An Analysis of Taxi Driver's Route Choice Behavior Using the Trace Records

Li Li, *Fellow, IEEE*, Shuofeng Wang, and Fei-Yue Wang, *Fellow, IEEE*

***Abstract*—Understanding travelers' route choice behavior is a key task in transportation studies. In this paper, we analyze the route choices of Beijing taxi drivers regarding four frequently mentioned cost-based route choice rules: pursuing shortest time, or distance, avoiding passing signalized intersections, or making left/right turnings. Test results show that route choices of drivers are not always optimal according to either of these rules. Instead, we argue that taxi drivers are bounded rational and usually choose a satisfactory route that belongs to one of the few routes that consume the shortest times. Test results show that more than 90% observed traces can be explained by this simple explanation.**

***Index Terms*—Big data, route choice, taxi.**

I. Introduction

UNDERSTANDING travelers' route choice behavior is a key element in transportation modeling and urban planning. An appropriate route choice model can help explain travelers' perceptions of route characteristics and predict their actions under certain hypothetical scenarios [1]–[4]. The analysis results will help to build better transportation networks and guidance information so as to relief traffic congestions.

The first study of route choice behavior goes back to Wardrop's paper in 1952 [5]. Wardrop assumed that drivers pursued routes with least costs and proposed the Wardrop's equilibrium in which drivers cannot reduce their travel costs by unilaterally choosing another route [6]. To reach the Wardrop's equilibrium, drivers need to be purely rational and have perfect knowledge [7]. However, this assumption has been criticized for ignoring the limitation of drivers' rationality and knowledge [8].

Then, stochastic user equilibrium (SUE) was proposed as an alternative to describe bounded rational route choices [9]. In SUE, drivers were assumed to be rational but have imperfect knowledge about route costs. Various discrete choice models have been adopted in the SUE problem, including C-logit model [10], path-size logit model [11], and so on. Most of these discrete choice models require an explicit candidate routes generation process and the candidate routes are always assumed to include optimal routes. Prato and Bekhor [12] showed that, in numerical simulation of route choices, the size and the composition of generated route sets could have a significant influence on model performance. Thus, an examination about whether drivers choose the optimal routes helps generate appropriate candidate routes set for discrete choice simulations in practices.

The rise of mobile services and big data techniques [13] make it easier to monitor the movements of massive vehicles and people. This brings opportunity to empirically investigate travelers' route choices.

The most frequently used data are mobile phone data [14] and GPS data [15]–[18]. Most empirical studies focus on route choice behavior of private car drivers [14], [15], [19]–[23] and public transport networks users [24]–[26]. Different hypothesis on travelers' perceptions of route characteristics and preferences of influencing factors were proposed. For example, the preference for routes with thrill and adventure was studied in [27]. The preference for routes with beautiful scenery was examined in [28]. The preference for easy navigation was studied in [29]. The preference for routes across the city center was analyzed in [30].

More studies address the cost of routes as the key factors in route choices. The proposed measure of routes' cost includes travel time [31], [32], travel distance [23], [33], the number of traffic lights along the routes [34], and the number of turns [35], [36]. Several studies have examined the above factors based on empirical traces. For example, Ciscal-Terry *et al.* [22] examined the private car traces with four optimal candidate routes (minimize distance, travel time, number of turns, and number of traffic lights). They compared the difference of travel costs between the observed trace and the calculated optimal candidate route. However, they did not consider the situations in which the observed trace and an optimal candidate route may have very similar travel cost but do not overlap with each other. Zhu and Levinson [23] compared the observed private car traces with the shortest travel time candidate routes. However, they only examined the shortest travel time candidate routes but ignored other optimal routes.

Both these two studies were based on private car traces. Since private car drivers have various travel purposes and

Manuscript received September 11, 2017; revised February 10, 2018; accepted April 25, 2018. Date of publication May 17, 2018; date of current version May 25, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61533019 and Grant 91720000, in part by the Beijing Municipal Science and Technology Commission Program under Grant D171100000317002, and in part by the Beijing Municipal Commission of Transport Program under Grant ZC179074Z. *(Corresponding author: Fei-Yue Wang.)*

L. Li and S. Wang are with the Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: li-li@tsinghua.edu.cn; wsf17@mails.tsinghua.edu.cn).

F.-Y. Wang is with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China (e-mail: feiyue.wang@ia.ac.cn).

Digital Object Identifier 10.1109/TCSS.2018.2831285

different levels of knowledge of road networks, both the two studies [22], [23] found heterogeneous route choice behaviors.

To reduce the influence of heterogeneity, we studied route choices of taxi drivers based on massive taxi trace records collected in Beijing. Though taxi drivers might sometimes be required by some passengers to make certain detour behaviors, we expect them to show significantly higher preference of lower travel costs and we can assume these taxi drivers are familiar with the road network.

Similar to existing studies [22], [23], we examine four frequently mentioned rules: pursuing shortest time [31], [32], or distance [23], [33], avoiding passing signalized intersections [34], or making left/right turnings [35], [36]. Test results show that neither one rule can explain all the traces observed in practice. The four rules can only collectively explain 47.1% observed traces at most.

To further explain drivers' route choice behaviors, we argue that taxi drivers usually choose the route by their experiences of travel times and the finally chosen routes should be among the top few optimal candidate routes. We find that more than 90% observed traces can be explained by this simple explanation. Our findings indicate that taxi drivers yield bounded rationality in route choice behavior and choose satisfactory routes but not always the best route.

To clearly present our findings, the rest of this paper is organized as follows. In Section II, we present the method to transfer the empirical trace records into paths on a simplified network. We also explain how to filter out the abnormal taxi trip records data. In Section III, we analyze the empirical route choices based on conventional and new explanations. Finally, we conclude in Section IV.

II. Data and Methodologies

The raw of taxi trip records cannot be directly used for route choices analysis. Generally, we need to carry out the following processing steps.

1) Transfer the empirical trace records into route records on a simplified network.
2) Filter out abnormal data that may bias the conclusions. We will explain them in detail in Sections II-A and II-B.

A. Data Transfer

In this paper, we examine 22 675 origin-destination (OD) taxi trip records collected from about 7668 taxis that served in Beijing during June 1, 2014–June 26, 2014. These data have also been used in our previous studies [37], [38].

Each OD trip record contains the following information: taxi ID, a consecutive series of sampled GPS positions (longitude and latitude) along with the precise sampling times. The sampling time interval is not a constant but is around 60 s. Scanning the GPS positions and the sampling times, we can identify where and when a taxi trip begins and ends.

Since the GPS positions are often not so accurate in urban areas, we match the trip records onto the road map for Beijing using OpenStreetMap map data. Based on the road map, we formulate a simplified network, so that each trip record can be transferred into a route on the simplified network for future study.

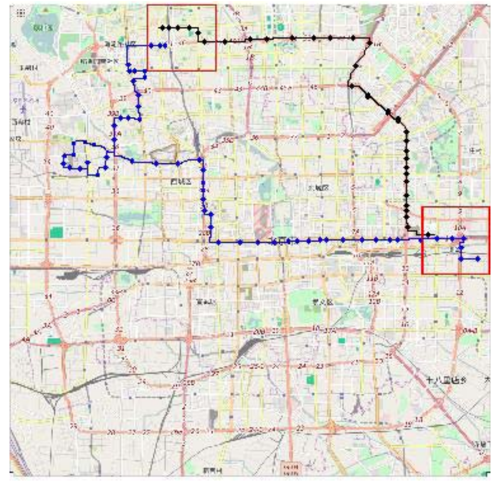


Fig. 1. Illustration of a grid pair, where we plot two traces in it.

In this simplified network, the nodes represent road segments and the edges denote the connections between road segments. Suppose a link L in the road network connects road segment E_1 and E_2, and it has a few attributes. $L \cdot$length denotes the geographical length of road segment E_2. $L \cdot$signals denotes the number of traffic signals locating along road segment E_2. $L \cdot$ turns denotes the turning angles (a value between 0° and 90°) between road segment E_1 and E_2.

B. Data Filtering

In this paper, we first filter out the abnormal data in which taxi drivers take significantly longer time (e.g., triple than usual time) to reach the destination. This is because taxi drivers may wait for the passengers to finish something during some trips. Such records should not be counted in.

Then, we group the remained traces that link neighboring ODs into several clusters, so that abnormal data can be further identified as outliers and be filtered.

Specially, we divide the area within the fifth ring road of Beijing into $11 * 11$ grids. Each grid covers an area of about 5.8 km^2. We will discuss the influence of grid number in Section III-C.

We define the grid pair as the union set of OD pairs between two grids. We call a taxi OD trip record belonging to the grid pair (G_1, G_2) if its origin is in grid G_1 and its destination is in grid G_2.

22 675 taxi trip records seem not so big. However, further analysis shows that more than 95% of taxi trace records collected for this paper is within 66 grid pairs. Each of these 66 grid pairs contains more than 250 OD trip records when the size of the grid is 5.8 km^2. Therefore, these records are enough to retrieve the common characteristics in route choice behaviors.

Fig. 1 gives an illustration of a grid pair and plots two observed traces in this grid pair. The black points and the blue points denote the sampled GPS positions of two trip records. The red rectangles denote the origin and the destination grids. The blue observed trace is indeed an abnormal trip record because there is a ring in it.

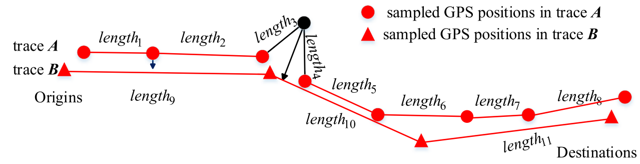


Fig. 2. Illustration of coverage score algorithm.

To filter out the abnormal data, we group the observed traces in a grid pair into several clusters, so that abnormal data can be identified as outliers and be filtered.

To this end, we adopt the so-called coverage score derived from [39] as the similarity between traces. This calculating algorithm of coverage score works as follows.

1) For each sampled position in trace $\boldsymbol{A}$, we calculate the minimum distance from it to trace $\boldsymbol{B}$. Considering the error caused by GPS accuracy, if the minimum distance is below 30 m, we think the corresponding sampled position locates in trace $\boldsymbol{B}$. Otherwise, it does not locate in route $\boldsymbol{B}$.
2) We then calculate the coverage score Score_{AB} as

$$\text{Score}_{AB} = \frac{2 * \text{len}_C}{\text{len}_A + \text{len}_B} \tag{1}$$

where len_C denotes the geographical length of the coverage part between trace $\boldsymbol{A}$ and trace $\boldsymbol{B}$, len_B denotes the length of trace $\boldsymbol{B}$, len_A denotes the length of trace $\boldsymbol{A}$.

Fig. 2 illustrates how to calculate the coverage score. The red round points in Fig. 2 denote position samples in trace $\boldsymbol{A}$ which locate in trace $\boldsymbol{B}$, and the black round point denotes the position sample which does not locate in trace $\boldsymbol{B}$. In the illustration, $\text{len}_A = \sum_{i=1}^{i=8} \text{length}_i$, $\text{len}_B = \sum_{i=9}^{i=11} \text{length}_i$, and $\text{len}_c = \sum_{i=1}^{i=2} \text{length}_i + \sum_{i=5}^{i=8} \text{length}_i$.

The coverage score below a threshold value S_0 is finally set to 0 when grouping the observed traces into clusters. This trick makes all the abnormal traces be grouped into one cluster. Test results show that $S_0 = 0.73$ is an appropriate empirical threshold to distinguish different clusters and meanwhile allow enough diversity within clusters. We will discuss the influence of threshold value S_0 in Section III-C.

Based on the coverage scores, we use the "community detection" algorithm proposed in [40] to group the observed traces in a grid pair into several clusters.

As shown in [41] and [42], the fewer traces a cluster contains, the more likely this cluster denotes the set of outlier (abnormal traces).

When S_0 is set to 0.73, we totally filter out 6.4% traces and get 21 218 valid traces data after discarding the abnormal traces. Fig. 3 gives an illustration of observed traces for a grid pair and plots three clusters in it when S_0 is set to 0.73. Traces in a cluster follow similar routes but meanwhile allow tolerable diversity. The clusters containing few traces have not been plotted in Fig. 3. Especially, we can see the trace with a ring shown in Fig. 1 is discarded.

Moreover, most observed routes shown in Fig. 3 are part of expressways. This is partly because expressways have fewer turns/signalized intersections and it is often faster to drive on expressways than other roads. So, the plots in Fig. 3 are in agreement with both the conventional explanation and new explanation.

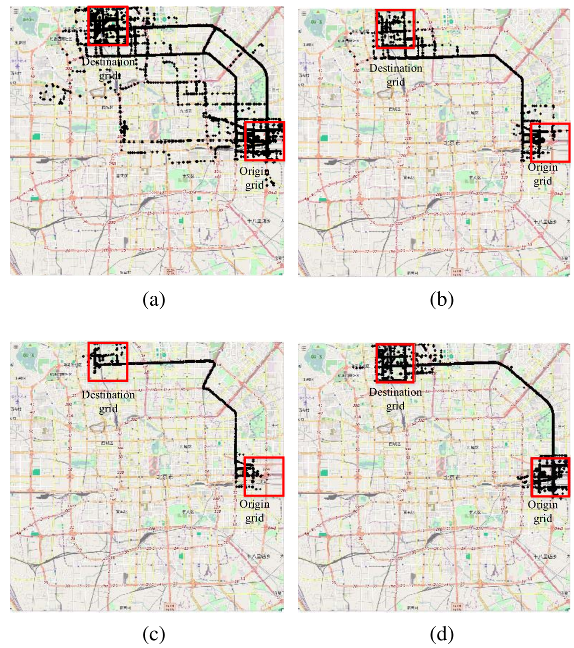


Fig. 3. (a) Traces observed for a grid pair, and (b)–(d) traces that belong to the found three clusters, respectively.

III. Numerical Testing Results

A. Performance of Conventional Explanations

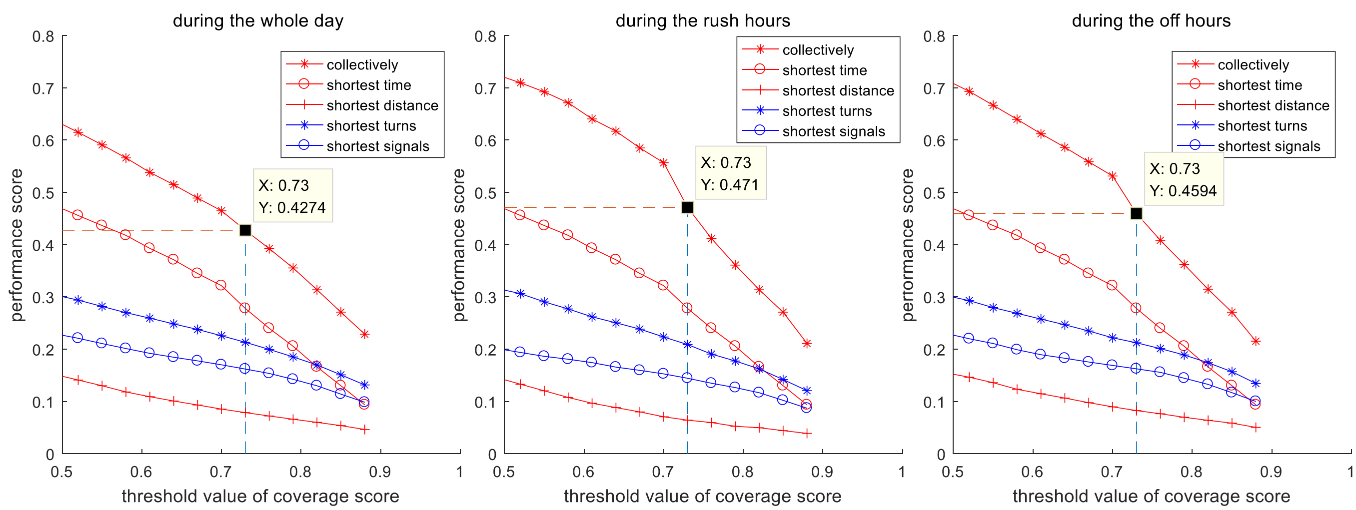
In this section, we check whether the aforementioned conventional explanations work for empirical route choice behaviors of taxi drivers. We first calculate the best candidate routes according to the four rules mentioned in Section I. Then, we compare the empirical trace with each obtained the best candidate route, respectively. The coverage score between empirical traces and a candidate route are calculated. If the coverage score is more than a preselected threshold S_0 (e.g., 0.73 used in the following), we say this empirical trace can be explained by the corresponding conventional rule. If an observed trace can be explained by at least one of the four candidate routes, we say that the four rules can collectively explain the trace route.

In the following, the four best candidate routes are named as "shortest distance," "shortest time," "shortest turns," and "shortest signals," respectively.

Given the starting node s and the terminal node t, a route that connects s and t is a sequence of nodes $P = (v_1, v_2, \ldots v_n)$ (where $v_1 = s$ and $v_n = t$) such that v_i is adjacent to v_{i+1} for $1 \le i \le n-1$. We use L_{v_i,v_j} to denote the network link between v_i and v_j.

The "shortest distance" route that connects s and t is obtained by solving the following optimization problem:

$$\min \sum_{i=1}^{i=n-1} L_{v_i,v_j} \cdot \text{length}. \tag{2}$$



(a) (b) (c)

Fig. 4. Performance scores for different candidate routes. (a) Performance score during the whole day. (b) Performance score during the rush hours. (c) Performance score during the off hours.

The "shortest signals" route that connects s and t is obtained by solving the solution of the following optimization problem:

$$\min \sum_{i=1}^{i=n-1} L_{v_i,v_j} \cdot \text{signals}. \tag{3}$$

The "shortest turns" route that connects s and t is obtained by solving the following optimization problem:

$$\min \sum_{i=1}^{i=n-1} L_{v_i,v_j} \cdot \text{turns}. \tag{4}$$

The shortest time route is the route that connects s and t, and meanwhile consumes the least travel time among history travel time.

As travel time always changes with time, we calculate three candidates "shortest time" routes for each OD trip record, one for the whole day, one for the rush hours, and one for the off hours.

In this paper, rush hours are defined as the time interval during which the amount of taxi orders per hour is highest. Off hours are defined as the time interval during which the amount of taxi orders per hour is lowest. For different grid pairs, the rush hours or off hours may be different, but for most grid pairs, the rush hours are 10:00–14:00, and the off hours are 00:00–6:00 and 22:00–24:00. There is a significant time delay between the rush hours for private cars and taxis. This is partly because the departure times of travelers with different travel modes may vary.

We define the so-called "*performance score*" to show the percent of observed traces that can be explained by one kind of candidate route. For example, the "*performance score*" of the "shortest distance" route is equal to the percent of observed traces that can be explained by the corresponding "shortest distance" candidate routes.

Fig. 4 gives the performance scores for different candidate routes. We can see that the "shortest distance" candidate route yields the worst performance than the other three candidate routes. If the threshold value is chosen as 0.73, the "shortest time" route performs the best among the four candidate routes.

The "collectively performance score" is below 0.47, no matter what kind of "shortest time" routes are considered. It indicates neither one rule nor their combinations can well explain all the traces observed in practice.

B. Performance of New Explanation

In this paper, we give a new explanation for route choice behavior of taxi driver. We argue that taxi drivers usually choose the route by their experiences of travel times and they finally choose routes among the top few optimal candidate routes.

To verify this hypothesis, we introduce the following measure. If two routes belong to the top α (e.g., 10% used in this paper) shortest travel time routes, we say none of them is significantly better than the other in the minimum travel time measure. Moreover, if the percentage difference in mean travel time between two routes is below a threshold value β (e.g., 10% used in this paper), we say that none of them is significantly better than the other in the mean travel time measure. If two routes are not significantly better than each other in both the minimum travel time and the mean travel time measure, we say that none of them is significantly better than the other in the travel time measure.

We choose α as 10% because the top 10% shortest travel time routes are fast enough. We choose β as 10% to guarantee that the mean travel times of two routes are approximately the same.

As mentioned in Section II, the observed traces in a cluster follow the same route excepting tolerable diversity. Therefore, to examine our explanation is equivalent to examine the following two hypotheses.

1) For a grid pair, every cluster contains at least one trace which belongs to the top α (e.g., 10% used in this paper) shortest travel time empirical traces.

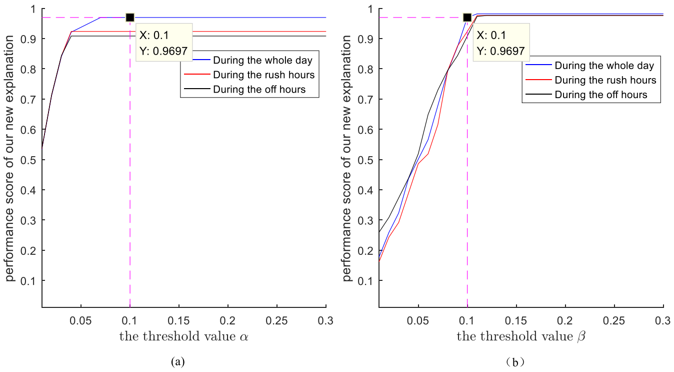


Fig. 5. Performance scores for our new explanation. (a) β is set to 0.1, S_0 is set to 0.73, and the grid number is 121. (b) α is set to 0.1, S_0 is set to 0.73, and the grid number is 121.

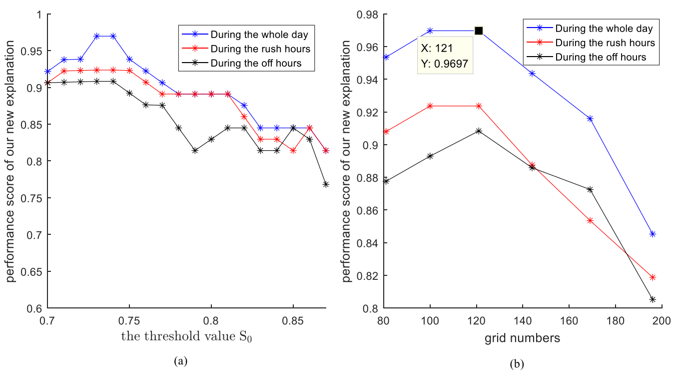


Fig. 6. Performance scores for our new explanation. (a) α is set to 0.1, β is set to 0.1, and the grid number is 121. (b) α is set to 0.1, β is set to 0.1, and S_0 is set to 0.73.

2) For every two clusters in a grid pair, the percentage difference in mean travel time of empirical traces is below the threshold value β. This hypothesis is respectively checked for the whole day, for the rush hours and for the off hours.

If both the two hypotheses are correct for a grid pair, we say that traces in the grid pair can be explained by our new explanation.

The "performance score" of our new explanation is defined as the percent of observed traces that can be explained our new explanation. Fig. 5 gives the performance score for our new explanation. We can see that 96.97% (20 576 in 21 218 valid traces) observed traces can be explained by our explanation. It also indicates that taxi drivers choose a satisfactory route not the best route both in the minimum travel time and in the mean travel time.

C. Sensitivity Analysis

In this section, we present the sensitivity analysis of parameters such as the threshold value S_0 and grid number. Fig. 6 plots the variation of performance score in terms of the threshold value S_0 and grid number.

First, we can see that, no matter what values of S_0 and grid number are chosen, the performance score for our new explanation is always higher than the performance score for conventional explanations. This has already verified the effectiveness of the proposed new explanation.

Second, if S_0 is selected to be too large, the tolerance for route diversity decreases noticeably and thus leads to the decrease of overall performance score. If S_0 is selected to be too small, we cannot easily categorize traces into appropriate clusters. According to Fig. 6, the optimal value of S_0 is 0.73 which leads to the highest performance score for the new explanation. So, we adopt $S_0 = 0.73$ in all the above-mentioned study.

Third, the grid number cannot be chosen as too large, since we do not collect millions of trace records in this paper. If we set a too large grid number, each grid may just contain tens of trace records and we cannot correctly analyze such cases. On the other side, if the grid number is chosen to be too small, several traces with significantly different ODs will be categorized into one cluster and we cannot handle such cases, either. According to Fig. 6, the optimal value of grid number is 121 (or $11 * 11$).

Under the choice $S_0 = 0.73$ and grid number is 121, the performance score of our new explanation is as high as 96.97%.

IV. Conclusion

In this paper, we examine two kinds of explanations of route choice behaviors based on massive taxi trace records collected in Beijing. The first kind of explanations argues that traveler choose the routes either with the shortest travel distance, the shortest travel time, the least number of turns, or the least number of traffic signals. However, test results show that even the combination of all these four rules can only explain at most 47.1% observed traces.

The second kind of explanations proposed in this paper argues that drivers usually choose the satisfactory route but not always the best routes. Test results show that more than 90% observed traces can be explained under this assumption. This indicates that, though taxi drivers are with rich experience and familiar with the road network, they are with bounded rationality in route choices.

Moreover, our study also proves two benefits of big data analysis. First, it becomes easy for us to distinguish anomaly when enough data are available. Second, a comprehensive analysis of big data may prevent us from conforming to some stereotype conclusions.

However, the method proposed in this paper may be improved in some ways, because of the limitation of taxi trace records that we collected. For example, some customers may ask taxi drivers to select some special routes. However, current taxi trip records do not support us to differ customers. So, we cannot tell who dominate the selection of routes. In the future, we would like to investigate Internet-based taxi trip records which differ customers to examine this interesting problem.

References

[1] J. N. Prashker and S. Bekhor, "Route choice models used in the stochastic user equilibrium problem: A review," *Transp. Rev.*, vol. 24, no. 4, pp. 437–463, 2004.

[2] C. G. Prato, "Route choice modeling: Past, present and future research directions," *J. Choice Model.*, vol. 2, no. 1, pp. 65–100, 2009.

[3] W. Leong and D. A. Hensher, "Embedding decision heuristics in discrete choice models: A review," *Transp. Rev.*, vol. 32, no. 3, pp. 313–331, 2012.

[4] G. M. Ramos, W. Daamen, and S. Hoogendoorn, "A state-of-the-art review: Developments in utility theory, prospect theory and regret theory to investigate travellers' behaviour in situations involving travel time uncertainty," *Transp. Rev.*, vol. 34, no. 1, pp. 46–67, 2014.

[5] J. G. Wardrop, "Road paper. Some theoretical aspects of road traffic research," *Proc. Inst. Civil Eng.*, vol. 1, no. 5, pp. 767–768, 1952.

[6] S. C. Dafermos and F. T. Sparrow, "The traffic assignment problem for a general network," *J. Res. Nat. Bureau Standard B*, vol. 73, no. 2, pp. 91–118, 1969.

[7] T. Gärling, "Behavioral assumptions overlooked in travel-choice modeling," in *Travel Behaviour Research: Updating the State of Play*. Amsterdam, The Netherlands: Elsevier, 1998, pp. 3–18..

[8] E. J. Manley, S. W. Orr, and T. Cheng, "A heuristic model of bounded route choice in urban areas," *Transp. Res. C, Emerg. Technol.*, vol. 56, pp. 195–209, Jul. 2015.

[9] C. F. Daganzo and Y. Sheffi, "On stochastic models of traffic assignment," *Transp. Sci.*, vol. 11, no. 3, pp. 253–274, 1977.

[10] E. Cascetta, A. Nuzzolo, F. Russo, and A. Vitetta, "A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks," in *Proc. Int. Symp. Transp. Traffic Theory*, 1996, pp. 697–711.

[11] M. Ben-Akiva and M. Bierlaire, "Discrete choice methods and their applications to short term travel decisions," in *Handbook of Transportation Science*, vol. 23. Boston, MA, USA: Springer, 1999, pp. 5–33.

[12] C. Prato and S. Bekhor, "Modeling route choice behavior: How relevant is the composition of choice set?" *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2003, pp. 64–73, Jan. 2007.

[13] Y. Zheng, "Urban computing: Enabling urban intelligence with big data," *Frontiers Comput. Sci.*, vol. 11, no. 1, pp. 1–3, 2017.

[14] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González, "The path most traveled: Travel demand estimation using big data resources," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 162–177, Sep. 2015.

[15] K. Ramaekers, S. Reumers, G. Wets, and M. Cools, "Modelling route choice decisions of car travellers using combined gps and diary data," *Netw. Spatial Econ.*, vol. 13, no. 3, pp. 351–372, 2013.

[16] M. Zimmermann, T. Mai, and E. Frejinger, "Bike route choice modeling using GPS data without choice sets of paths," *Transp. Res. C, Emerg. Technol.*, vol. 75, pp. 183–196, Feb. 2017.

[17] Z. He, K. Chen, and X. Chen, "A collaborative method for route discovery using taxi drivers' experience and preferences," *IEEE Trans. Intell. Transp. Syst.*, to be published. [Online]. Available: https://ieeexplore.ieee.org/document/8064204/, doi: 10.1109/TITS.2017.2753468.

[18] S. Gong, J. Cartlidge, Y. Yue, G. Qiu, Q. Li, and J. Xin, "Geographical huff model calibration using taxi trajectory data," in *Proc. 10th ACM SIGSPATIAL Workshop Comput. Transp. Sci.*, 2017, pp. 30–35.

[19] N. Dhakar and S. Srinivasan, "Route choice modeling using GPS-based travel surveys," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 9, no. 2413, pp. 65–73, 2014.

[20] M. Abdel-Aty and Y. Huang, "Exploratory spatial analysis of expressway ramps and its effect on route choice," *J. Transp. Eng.*, vol. 130, no. 1, pp. 104–112, 2004.

[21] D. Sun, C. Zhang, L. Zhang, F. Chen, and Z.-R. Peng, "Urban travel behavior analyses and route prediction based on floating car data," *Transp. Lett.*, vol. 6, no. 3, pp. 118–125, 2014.

[22] W. Ciscal-Terry, M. Dell'Amico, N. S. Hadjidimitriou, and M. Iori, "An analysis of drivers route choice behaviour using GPS data and optimal alternatives," *J. Transp. Geograph.*, vol. 51, pp. 119–129, Feb. 2016.

[23] S. Zhu and D. Levinson, "Do people use the shortest path? An empirical test of Wardrop's first principle," *PLoS ONE*, vol. 10, no. 8, p. e0134322, 2015.

[24] S. Raveau, Z. Guo, J. C. Muñoz, and N. H. M. Wilson, "A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics," *Transp. Res. A, Policy Pract.*, vol. 66, pp. 185–195, Aug. 2014.

[25] T. Brands, E. de Romph, T. Veitch, and J. Cook, "Modelling public transport route choice, with multiple access and egress modes," *Transp. Res. Procedia*, vol. 1, no. 1, pp. 12–23, 2014.

[26] C. Viggiano, H. N. Koutsopoulos, J. Attanucci, and N. H. Wilson, "Inferring public transport access distance from smart card registration and transaction data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1, no. 2544, pp. 55–62, 2016.

[27] S. Bekhor and G. Albert, "Accounting for sensation seeking in route choice behavior with travel time information," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 22, pp. 39–49, Jan. 2014.

[28] D. Quercia, R. Schifanella, and L. M. Aiello, "The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city," in *Proc. 25th ACM Conf. Hypertext Social Media*, 2014, pp. 116–125.

[29] J. Shao, L. Kulik, E. Tanin, and L. Guo, "Travel distance versus navigation complexity: A study on different spatial queries on road networks," in *Proc. 23rd ACM Int. Conf. Conf. Inf. Knowl. Manage.*, 2014, pp. 1791–1794.

[30] T. Thomas and B. Tutert, "Route choice behavior in a radial structured urban network: Do people choose the orbital or the route through the city center?" *J. Transp. Geograph.*, vol. 48, pp. 85–95, Oct. 2015.

[31] K. L. Cooke and E. Halsey, "The shortest route through a network with time-dependent internodal transit times," *J. Math. Anal. Appl.*, vol. 14, no. 3, pp. 493–498, 1966.

[32] S. E. Dreyfus, "An appraisal of some shortest-path algorithms," *Oper. Res.*, vol. 17, no. 3, pp. 395–412, 1969.

[33] R. Bellman, "On a routing problem," *Quart. Appl. Math.*, vol. 16, no. 1, pp. 87–90, 1958.

[34] R. G. Golledge and T. Gärling, "Cognitive maps and urban travel," in *Handbook Transport Geography Spatial System*. Bingley, U.K.: Emerald Group Publishing Limited, 2004, pp. 501–512.

[35] D. M. Mark, "Finding simple routes:'Ease of description'as an objective function in automated route selection," in *Proc. CAIA*, 1985, pp. 577–581.

[36] B. Jiang and X. Liu, "Computing the fewest-turn map directions based on the connectivity of natural roads," *Int. J. Geograph. Inf. Sci.*, vol. 25, no. 7, pp. 1069–1082, 2011.

[37] B. Leng, H. Du, J. Wang, Z. Xiong, and L. Li, "Analysis of taxi drivers' behaviors within a battle between two taxi apps," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 296–300, Jan. 2015.

[38] Z. Li, S. Jiang, J. Dong, S. Wang, Z. Ming, and L. Li, "Battery capacity design for electric vehicles considering the diversity of daily vehicles miles traveled," *Transp. Res. C, Emerg. Technol.*, vol. 72, pp. 272–282, Nov. 2016.

[39] J. Froehlich and J. Krumm, "Route prediction from trip observations," SAE Tech. Paper 2008-01-0201, 2008.

[40] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech., Theory Experim.*, vol. 2008, no. 10, p. P10008, 2008.

[41] D. Zhang, N. Li, Z.-H. Zhou, C. Chen, L. Sun, and S. Li, "iBAT: Detecting anomalous taxi trajectories from GPS traces," in *Proc. Int. Conf. Ubiquitous Comput.*, 2011, pp. 99–108.

[42] Y. Dong, S. Wang, L. Li, and Z. Zhang, "An empirical study on travel patterns of Internet based ride-sharing," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 1–22, Jan. 2018.

Li Li (S'05–M'06–SM'10–F'17) is currently an Associate Professor with the Department of Automation, Tsinghua University, Beijing, China, where he is involved in intelligent control and sensing, intelligent transportation systems, and intelligent vehicles. He has authored or co-authored over 70 SCI indexed international journal papers and over 70 international conference papers as a first/corresponding author.

Shuofeng Wang is currently pursuing the Ph.D. degree with the Department of Automation, Tsinghua University, Beijing China.

His current research interests include travel behavior analysis.

Fei-Yue Wang (S'87–M'89–SM'94–F'03) received the Ph.D. degree in computer and systems engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990.

In 1990, he joined the University of Arizona, Tucson, AZ, USA, and then became a Professor and the Director of the Robotics and Automation Laboratory, Tucson, and Program in Advanced Research for Complex Systems. In 1999, he founded the Intelligent Control and Systems Engineering Center, Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, under the support of the Outstanding Oversea Chinese Talents Program from the State Planning Council and 100 Talent Program from CAS. In 2002, he was appointed as the Director of the Key Laboratory of Complex Systems and Intelligence Science, CAS, where he became the State Specially Appointed Expert and the Director of the State Key Laboratory of Management and Control for Complex Systems in 2011. His current research interests include methods and applications for parallel systems, social computing, and knowledge automation.

Dr. Wang is a Fellow of INCOSE, IFAC, ASME, and AAAS. He was a recipient of the Second Class National Prize in Natural Sciences of China and the Outstanding Scientist by ACM for his work in intelligent control and social computing in 2007, the IEEE ITS Outstanding Application and Research Awards in 2009 and 2011, and the IEEE SMC Society Norbert Wiener Award in 2014.Since 1997, he has served as the General or Program Chair of more than 20 IEEE, INFORMS, ACM, and ASME conferences. He was the President of IEEE Intelligent Transportation Systems (ITS) Society from 2005 to 2007, the Chinese Association for Science and Technology, USA, in 2005, the American Zhu Kezhen Education Foundation from 2007 to 2008, and the Vice President of the ACM China Council from 2010 to 2011. Since 2008, he has been the Vice President and the Secretary General of Chinese Association of Automation. He was the Founding Editor-in-Chief of the *International Journal of Intelligent Control and Systems* from 1995 to 2000 and the IEEE INTELLIGENT TRANSPORTATION SYSTEMS MAGAZINE from 2006 to 2007 and the Editor-in-Chief of the IEEE INTELLIGENT SYSTEMS from 2009 to 2012 and the IEEE TRANSACTIONS ON ITS from 2009 to 2016. He is currently the Editor-in-Chief of the IEEE/CAA JOURNAL OF AUTOMATICA SINICA and the *Chinese Journal of Command and Control*.