Path-finding through flexible hierarchical road networks: An experiential approach using taxi trajectory data

Qingquan Li [a,b,**], Zhe Zeng [c,*], Tong Zhang [a,b], Jonathan Li [d], Zhongheng Wu [e]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei, PR China
[b] Transportation Research Center, Wuhan University, Wuhan, Hubei, PR China
[c] Department of Mapping and Geographic Information Engineering, College of Geo-Resources and Information, China University of Petroleum, Qingdao, Shandong, PR China
[d] Department of Geography & Environmental Management, Faculty of Environment, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1
[e] NavInfo Co. Ltd., HongYu Mansion, 7 Xueyuan Road, Haidian District, Beijing, PR China

ARTICLE INFO

ABSTRACT

Optimal paths computed by conventional path-planning algorithms are usually not "optimal" since realistic traffic information and local road network characteristics are not considered. We present a new experiential approach that computes optimal paths based on the experience of taxi drivers by mining a huge number of floating car trajectories. The approach consists of three steps. First, routes are recovered from original taxi trajectories. Second, an experiential road hierarchy is constructed using travel frequency and speed information for road segments. Third, experiential optimal paths are planned based on the experiential road hierarchy. Compared with conventional path-planning methods, the proposed method provides better experiential optimal path identification. Experiments demonstrate that the travel time is less for these experiential paths than for paths planned by conventional methods. Results obtained for a case study in the city of Wuhan, China, demonstrate that experiential optimal paths can be flexibly obtained in different time intervals, particularly during peak hours.

1. Introduction

Navigation systems are an important component of intelligent transportation systems and have become a standard device in vehicles, cell phones and other mobile devices. Many web-based mapping services also provide navigation tools for regular users. Path planning, a core component of various navigation applications, involves identification of the shortest path for any given origin–destination pair in a directed graph in which a non-negative weight is applied to the length or travel time for road segments. The Dijkstra (1959) algorithm and label correcting (LC) algorithm (Bellman, 1958) routing are two classical methods used to solve the shortest path problem. Variants of these algorithms have been extensively studied (Cherkassky et al., 1996; Thorup, 2004). Depending on whether the edge weights are static or dynamic, we can classify theoretical computation schemes for the shortest path into two categories. In recent years, researchers realized that preprocessing of road networks can significantly improve

the performance of the Dijkstra (LS) algorithm (Gutman, 2004; Goldberg and Harrelson, 2005; Kohler et al., 2006; Sanders and Schultes, 2006). Preprocessing is usually performed for road networks with static hierarchies. Computation for static networks can yield the exact shortest paths using Euclidean distance-based measurements. Computation for dynamic traffic conditions is a theoretically complicated operations research problem (Ahuja et al., 1993). The uncertainty of real traffic situations means that the best search results are not necessarily computed in reality. Finding the exact shortest path in road networks with dynamic traffic conditions (dynamic road networks) is a non-deterministic polynomial-time hard (NP-hard) problem (Ahuja et al.,2003). A few simplifying assumptions can be made, such as transformation of dynamic networks to time-dependent networks using the first in, first out (FIFO) condition. The shortest path in a FIFO network can be computed using the label algorithm (Kaufman and Smith, 1993). However, the theoretical assumptions do not completely hold in a real road network. Thus, there is a gap between theoretical algorithm research and real-world applications.

Taxi drivers usually disregard the planned routes computed by navigation systems. They tend to adopt their own "best" routes according to their driving experience. Very often, their solutions are more cost-effective (less travel time and lower costs) than the shortest paths identified by traditional path-planning algorithms. Their experience implicitly comprises their familiarity with local

* Corresponding author at: College of Geo-resouces and information, China University of Petroleum, Qingdao, Shandong, PR China.
Tel.: +86 532 86980304; fax: +86 532 86980304.
** Corresponding author. Tel.: +86 27 68778222; fax: +86 27 68778043.
E-mail addresses: qqli@whu.edu.cn (Q. Li), Zhe.Zeng.79@gmail.com (Z. Zeng).

road networks and real-world traffic conditions during different time intervals, such as peak and off-peak hours. This type of driving experience is accumulated from thousands of validated route choices made in daily life. Therefore, intuitively, this highly reliable driving experience, which is represented by a huge number of taxi trajectories, may support dynamic path-planning decision-making.

In this paper, a flexible road network hierarchy is constructed using the experience of taxi drivers to support path-planning. Construction of this hierarchy depends on the speed and frequency for roads traveled by taxi vehicles. Roads categorized as upper level are traveled at high frequency and high speed and have a higher probability of being selected than those at lower levels. Consequently, the experience of taxi drivers can be integrated into an experiential road hierarchy. Path-finding based on such a hierarchy represents an approximation of selection of an experiential route by taxi drivers. Empirical results show that our new approach can provide an experiential optimal path. The travel time is less than traditional algorithms applied in various off-the-shelf navigation systems. Our novel approach is a pilot study for effective integration of realistic traffic information in theoretical optimal path-planning.

The remainder of the paper is organized as follows. Section 2 describes the methodology for integrating taxi driver experience into optimal path computation. Section 3 presents the experimental results. Section 4 discusses our method compared to related work on the path-finding problem in road networks. Section 5 concludes the paper with a summary of our study and a discussion of future research directions.

2. Methodology

The knowledge of experienced drivers is rarely integrated in path-finding. Driver experience can be mined from the data generated by extensive floating car applications in transportation. In this section, we present a novel optimal path-finding approach using a dynamic road hierarchy based on taxi driver experience.

Fig. 1 presents a framework for the identification of experiential optimal paths based on floating car trajectories. First, preprocessing of road network data is performed to construct a dynamic hierarchical road network by incorporating taxi driver experience. Second, the optimal path is computed based on the experiential road network. The framework consists of three main procedures (rectangular boxes in Fig. 1): recovery of taxi routes, construction of the experiential road hierarchy, and computation of the experiential optimal path.

The first two procedures represent preprocessing of road network data. In China, floating cars are largely taxis. Therefore, floating car data mainly refer to taxi GPS trajectories. Taxi routes, which implicitly incorporate optimal paths chosen by experienced drivers, are recorded as GPS trajectories. Thus, floating car data can be used as input data for constructing an experiential road hierarchy because they reflect the prior knowledge of taxi drivers for optimal path selection. Experiential routes are first recovered from original GPS coordinates. The original floating car data are tracking points recorded by a GPS receiver mounted in a taxi. A map-matching algorithm must be executed to assign original GPS coordinates to corresponding road segments since the original GPS data are not accurate enough to fall into correct road segments. Then a set of taxi routes can be recovered by connecting these road segments together. If the sampling rate is very low and road segments cannot be connected in a complete sequence, additional road segments should be interpolated to construct a continuous route in the road network. A set of experiential routes are therefore obtained from a huge number of recovered taxi trajectories. Second, an experiential road hierarchy for the network is constructed using recovered experiential routes. The more frequently that road seg-
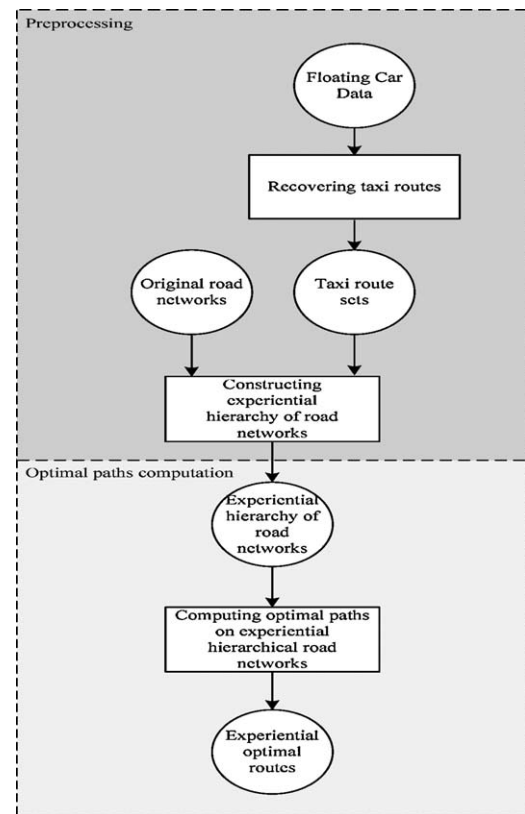

Fig. 1. Framework for the identification of experiential optimal paths. Oval boxes represent data. Rectangular boxes represent computation processes or functions. The preprocessing step incorporates taxi driver experience into the road network hierarchy so that the path computation step can produce experiential optimal paths.

ments are traveled as part of the experiential routes, the more likely it is that the segments are assigned to the upper level road category. An experiential road hierarchy can then be constructed statistically. Each level must be strongly connected so that an optimal path can be computed between each pair of nodes in the road network. Using this experiential hierarchy, experiential path computation can be performed using a bidirectional shortest path search algorithm. We describe the three procedures in the following sub-sections.

2.1. Recovery of taxi routes from GPS trajectories

Experiential taxi routes are recovered to obtain optimal routes chosen by taxi drivers from their GPS trajectories. A recovered route starts at the location where a passenger enters the taxi and ends when they leave, as recorded in the GPS trajectories. First, original tracking points are map-matched to corresponding road segments. We used the similarity metric of Greenfeld (2002) to determine the road segment to which a tracking point belongs. If GPS tracking data are sparse, an additional interpolation step is performed between two neighboring road segments that are not topologically connected in the road network. Assume that point p_i' is the match point for the original tracking point p_i at time t_i. Path pth is the route between match points p_i' and p_{i+1}' in a sequence of time. The path length is the sum of the length from point p_i' to p_{i+1}' along the route in the network (Fig. 2). The lengths of all paths between p_i' and p_{i+1}' are computed and compared to one estimated by multiplying the mean velocity of the two GPS track points by their time interval. The path whose length is closest to the estimated length is chosen as a recovered trajectory segment. Road segments between the two segments that match points belong to are interpolated into the recovery routes. Thus, through interpolation, complete taxi routes
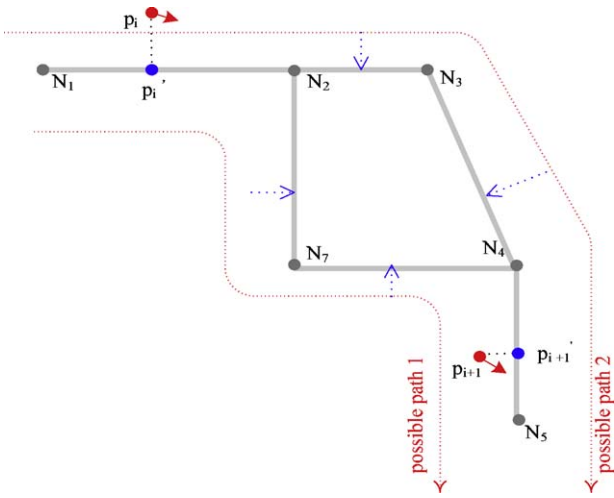
Fig. 2. Interpolation of GPS paths.

can be recovered through sequential matching of road segments in time series.

2.2. Construction of the experiential road hierarchy

Given a directed graph $G = \{N, E\}$, N is a set of intersections of roads and E is a set of road segments between two neighboring junctions. A continuous recovery route in a graph G is denoted as:

$$\xi = \langle e_1, \ldots, e_j, \ldots, e_m \rangle \, (e_j \in E, \, 1 \leq j \leq m), \tag{1}$$

where j $(1 \leq j \leq m)$ is a time index from the start point to the end point of ξ in chronological order. ξ contains no cycles and is essentially a set of acyclic edges (i.e. ξ satisfies the condition $e_i \neq e_j$, $\forall e_i$, $\forall e_j \in \xi$ and $i \neq j$).

An experiential route can be determined using speed information extracted from a huge number of recovered taxi routes. A route with higher speed implies that this is preferable for driving. Conceptually, a route can be associated with edges of G. Therefore, for a large number of routes, experiential routes statistically reflect a driver's choice of route. An experiential road hierarchy can then be constructed based on the preference information. The hierarchy is flexible for different time intervals, since traffic status continually changes with time. Each level of the hierarchy should be strongly connected to ensure topological completeness for optimal path computation. There are two steps for road hierarchy construction: building of experiential route sets and construction of the experiential road hierarchy.

2.2.1. Building of experiential route sets

Let $l: E \to R^+$ be a length function in graph G. $l(e)$ represents the length of edge e. Assume that p' is a point matching GPS point p on matched edge $e = \{v_{head}, v_{tail}\}$. $l_f(e)$ is the length of the segment from p' to v_{tail} on e, and $l_b(e)$ is the length of the segment from v_{head} to p' on e. The length of a recovery route ξ can be computed as:

$$l_\xi = l_f(e_1) + \sum_{i=2}^{m-1} l(e_i) + l_b(e_m). \tag{2}$$

The length of the first edge is computed using the function $l_f(.)$, since e_1 is split by its matching point. Similarly, the length of the last edge is computed using the function $l_b(.)$ because it is also broken at the last matching point. The speed for route ξ is given by:

$$v_\xi = \frac{l_\xi}{\delta t}, \quad (\delta t : t_m - t_1), \tag{3}$$

where t_1 and t_m denote the start and end times for route ξ, respectively. Conceptually, v_ξ is a critical factor in choosing preferred routes. Empirically, taxi drivers select a route that they can drive at higher speed. Therefore, we can determine whether or not a route is empirically preferred from its associated speed. Let v_T be the speed threshold, which can be determined by statistical analysis of all trajectories (see Section 3). An experiential route set is given by:

$$S_{ER} = \{\xi | v_\xi \geq v_T\}. \tag{4}$$

Time should be constrained to the time interval $\Delta t_{interval}$, so an experiential route set in $\Delta t_{interval}$ is given by:

$$S_{ER}^{\Delta} = \{\xi | v_\xi \geq v_T, \quad \text{and} \quad t_1, t_m \in \Delta t_{interval}\}, \tag{5}$$

where $\Delta t_{interval}$ is a time interval (t_{begin}, t_{end}), and t_1 and t_m are as defined above. Given any time interval, a flexible experiential route set can be built.

2.2.2. Construction of the experiential road hierarchy

All roads are categorized as heavily used, frequently used or rarely used roads, according to driving preference information derived from the frequency for edge e in all experiential routes in S_{ER}^{Δ}. An experiential road hierarchy for the network can be constructed based on classification of the travel frequency for edge e, provided that all road edges at each level are strongly connected.

An interesting finding on road hierarchy was reported by Jiang (2007) topological, who demonstrated that the topological character of urban streets conforms to the 80/20 principle, whereby a 20% of streets account for 80% of traffic flow, implying that city road networks are inherently hierarchically organized (Jiang, 2009). Inspired by these findings, we used the quantile as a statistical reasoning tool to determine the driving preference of taxi drivers. Each route in the set S_{ER}^{Δ} is associated with corresponding edges in G. Driving preferences during specific time intervals, such as peak hours, can be derived from travel frequency quantiles for edges for all experiential routes. Travel frequency is thus a criterion for extracting upper level experiential roads from graph G.

Let function $c(e)$ be the travel frequency for edge $e \in E$, determined by counting the number of routes that pass through e in all trajectories of S_{ER}^{Δ}.

$$p(g) = P\{c(e) = g\}, \quad (g \geq 0) \tag{6}$$

denotes the probability distribution for travel frequency in edge set E. Let $N(g)$ be the number of edges that satisfy $c(e) = g$. Let m be the total number of edges in E. Then the above formula can be rewritten as:

$$p(g) = \frac{N(g)}{m}, \quad (g \geq 0). \tag{7}$$

Let $F_n(q)$ be the cumulative distribution function (cdf) of g, i.e. $c(e)$ (where q denotes the variable for integration).

$$F_n(q) = \int_0^q p(g) dg, \quad (g \geq 0), \tag{8}$$

where $\int_0^{+\infty} p(g) dg = 1$, $(g \geq 0)$. We can statistically use F_n to derive the driving preference for the roads (Fig. 3). Let q_1, q_2 $(0 < q_2 < q_1 < 1)$ be two quantile thresholds. Two corresponding g values (g_1 and g_2) can be derived given q_1 and q_2. Based on the two thresholds, all roads can be classified into three levels according to driving preference. Heavily used roads are edges with $c(e)$ values g above the threshold g_1. Frequently used roads have $c(e)$ values between g_1 and g_2, and rarely used roads have $c(e)$ values below g_2. We can extract two upper levels in terms of $c(e)$ values from graph G. The road segments at these two levels should be strongly connected for hierarchical route planning. Therefore, extraction of upper level roads and forcing of strong connectedness are the two steps for
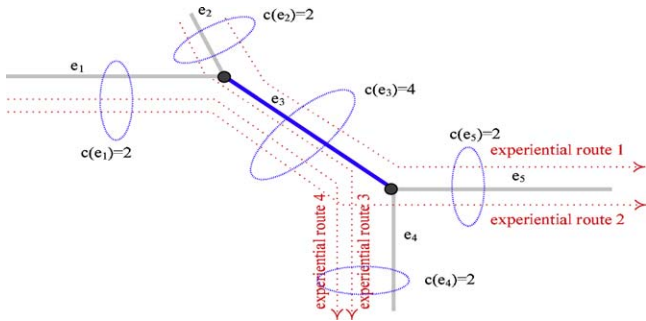
Fig. 3. Extraction of the most-traveled roads to build upper level road sets. There are four experiential routes in the figure, represented by dashed lines. Four routes pass edge e_3, so $c(e_3)$ has the greatest travel frequency. Thus, edge e_3 is the 0.8 quantile of the cdf F_n. Assuming that $q_1 = 0.8$, edge e_3 should belong to the upper level of the experiential road hierarchy.

construction of an experiential hierarchical graph for a road network.

Assume that the hierarchical graph:

$$H = \{G, H^0, H^1, H^2\} \tag{9}$$

is a partition of graph $G = \{V, E\}$ based on frequency $c(e)$ for a road. The two steps are described as follows (Fig. 4):

Step 1. *Extraction of upper level roads.* For each edge $e \in E$, we compare $c(e)$ with thresholds g_1 and g_2. The edge set E for the road network is grouped into three subsets, heavily used set $E_{he} = \{e \mid c(e) \geq g_1\}$, frequently used set $E_{fr} = \{e \mid g_1 > c(e) \geq g_2\}$ and rarely used set $E_{ra} = \{e \mid g_2 > c(e) \geq 0\}$, where $E = E_{he} \cup E_{fr} \cup E_{ra}$ and $E_{he} \cap E_{fr} \cap E_{ra} = \varnothing$. The three subsets of E constitute a road hierarchy for the network. The bottom level is $H_0 = G$. Level H^1 is constructed as $E_{he} \cup E_{fr}$, where $H^1 = \{V^1, E^1\}$, $E^1 = E_{he} \cup E_{fr}$, and $V^1 = \{v \mid v$ is the head or tail vertex of $e^1 \in E^1\}$. This ensures that edge $e^1 \in E^1$ and its two vertexes are also in H^1. Similarly, top level H^2 is constructed as E_{he}, where $H^2 = \{V^2, E^2\}$, $E^2 = E_{he}$, and $V^2 = \{v \mid v$ is the head or tail vertex of $e^2 \in E^2\}$. The subgraph $H^i = \{V^i, E^i\}$ is level i of the hierarchy graph H and satisfies $G = H^0 \supset \ldots \supset H^i \supset \ldots \supset H^r$ for $0 \leq i \leq r$ (in this case, $r = 3$). We attach the node associations of neighboring levels, which record the duplicate copy of the node in neighboring upper level H^{i+1}, to a node $v \in V^i$ in level H^i. The association between levels i and $i+1$ is established so that hierarchical route planning can be performed in the road network.

Step 2. *Forcing of strong connectedness.* Let C^i be a maximal connected subgraph of level H^i constructed according to step 1. The vertexes and edges that do not belong to C^i in level H^i are deleted. C^2 is a maximal strongly connected component (SCC) of H^2. E^2/E_{C^2} and V^2/V_{C^2} are first deleted from H^2. Then C^2 is a new H^2. The same operation is performed on H^1. After deletion of these vertexes and edges that do not belong to C^i, the node association between neighboring levels should be upgraded because associated edge sets and vertex sets have changed.
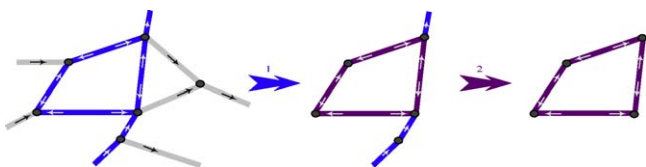

Fig. 4. Construction of the experiential road hierarchy. 1. Extraction of upper level roads based on travel frequency. 2. Forcing strong connectedness for extracted roads.
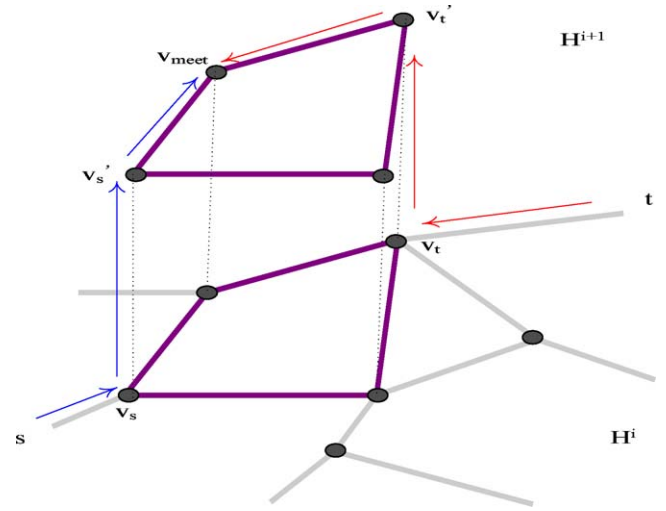

Fig. 5. Hierarchical route-planning. H^i and H^{i+1} are two neighboring levels. s,t are the source and target vertexes. v'_s is a duplicate of v_s in the upper level H^{i+1} and v'_t is a duplicate of v_t in the upper level H_{i+1}. The hierarchical search terminates at v_{meet} in level H_{i+1}.

2.3. Experiential optimal path planning

We extend the paradigm of Car and Frank (1994) by incorporating taxi driver experience. A taxi driver identifies optimal paths based on hierarchical reasoning. Roads that are frequently traveled at high speed are more likely to be selected first. An experiential hierarchy H for the road network $G = (V, E)$ is constructed by merging experiential information recovered from taxi trajectories with the actual road network. An experiential optimal path can then be computed by applying conventional LS algorithms to H, which consists of levels H^0, H^1 and H^2. Each node $v \in V^i$ is extracted from graph G for each level $H^i = \{V^i, E^i\}$, so v in the original graph has duplicate copies for the associated levels. The same procedure is applied for edge set E^i. Two duplicate vertexes in the two neighboring levels H^i and H^{i+1} should be associated via an association mapping:

$$f_a : v^i \to v^{i+1} \tag{10}$$

for $v_i \in V^i$. Accordingly, when the path computation algorithm finds a vertex that has an association mapping record, it continues to search in the upper level starting from the duplicate for this vertex. Reversing all edges of G, we can denote $G_r = \{V, E_r\}$ as the reverse graph of directed graph G, where $E_r = \{(u, v) \mid (v, u) \in E\}$ and $l((v, u)) = l((u, v))$. We thereby construct a reverse

$$H_r = \{G_r, H_r^0, H_r^1, H_r^2\} \tag{11}$$

of the experiential road hierarchy H.

We used bidirectional search methods (Nicholson, 1966) and hierarchical reasoning to compute the experiential optimal path. The forward search is performed in H and the backward search is simultaneously performed in H_r. Each level is strongly connected, so that the route planning can be performed via the following steps based on H. Let s be a source point and t be a target point (Fig. 5):

(1) The forward search starts from s and the backward search from t in level H^i until the nearest node v_s with an associated mapping record for the upper level H^{i+1} and the nearest node v_t with an associated mapping record for H_r^{i+1}, respectively, are identified.
(2) The bidirectional search subsequently continues from v'_s and v'_t, which are the duplicate nodes for v_s and v_t, regarded as a new pair of source and target nodes in the upper level H^{i+1}.
(3) If the forward search meets the backward search at a node v_{meet}, the search process is terminated.

Table 1
Road class codes.

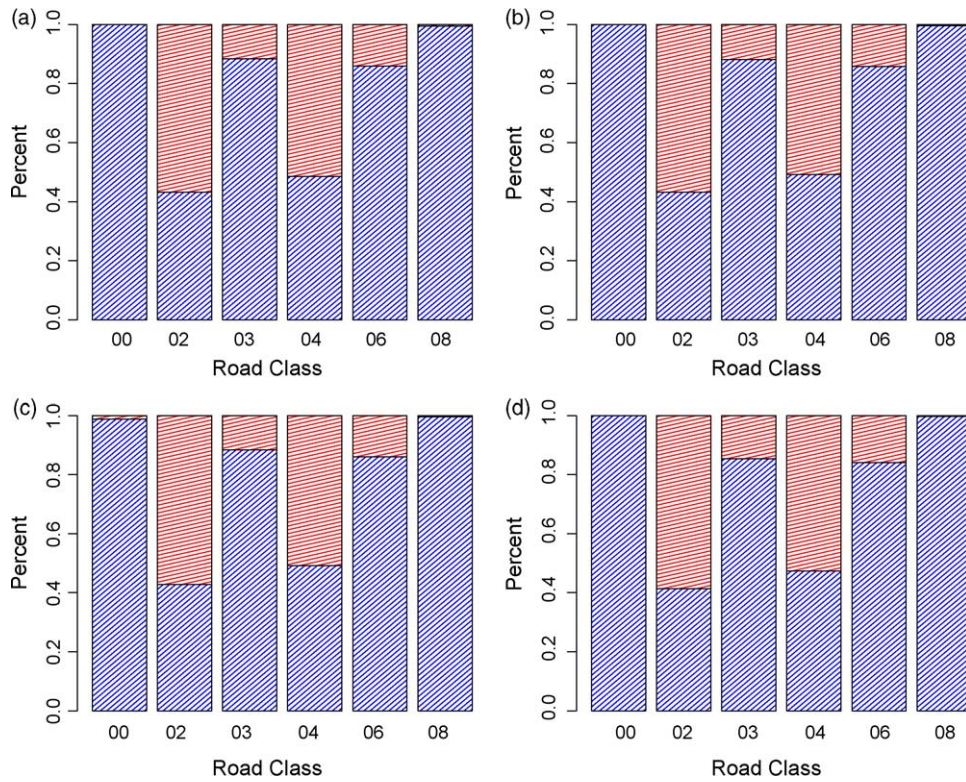| Code	00	02	03	04	06	08
Road class	Freeway	City expressway	State road	Urban major arterials	Urban minor arterials	Local streets
Number	325	1250	1486	1547	19,482	21,096

Fig. 6. Comparison of pre-defined road classes and the experiential road hierarchy. The *y*-axis denotes the percentage of experiential levels for each road class and the *x*-axis represents road class codes. Red shading represents frequently used roads and blue shading denotes rarely used roads. (a) 7:00–9:00; (b) 10:00–12:00; (c) 17:00–19:00 and (d) 20:00–22:00. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)
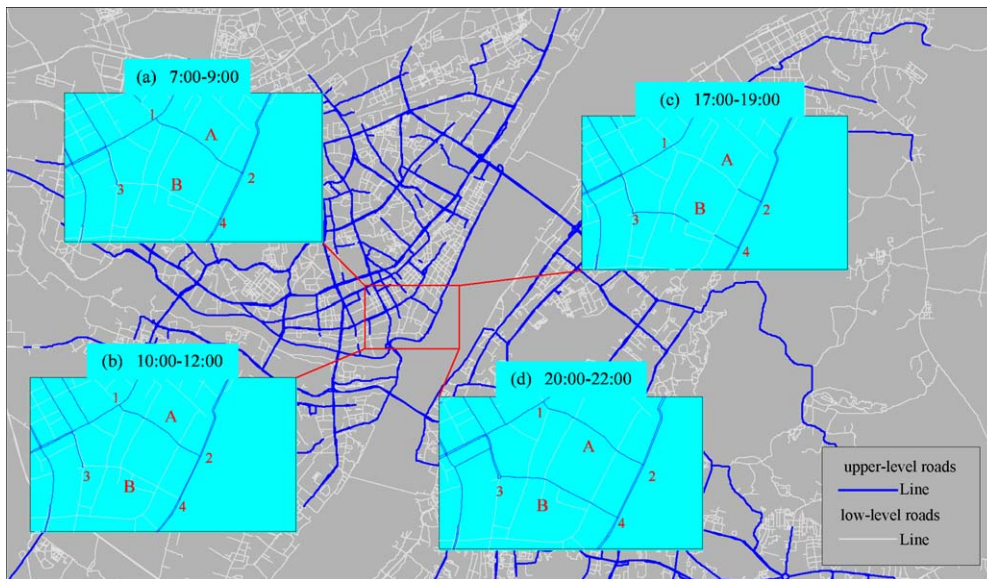

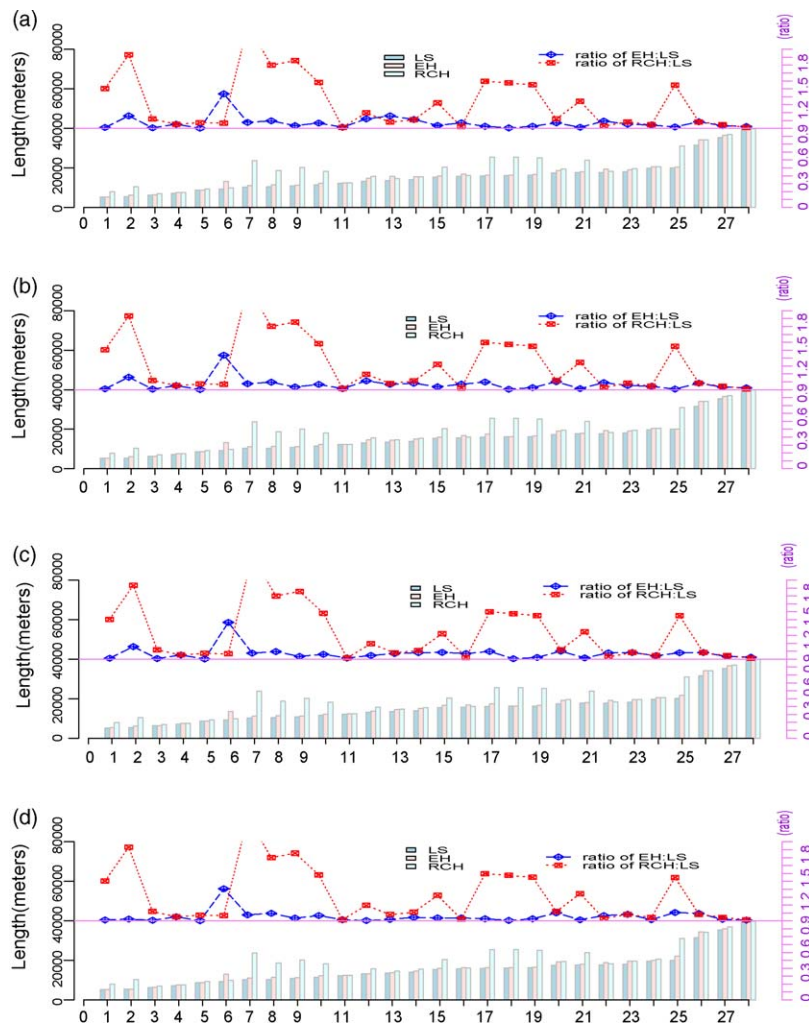**Fig. 7.** Flexible experiential road hierarchy for four time intervals.

Fig. 8. Route length comparison for the three algorithms. The bar height represents route length. The right-hand vertical axis denotes the length ratio for two algorithms for the 28 experiments. (a) 7:00–9:00; (b) 10:00–12:00; (c) 17:00–19:00 and (d) 20:00–22:00.

(4) If the search finds nodes with associated mapping records for the upper level H^{i+1}, steps 1 and 2 are repeated until the search reaches the highest level.

3. Experimental results

We collected real-world taxi trajectories recorded by GPS receivers in 12,314 taxis during 1 week in the city of Wuhan, China. The GPS sampling rate was one signal every 40 s. The road network of Wuhan is represented by a graph with 45,186 edges and 20,950 nodes. We recovered taxi routes in the graph using the method in Section 2.1. We grouped all recovered routes into four sets by time intervals: 07:00–09:00 h (morning peak hours), 10:00–12:00 h (morning off-peak hours), 17:00–19:00 h (afternoon peak hours) and 20:00–22:00 h (evening peak hours). A taxi journey with start and end times within a specified time interval belonged to the corresponding route set for that time interval. We built experiential routes in the four sets according to Eq. (5). All GPS tracking points for original trajectories in the four time intervals were used to statistically compute the speed thresholds using mean speed data.

3.1. Comparison of the experiential road hierarchy and pre-defined road classes

We constructed an experiential road hierarchy for the Wuhan road network using the method in Section 2.2. Only two levels

were built since the Wuhan road network is relatively simple. We set $q_1 = 0.8$. All roads were categorized as frequently used or rarely used. Edges with a cdf (Eq. (8)) greater than q_1 were classes as frequently used roads, and the remaining edges as rarely used roads. Therefore, frequently used roads were considered as upper level roads and all roads as low-level roads. There are two main differences between the pre-defined road classes (Table 1) and our experiential hierarchy.

First, the distribution of upper level roads differs between the experiential hierarchy and the pre-defined road classes. Fig. 6 shows the differences between the pre-defined road classes and the experiential hierarchy for the four time intervals. Table 1 lists pre-defined road class codes specified by map data suppliers. The distribution of the upper level roads in our experiential road hierarchy is mainly centered on city expressways, state roads, urban minor arterials and local streets. Nearly 50% of the city expressways and urban main arterials are extracted as upper level roads in the experiential hierarchy. The percentage of upper level roads in class code 06 (urban minor arterials) is less than 20%, even though the number of experiential upper level roads in class code 06 is very large.

Second, the pre-defined road class hierarchy is static, whereas the hierarchy based on taxi driver experience is flexible. In Fig. 6, the experiential road hierarchy distribution is similar for the four time intervals, although slight differences exist. The differences result
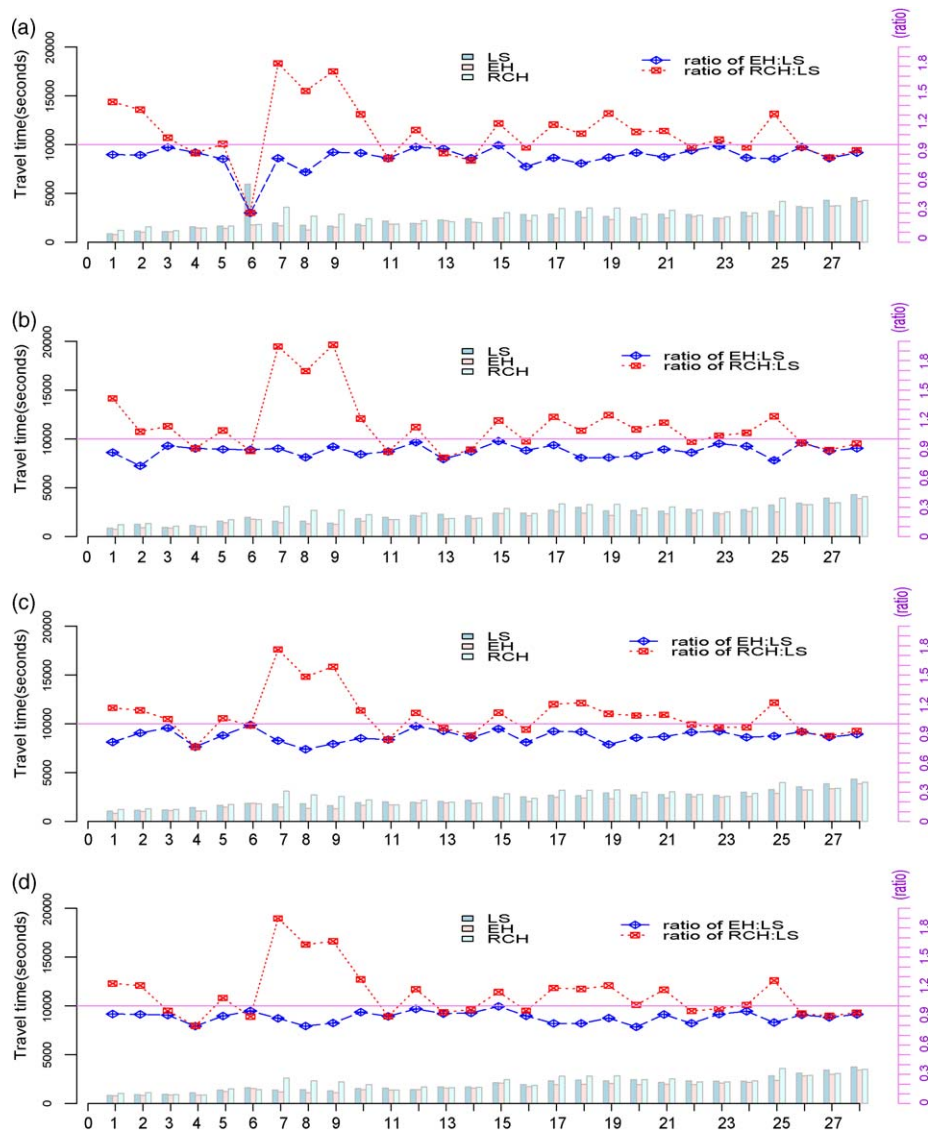
Fig. 9. Route travel time comparison for three algorithms, analogous to Fig. 8. (a) 7:00–9:00; (b) 10:00–12:00; (c) 17:00–19:00 and (d) 20:00–22:00.

from flexible experiential route sets that reflect the preference of taxi drivers in different time intervals. Fig. 7 presents an example. Road A contains a few segments between intersections 1 and 2 and road B contains a few segments between intersections 3 and 4. Road A is an experiential upper level road but road B is not during morning peak and off-peak hours. Roads A and B are not completely connected in the experiential upper level road network during evening peak hours, but are during evening off-peak hours. Roads in the experiential upper level have higher priority and are more likely to be chosen for computing optimal paths. Road B, which does not belong to the experiential upper level roads during morning peak and off-peak hours, might not be chosen by taxi drivers at these times. Roads A and B lose their priority during evening peak hours. However, they recover their priority during evening off-peak hours. These changes in upper level hierarchy reflect differences in driving preferences of taxi drivers with knowledge of realistic traffic conditions on roads A and B. This example demonstrates that the experiential road hierarchy is flexible in that road priority can be adjusted for different time intervals according to a driver's experience.

3.2. Comparison of experiential path-finding and two traditional algorithms

We randomly chose 28 origin–destination (OD) pairs to implement three algorithms: the Dijkstra (label setting) algorithm (LS) using the original road network, the route-planning algorithm for the pre-defined road class hierarchy (RCH), and the route-planning algorithm for our experiential road hierarchy (EH). The RCH algorithm based on pre-defined road class originated from Car and Frank (1994). The road class hierarchy is as defined by NavInfo Inc., a leading map data supplier for navigation systems in China. The road class hierarchy comprises two levels: the upper level includes freeways, city expressways, state roads and urban major arterials; the lower level consists of all road classes. Path computation based on road class hierarchy is widely used in various navigation systems. The EH algorithm is based on the hierarchy constructed by the method in Section 2.2. The weight for path-finding computation is the length of road segments. The travel time for each road segment was obtained using floating car data collected over a week. The path-finding results are compared
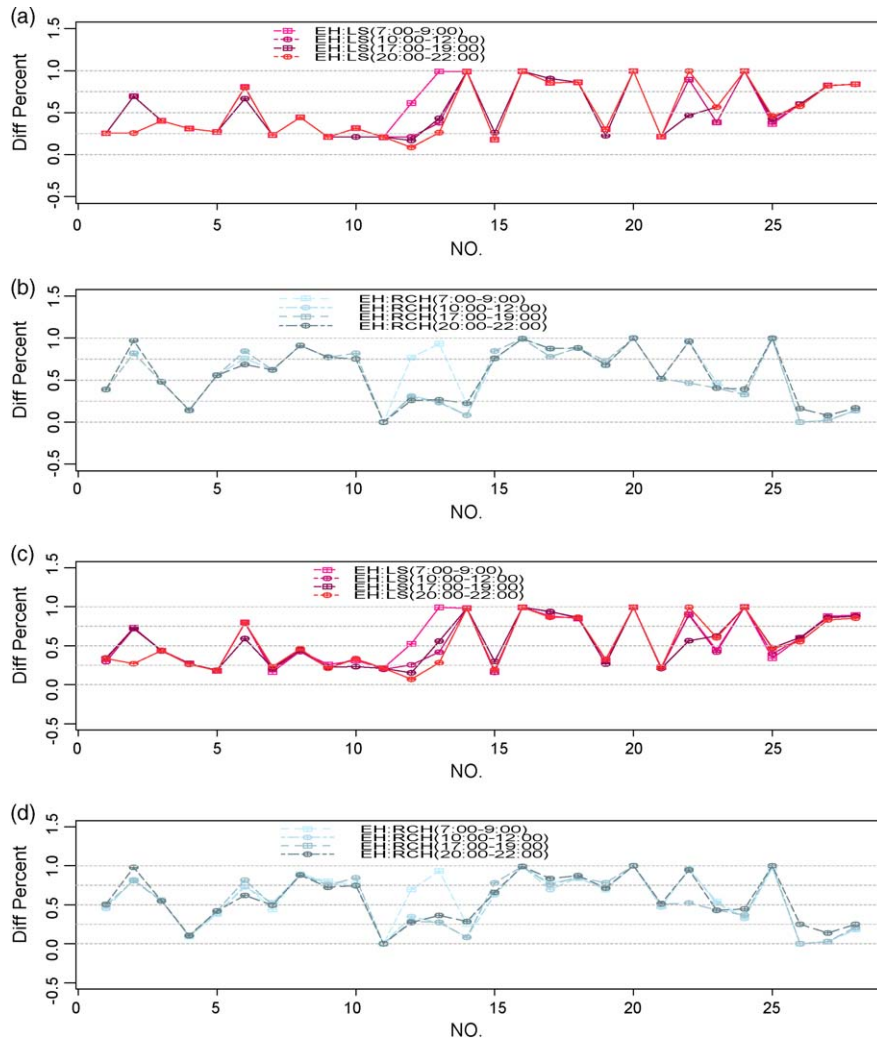
Fig. 10. Degree of difference between the EH results and the results for the RCH and LS algorithms. (a) Degree of length difference between EH and LS routes; (b) degree of length difference between EH and RCH routes; (c) degree of travel time difference between EH and LS routes; (d) degree of travel time difference between EH and RCH routes.

for the algorithms in Figs. 8–10. The 28 OD pairs are numbered by Dijkstra distance in ascending order.

Fig. 8 shows length differences between the three algorithms. The bars represents route lengths and connected dots depict the EH/LS and RCH/LS ratios. The LS and RCH lengths for the four time intervals are all identical since they do not take into account real traffic information. Thus, the RCH/LS ratio does not change. However, the EH lengths change during four time intervals because path-finding is based on a flexible experiential road hierarchy. The RCH route lengths exceed the shortest length more significantly than those for the EH approach. All the EH/LS ratios are less than 1.3 apart from one (1.5). The EH path lengths are thus very close to the shortest length computed by the LS algorithm. Experiments demonstrate that the EH algorithm can yield flexible paths that are only slightly longer than the shortest paths for different time intervals, as evidenced by a more stable EH/LS ratio compared to RCH/LS.

Fig. 9 shows travel time differences between the three algorithms. The EH algorithm yields optimal routes with the shortest travel time among all the algorithms for all four time intervals. All but one RCH routes have a longer travel time than for the shortest path. The EH/LS ratios for all 28 experiments are below 1, demonstrating that the flexible EH approach can identify paths that are optimal in terms of travel time for different time intervals. However, the RCH and LS algorithms do not have this flexibility.

Fig. 10 shows the degree of difference for paths computed by the EH algorithm and the other two algorithms. Assume that paths A and B are two results computed by different algorithms. Let $l(A)$ be the sum of weights (length or travel time) for all road segments on path A. Let $l_{A:B}(A)$ be the sum of weights for the parts on path A that differ from path B. $l_{A:B}(A)/l(A)$ denotes the degree of difference between paths A and B. The EH algorithm is very flexible for computing paths that vary by length or travel time in response to different planning ODs and time intervals.

Fig. 10(a) shows degree of difference in path length for the EH and LS algorithms. The index numbers are ranked by Dijkstra distance in ascending order. The LS algorithm is static, so that lengths computed are the same in different time intervals. The results reveal that the EH paths vary with time. The degree of difference is larger than 0.5 for 10 paths for the four time intervals (index numbers 6, 14, 16–18, 20, 24, 26, 27, and 28). Eight of these paths are relatively long (longer than the median route length). For these 10 paths, more than half of their segments (measured by length) do not overlap with the corresponding shortest path. The degree of difference is less than 0.5 for 13 paths (index numbers 1, 3, 4, 5, 7–11, 15, 19, 21, 25). Nine of these paths are relatively long (longer than the median length). For these 13 paths, more than half of their segments are identical to those of the corresponding shortest path. The results imply that a long path is more likely deviate from the shortest path for more than half its length and that a short path is

more likely to overlap with the shortest path for more than half of its segments.

Fig. 10(b) shows the degree of difference in path length for the EH and RCH algorithms. The degree of difference is larger than 0.5 for 15 paths, indicating that a majority of the EH paths do not overlap the RCH paths.

Fig. 10(c) and (d) shows the degree of difference in path travel time. The trends are similar to those for path length differences.

In Fig. 10(a) and (c), the degree of difference is larger than 0.5 for the three long-path experiments (nos. 26–28), which are longer than 30 km, even for the shortest path. In Fig. 10(b) and (d), the degree of difference is less than 0.5 for these paths. This indicates that for long paths, the EH algorithm yields greater similarity to paths computed by the RCH algorithm than to paths from the LS algorithm.

4. Discussions

The shortest path problem (SPP) has been extensively investigated by researchers working in various disciplines (e.g., computer science, transportation engineering, and geographical information science). We introduce a novel experiential path-finding method based on road hierarchies derived from the experience of taxi drivers. Experiential hierarchies are constructed by mining taxi GPS trajectories. Roads belonging to the upper level are given high priority when a route is planned. Compared to current methods, this approach uses a different perspective for solving the SPP in road networks. Experiments demonstrate that our methodology can produce highly dynamic, experientially optimal and computationally efficient paths for navigation applications.

Most existing preprocessing methods (Car and Frank, 1994; Jing et al., 1998; Jung and Pramanik, 2002; Sanders and Schultes, 2006; Li et al., 2008) are usually performed on static road networks. However, traffic conditions for specific road networks can change significantly if real-time traffic information is taken into account. Updating of road network hierarchies is difficult for on-line optimal path computation because preprocessing is generally a time-consuming procedure. In our approach we use the experience of taxi drivers and thus different optimal paths can be obtained for peak and off-peak hours in response to varied traffic conditions according to the flexible experiential road hierarchy.

The labeling algorithms (Dijkstra, 1959; Bellman, 1958) routing and their variants (Cherkassky et al., 1996; Thorup, 2004) compute an exact minimum cost-path between a source and a target. However, the exact shortest path in metric space is usually not an optimal one in reality. There is a difference between the exact shortest path and an experiential optimal path, which our method uses. The experiential method results approximate to those for the Dijkstra algorithm in terms of length. Furthermore, the travel time is less for the experiential paths than for the corresponding shortest paths. Road hierarchy is used to improve performance, as in current hierarchy methods Car and Frank (1994) and Li et al. (2008). The hierarchy of our method is derived from the experience of taxi drivers and reflects priorities in selecting routes at different times. Therefore, we can compute an experiential optimal path with performance similar to that of traditional hierarchical methods.

All these advantages of our approach are illustrated by Fig. 11. The EH length varies for the different time intervals, but the length for the other two algorithms remains constant. The EH travel time is less than for the shortest paths and the path length is approximately similar. Furthermore, the running time and number of nodes for the EH approach are close to those for the RCH algorithm. Thus, the performance of our method is close to that of hierarchical route-planning since we use a road hierarchy derived from taxi driver experience.
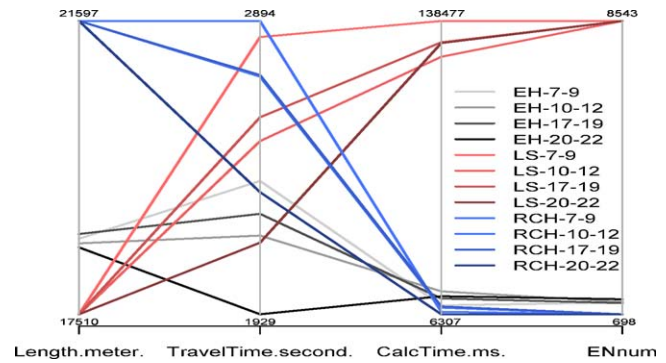

Fig. 11. Parallel coordinates plot based on four criteria: travel time, length, computing time and expanded node numbers. The four criteria are compared for LS, EH and RCH methods using arithmetic mean values for 28 experiments.

5. Conclusion

Taxi driver experience is rarely used in route planning in current navigation systems. We have proposed a new method that uses local taxi driver experience to compute optimal paths. Taxi GPS tracking points are first used to recover original taxi routes with map-matching and interpolating techniques. These taxi routes that reflect driver experience are used to construct a time-variant flexible road hierarchy. Based on this road hierarchy, an optimal path-finding algorithm is introduced. It can dynamically provide various path-finding results during different time intervals (such as peak and off-peak hours). Experimental results indicate that an experiential road hierarchy can be constructed for different time intervals and that our experiential approach for optimal path-finding can provide faster and shorter paths in comparison with two traditional algorithms. Future research may focus on a more extensive evaluation of our approach using floating car data from other cities and more accurate computation of travel speed.

References

Ahuja, R.K., Magnanti, T.L., Orlin, J.B., 1993. Network Flows: Theory, Algorithms and Applications. Prentice Hall, New Jersey.
Ahuja, R.K., Orlin, J.B., Pallottino, S., Scutella, M.G., 2003. Dynamic shortest paths minimizing travel times and costs. Networks 41 (4), 197–205.
Bellman, R., 1958. On a routing problem. Quarterly of Applied Mathematics 16, 87–90.
Car, A., Frank, A.U., 1994. Modelling a hierarchy of space applied to large road networks. In: Lecture Notes in Computer Science:IGIS'94, vol. 884/1994. Springer, Berlin, pp. 15–24.
Cherkassky, B., Goldberg, A., Radzik, T., 1996. Shortest paths algorithms: theory and experimental evaluation. Mathematical Programming 73 (2), 129–174.
Dijkstra, E., 1959. A note on two problems in connexion with graphs. Numerische Mathematik, 269–271.
Goldberg, A., Harrelson, C., 2005. Computing the shortest path: a search meets graph theory. In: Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics Philadelphia, PA, USA, pp. 156–165.
Greenfeld, J., 2002. Matching gps observations to locations on a digital map. In: Proceedings of 81st Annual Meeting of the Transportation Research Board.
Gutman, R., 2004. Reach-based routing: a new approach to shortest path algorithms optimized for road networks. In: Proceedings of the 6th Workshop on Algorithm Engineering and Experiments, SIAM, pp. 100–111.

Jiang, B., 2007. A topological pattern of urban street networks: Universality and peculiarity. Physica A: Statistical Mechanics and its Applications 384 (2), 647–655.

Jiang, B., 2009. Street hierarchies: a minority of streets account for a majority of traffic flow. International Journal of Geographical Information Science 23 (8), 1033–1048.

Jing, N., Huang, Y.-W., Rundensteiner, E.A., 1998. Hierarchical encoded path views for path query processing: an optimal model and its performance evaluation. IEEE Transactions on Knowledge and Data Engineering 10 (3), 409–432.

Jung, S., Pramanik, S., 2002. An efficient path computation model for hierarchically structured topographical road maps. IEEE Transactions on Knowledge and Data Engineering 14 (5), 1029–1046.

Kaufman, D., Smith, R., 1993. Fastest paths in time-dependent networks for intelligent vehicle-highway systems application. Journal of Intelligent Transportation Systems 1 (1), 1–11.

Kohler, E., Möhring, R., Schilling, H., 2006. Fast point-to-point shortest path computations with arc-flags. In: Demetrescu, C., Goldberg, A.V., Johnson, D.S. (Eds.), 9th DIMACS Implementation Challenge – Shortest Paths, pp. 41–72.

Li, H., Sheng, X., Li, G., 2008. Principles of adjusting road hierarchy for route planning in Navigation maps. In: Proceedings of 7th Annual Meeting of the China Technical Application Association for Global Positioning System, vol. 1, Beijing, pp. 209–214 (in Chinese).

Nicholson, T.A.J., 1966. Finding the shortest route between two points in a network. The Computer Journal 9 (3), 275–280.

Sanders, P., Schultes, D., 2006. Engineering highway hierarchies. In: ESA'06: Proceedings of the 14th conference on Annual European Symposium. Springer-Verlag, London, UK, pp. 804–816.

Thorup, M., 2004. Integer priority queues with decrease key in constant time and the single source shortest paths problem. Journal of Computer and System Sciences 69 (3), 330–353 (special issue on STOC 2003).