */
      RoadTopoPoint  EndPoint;      /* the end point of
                                       the road topology */
      long   RoadLength;            /* the length of the
                                          road  section */
}
Struct  RoadExperience
{
      long      RoadID;         /* the id of the road */
      long     AverageSpeed;      /* the average speed of
                                        the road */
      long     PassingTime;       /* the passing time of
                                       the road */
      int      ExpeLevel;         /* the experience level of
                                       the road section */
}
```

4. EXPERIMENT AND ANALYSIS

This paper adopts Wuhan City's traffic navigation digital map (scale is 1:1000) as basic experimental data. The road network contains 11,598 road sections and 8,471 nodes. The historical floating car data is the GPS sampling data of over 4000 taxies in December 2008. Sampling model is time dependant and the sampling time interval is 40 seconds. This paper chooses C#.net as developing tool and ArcGIS 9.2 as GIS platform to conduct route planning experiments. It finishes 350 times and with the index ratio B_s of road distance to experience level as 1 and α as 3. We compare the route planning algorithm based on TEKM with the classical shortest path algorithm, and the statistical analysis and comparison results is shown in Tab. 2. The results show that the route length which is calculated by TEKM algorithm increases by about 4% comparing to the shortest path algorithm, while the passing time could be reduced nearly one quarter (24%), and the road experience level increases 16%. The average road weight narrow gap reduced to 17%. That is to say that the route length which is planned by TEKM doesn't increase sharper than the route which is planned by the classical shortest path algorithm, while the travel time can be reduced to a large extent. At the same time, the route planned by TEKM can effectively avoid road level's dramatic changes comparing to the route planned by the shortest path algorithm, which make that the route maintains road level's consistency and coherence more reasonable and is more accordant with people travelling habits and recognition.

We randomly selected six discrete target points which are marked with stars (As shown in Fig.4). This experiment respectively adopted the traditional shortest path algorithm and TEKM route planning algorithm that is based on experience knowledge to plan routes between every two target points, and compare two routes obtained separately by two algorithms. The red path is the route calculated by route planning algorithm based on TEKM, and the green one is calculated by the traditional shortest path algorithm.

We obtained 15 routes which are accessible paths between two of the six target points that are planned by TEKM algorithm and the traditional shortest path algorithm, and compared four indicators: road length L, passing time T, average experience level $\overline{C(t)}$ which can be calculated by eq.4 and average road weight $\overline{Dest(t)}$, the results are shown in table 3.

Tab. 2 Comparing the TEKM algorithm with the shortest path algorithm

Increase range of route length	Decrease range of passing time	Increase range of road experience level	Decrease range of average road weight
4%	24%	16%	17%

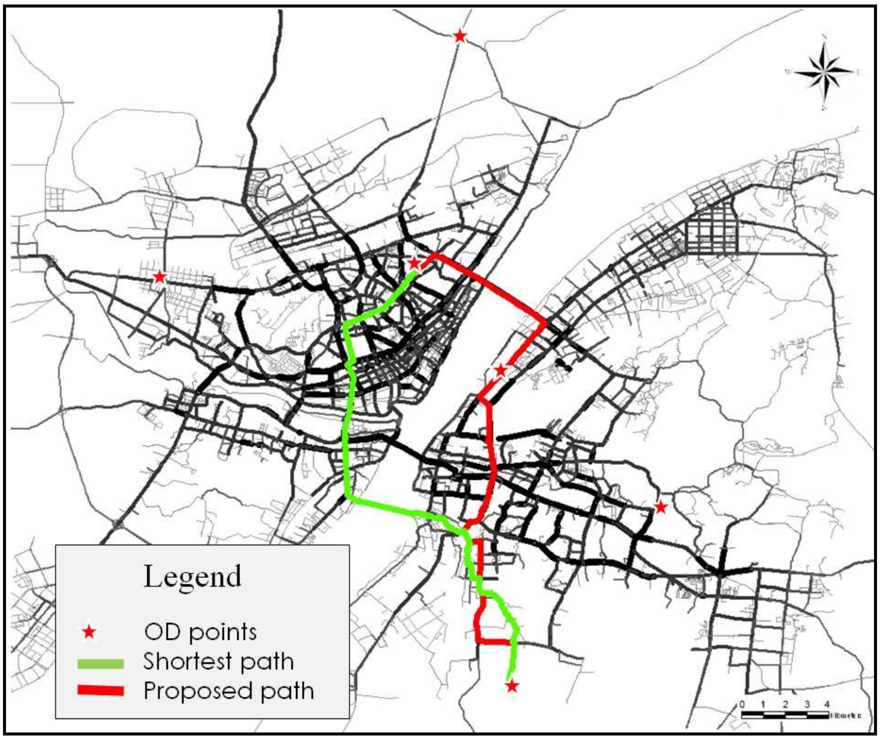

Fig. 4 Comparison of the routes with different algorithms

Tab. 3 Comparison of the 15 routes

序号	O→D	Shortest path algorithm				Proposed algorithm			
		L	T	$\overline{C(t)}$	$\overline{Dest(t)}$	L	T	$\overline{C(t)}$	$\overline{Dest(t)}$
1	1→2	9220.4	1297.0	3.7	1.6	9611.9	1004.2	4.3	1.3
2	1→3	17334.4	6569.4	4.1	2.2	18265.8	2122.5	4.8	1.3
3	1→4	9612.1	2197.0	4.4	1.5	10039.2	1685.1	4.5	1.4
4	1→5	10180.7	1650.1	3.5	1.7	10632.3	1170.9	3.9	1.4
5	1→6	21784.1	4079.9	3.6	1.7	22773.8	2825.7	4.6	1.3
6	2→3	9542.3	1665.4	4.6	1.3	9717.7	1345.2	4.9	1.2
7	2→4	17997.7	3203.3	4.1	1.5	19291.8	2147.3	4.3	1.4
8	2→5	17845.3	2230.9	3.8	1.5	18266.7	1939.6	4.2	1.3
9	2→6	13545.9	2997.9	3.3	1.8	14225.6	2048.3	4.7	1.3
10	3→4	21473.6	3373.4	4.2	1.5	22042.3	2665.0	4.5	1.3
11	3→5	25722.3	8210.3	4.1	2.0	26920.6	3057.9	4.6	1.3
12	3→6	9377.0	1442.5	4.1	1.5	9450.3	1186.0	4.2	1.3
13	4→5	17222.2	2785.9	2.8	2.1	17980.5	2028.2	4.1	1.5
14	4→6	24187.3	4409.4	3.3	1.8	26523.4	3380.2	4.4	1.4
15	5→6	30704.2	4812.5	3.8	1.6	31428.6	3761.0	4.5	1.3

5. CONCLUSION

It can obtain the taxis' experience and regulations of route planning by analysis the historical traffic data collected by floating cars, based on this principle, this paper proposes a taxis' experience knowledge model (TEKM) for route planning, and established a route planning algorithm based on TEKM. With the Wuhan city's traffic network and floating car data, this paper perform350 route planning experiments. The result shows that the route length which is planned by TEKM does not increase sharper than the route which is planned by the classical shortest path algorithm, while the travel time can be reduced to a large extent. At the same time, the route planned by TEKM can effectively avoid road level's dramatic changes comparing to the route planned by the shortest path algorithm, which make that the route maintains road level's consistency and coherence more reasonable and is more accordant with people travelling habits and recognition. Due to the limited information of floating car data, some factors which may influence model such as different region's occupational characteristics (for example, Wuchang District of Wuhan City is an education area, while Hankou District is a commercial district) and travel costs are not taken into account. So, these questions need further study.

6. ACKNOWLEDGEMENTS

The work descried in this paper was supported by the funding from the National Natural Science Foundation of China (No. 40801155, No. 40830530 and No. 60872132), and the projects funding of LIESMARS.

Reference

CHERKASSKY B V, GOLDBERG A V, RADZIK T. Shortest Paths Algorithms: Theory and Experimental Evaluation [J]. Mathematical Programming, 1996, 73: 129-174.

FISHER P. A Primer of Geographic Search Using Artificial Intelligence. Computers & Geosciences, 1990, 16(6): 753-776.

GONZALEZ H, HAN J, LI X. Adaptive Fastest Path Computation on a Road Network: A Traffic Mining Approach [A]. Proceedings of the 33rd international conference on Very large data bases table of contents [C], [s.n.], 2007.

HAN G., JIANG J, CHEN J. An Arc Based Dijkstra Algorithm for Road Turning Penalty in Vehicle Navigation System [J]. Acta Geodaetica et Cartographic Sinica, 2002, 31(4): 366-368.

LU F, LU D, CUI W. Time Shortest Path Algorithm for Restricted Searching Area in Transportation Networks [J]. Journal of Image and Graphics, 1999, (A): 849-853

LU F. Shortest Path Algorithms: Taxonomy and Advance in Research [J]. Acta Geodaetica et Cartographic Sinica, 2001, 30(3): 269-275.

MCQUEEN B, SCHUMAN R, CHEN K. Advanced Traveler Information Systems [M]. Boston: Artech House, Inc. 2002.

REN G, WANG W, DONG W. Shortest path problem with turn penalties and prohibitions and its solutions [J]. Journal of Southeast University (Natural Science Edition), 2004, 34(1): 104-108.

SUN Jinlin. Research on Vehicle Route Optimization Based on Probe Car GPS/GIS[D]. Beijing: Beijing Jiaoong University.

XIN F, CHEN X, LIN H. Research on time space distribution characteristics of floating car data in road network. China Journal of Highway and Transport, 2008，21（4）：105-110.

ZHENG N, LI Q, XU J. A Bidirectional Heuristic Shortest Path Algorithm with Turn Prohibitions and Delays. Geomatics and Information Science of Wuhan University, 2006, 31 (3) : 256-259.