Research on Taxi Search Strategy based on Passenger Distribution and Traffic Conditions

Ben Hao YANG*, Ruo Qian WU
School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
*Email: yangbenhao2015@buaa.du.cn

Abstract: With the rising smart transportation and sharing economy, taxis play an increasingly important role in urban public transportations. While researches focus mostly on the operation of taxis based on the long-term traffic supply and demand equilibrium, it is essentially unknown how the real-time recognition of passenger demand and traffic conditions will influence the outcome of the taxi search strategies. Based on Beijing traffic data, we study the optimal search strategy of taxis through simulation analysis in real network, considering both the passenger distribution and the traffic conditions. The results show that the taxi driver's passenger empirical index will benefit the search efficiency. More important, it is suggested that while weighing between the passenger distribution and traffic conditions, the optimal search strategy at low passenger density is moving towards areas with more passengers, and the optimal strategy at high passenger density is moving towards areas with better traffic conditions. Our findings can help on-line taxi operators achieve better management efficiency with real-time forecasting of passenger distribution and traffic conditions.

1. Introduction

Taxi plays an increasingly important role in the public travel of urban residents, and its operating efficiency is closely related to the residents' travel expenses. The taxi operation process can be divided into two stages: taxi service process while occupied and taxi search process while vacant, and the vacant taxi search efficiency is the main factor that determines the taxi operation efficiency. Therefore, research on the optimizing of the taxi drivers’ search strategies is very important. The rising of big data makes it possible to collect and forecast information such as urban passenger distribution and traffic conditions[1-5]. Based on passenger distribution and traffic conditions, improving the taxi operation efficiency and the overall traffic efficiency[6,7] of the city has become the future trends.

Most of the original studies of taxi operations ignore the temporal and spatial characteristics of taxi search behaviors, mainly studying from the economic view and trying to determine the price and taxi quantity to achieve the balance between taxi supply and passenger demand, and explain how the taxi supply and demand leads to high fees, low efficiency and other results[8-12]. These studies mainly illustrate the impact of pricing and taxi quantity on the taxi operation efficiency from the perspective of long-term planning, which is still not enough to provide the basis for real-time operational decision-making. In the actual taxi operation process, the driver's search strategy depends not only on the balance of supply and demand, but also on the real-time passenger distribution and road traffic conditions.

Yang et al. considered the spatial characteristics of the taxi operating route and established a static network model[13]. The model assumes that the passengers are distributed at nodes of the network and they used mathematical methods to acquire the relationship among some system properties such as
taxi quantity, taxi utilization rate, the average waiting time of the passengers and so on. The results show that the average taxi utilization rate decreases sharply with the increasing of taxi quantity, and the average passenger waiting time increases with the increasing of taxi utilization rate. Subsequent studies have further considered factors such as road congestion and passenger demand elasticity[14], the taxi service supply and demand balance under competition or monopoly market[15], and also established a multiperiod taxi service dynamic model[16]. The basic assumption of the above models is that once the passenger gets off and the taxi will go to the new target node and search continuously at the target node until the next passenger is found, ignoring the search process of the passenger during the journey to the target node. And in fact the taxi search strategy will also be constantly adjusted with the changing of surrounding information. Mu et al.[17-19] simulated the dynamic search process for passengers of vacant taxis on the road network, analyzed the impact of different search strategies on the taxi utilization rate, empty-run rate, average search time and other system properties, and found that the search information based on historical information can effectively improve system operational efficiency. The model considers the distribution of historical passenger demand, but do not consider the impact of road traffic conditions on the actual search process and search efficiency. The collection of floating vehicle data makes it possible to analyze the behavior of the taxi based on the real data. Wong et al.[20] used 460 city taxi GPS data to analyze taxi search behavior, and the results show that vacant taxi drivers often travel to high demand areas during early morning, but avoid crossing different areas during peak periods.

From the above analysis, we can see that most of the studies focus on the overall balance of supply and demand of the taxi system, or only consider the passenger distribution to optimize the search strategies, few works have comprehensively analyzed the optimal search strategies of taxis under the influence of both passenger distribution and road traffic conditions. Based on the Beijing traffic network (as shown in Fig. 1a), this study will establish a taxi search model in the urban traffic network and consider both the passenger distribution and the traffic conditions, and compare and analyze the taxi search efficiency under different search strategies, and finally acquire the optimal search strategy under different conditions.

2. Search strategy modeling

At a certain passenger distribution, the taxi's search process for passengers can be modeled as follows: once a taxi arrives at any road nodes, it will choose its direction in the next step, and the direction choice will be decided by the passenger number and traffic conditions in each direction. As shown in Fig. 1c, the direction choice reflects its search strategy, and this strategy depends largely on their working experience: experienced drivers have a higher recognition level on the city's passenger distribution based on their rich historical experience, thus they are more likely to move to areas with more passengers; but for less experienced drivers the direction choice may show some blindness. The taxi direction choosing probability includes two parts, the probability considering experience for passenger distribution and the random probability considering blindness. Here we define the driver empirical index $\alpha$, $0 \leq \alpha \leq 1$. When the taxi is located at the node $i$ with degree $k$ of the road network, the direction choosing probability can be expressed as follows:

$$
P_{i\rightarrow 1} = (1 - \alpha)P_{0,1} + \alpha P_{1,1}$$

$$
P_{i\rightarrow 2} = (1 - \alpha)P_{0,2} + \alpha P_{1,2}$$

$$
\vdots$$

$$
P_{i\rightarrow k} = (1 - \alpha)P_{0,k} + \alpha P_{1,k}
$$

where the random probability in the $k$th direction $P_{0,k} = P_{0,2} = \cdots = P_{0,k} = 1/k$, and the probability in each direction considering passenger distribution is in direct proportion to passenger quantity in that direction, that is $P_{1,1} = \frac{m_1}{m}$, $P_{1,2} = \frac{m_2}{m}$, $\cdots$, $P_{1,k} = \frac{m_k}{m}$, where $m_1$, $m_2$, $\cdots$, $m_k$ correspond to passenger quantity in each nodes and $m$ is the total passenger quantity of $k$ nodes. While $\alpha = 0$, the driver’s direction choice shows a complete blindness. While $\alpha = 1$, the driver’s choice completely follows passenger distribution, that is moving towards directions with more passengers.
On the basis of considering passenger distribution, the road traffic conditions will be added to the search model. Here we define the trade-off index $\beta$, $0 \leq \beta \leq 1$ of taxi drivers for passenger distribution and traffic conditions. This index reflects the tendency of the taxi driver when facing different passenger distributions and traffic conditions: moving towards nodes with good traffic conditions or towards nodes with a relatively large number of passengers. In order to reflect the coupling relationship between the passenger number of a given node and the road speed towards the node, here we assume the road speed towards node $j$, $V_j = 80/(1 + m_j \gamma)$, where $m_j$ is passenger quantity on node $j$, and $\gamma$ is the coupling strength between $V_j$ and $m_j$. As shown in Fig. 1c, the taxi direction choosing probability is as follows:

$$P_{i \rightarrow j} = (1-\beta)P_{i,j} + \beta P_{i,j+1}$$

$$P_{i \rightarrow j} = (1-\beta)P_{i,j} + \beta P_{i,j-1}$$

$$\vdots$$

$$P_{i \rightarrow k} = (1-\beta)P_{i,k} + \beta P_{i,k+1}$$

where $P_{i,j}$, $P_{i,j+1}$, $P_{i,j-1}$, $P_{i,k}$, $P_{i,k+1}$ correspond to probability considering passenger distribution, and $V_{i,j}, V_{i,j+1}, \ldots, V_{i,k}$ correspond to road speed in each direction, and $V$ is the total speed of the vehicle. $P_{i,j}$, $P_{i,j+1}$, $\ldots$, $P_{i,k}$ are the above probability considering passenger distribution. $\beta = 0$ means the drivers tend to move towards areas with more passengers without considering traffic conditions, while $\beta = 1$ means the drivers tend to move towards areas with better traffic conditions without considering passenger quantity.

With the above search model, we can define each taxi’s search efficiency as follows:

$$E = N_0/T_0 = 1/(\langle T \rangle)$$

where $N_0$ is the total passengers found within total vacant operating time $T_0$, $\langle T \rangle$ is the average search time. The taxi travel time along road $j$ is $t_j = l_j/V_j$, where $l_j$ is the road length and $V_j$ is current speed of road $j$. During the actual taxi operation, the vacant and occupied stages will appear alternately, thus in the model, each time a taxi finds a passenger, it will randomly appear at a node and starts its next search.

Here we consider four different passenger number distributions: uniform distribution, each passenger falls on any node with the same probability; exponential distribution, passenger number follows $p(n) = 1/n \ e^{-n}/\rho$, where $\rho$ is passenger density (average passenger number on each node); power-law distribution, passenger number follows $p(n) \sim n^{-2}$ (taking the reality into account, the passenger number on each node is limited within 0~20); actual distribution, passenger distribution obtained from the real taxi operating trajectory data.

3. Simulation results
At different passenger distributions and passenger density $\rho$, Fig.2 shows the simulation results of
how the search efficiency $E$ changes with the empirical index $\alpha$ on Beijing traffic network. Fig. 2a shows how $E$ changes with $\alpha$ for uniform passenger distribution. We can see that larger $\alpha$ results in higher $E$, which indicates that the experienced (larger $\alpha$) drivers tend to move towards areas with high $\rho$ based on their experience about surrounding information, which leads to high efficiency. We can also see that when $\rho$ is large, the taxi will find passengers in a shorter time and achieve higher efficiency. Fig. 2b and Fig. 2a have a basically similar trend. For these two passenger distributions, the passengers distribute very homogeneously on the network, and the curves have no significant difference. Comparing with Fig. 2a, the efficiency in Fig. 2c is relatively small. At power-law passenger distribution the passengers distribute very heterogeneously. The number of passengers is large on a small set of nodes, but is small or even 0 on a large set of nodes. In this case, only a small number of taxis happen to appear around nodes with large passenger number and these taxis will acquire high $E$, while the efficiency of the rest taxis will be significantly lower. This phenomenon is more pronounced in the actual passenger distribution of Fig. 2d, since in the actual taxi operation, the passengers are basically concentrated in the hotspot area of the central city, while most nodes in suburban area have very small number of passengers. This difference then further leads to lower search efficiency.

Figure 2 Model simulation results considering only passenger distribution. Each figure shows how search efficiency $E$ changes with empirical index $\alpha$. There are 500 taxis in the network, each taxi runs 5000 minutes, and results in (a)-(c) are averaged over 500 realizations (in each realization the passengers are regenerated). At a given $\alpha$, each taxi empirical index $\alpha_{\ell}$ obeys the normal distribution $\alpha_{\ell} \sim N(\alpha, 0.01)$. (a) uniform passenger distribution; (b) exponential passenger distribution; (c) power-law passenger distribution; (d) actual passenger distribution in Beijing, different passenger density $\rho$ correspond to passenger quantity statistics around 6 p.m. within 1h, 2h and 4h, respectively.

Considering both passenger distribution and traffic conditions, Fig. 3 shows how the efficiency $E$ changes with the trade-off index $\beta$ in Beijing traffic network. At uniform passenger distribution, we can see in Fig. 3a that $\beta_{\text{opt}} = 0$ at low $\rho$ and $\beta_{\text{opt}} = 1$ at high $\rho$. When $\rho$ is low and the number of passengers is small, the road traffic is generally in good condition. In this case, the optimal strategy is moving towards areas with more passengers, which corresponds to $\beta_{\text{opt}} = 0$. When $\rho$ is high and
the number of passengers is large, the road traffic is generally in very poor condition. In this case, the optimal strategy is moving towards areas with better traffic conditions, which corresponds to \( \beta_{\text{opt}} = 1 \). Fig. 3b and Fig. 3a share basically the same trend, and the difference is that \( E \) is relatively low at high \( \rho \) in Fig. 3b. At the same \( \rho \), the exponential distribution has more nodes with high \( \rho \) comparing to uniform distribution. These nodes will further increase the heterogeneity of the road velocity distribution, resulting in the decrease of \( E \). The efficiency will be further reduced because of more heterogeneous passenger distribution, which is more clearly shown in Fig. 3c. There are more nodes with more passengers in power-law passenger distribution, and the traffic heterogeneity caused by these nodes are more pronounced, thus \( E \) will be further reduced due to the heterogeneous passenger distribution. The results of the actual passenger distribution are very similar (see Fig. 3d).

![Figure 3 Model simulation results considering both passenger distribution and traffic conditions. Each figure shows how search efficiency \( E \) changes with trade-off index \( \beta \). There are 500 taxis in the network, each taxi runs 5000 minutes, and results in (a)-(c) are averaged over 500 realizations (in each realization the passengers are regenerated). At a given \( \beta \), each taxi trade-off index \( \beta \) follows the normal distribution \( \beta_t \sim N(\beta, 0.01) \). (a) uniform passenger distribution; (b) exponential passenger distribution; (c) power-law passenger distribution; (d) actual passenger distribution in Beijing, different passenger density \( \rho \) correspond to passenger quantity statistics around 6 p.m. within 1h, 2h and 4h, respectively.](image)

Fig. 4 shows how the search efficiency \( E \) changes with trade-off index \( \beta \) and coupling strength \( \gamma \) at actual passenger distribution. It can be seen that the coupling relationship between passenger quantity and traffic conditions determines the optimal search strategy. At lower \( \gamma \), the number of passengers in each node has little effect on the road speed. In this case, the optimal search strategy is moving towards areas with more passengers and considering less about traffic conditions. When \( \gamma \) is large, the passenger quantity in each node has a great influence on the road speed, and the search strategy considering the passenger distribution will cost more time than before, resulting in a significant reduction in search efficiency.
4. Conclusion and prospects

Based on the real traffic data of Beijing, this paper establishes a taxi search model and considers two factors: passenger distribution and traffic conditions. The simulation results show that the passenger’s empirical index $\alpha$ plays an important role in the search efficiency and the improvement of $\alpha$ will obviously increase the search efficiency. At different passenger distributions, the impact of $\alpha$ is also different. This result shows that effective use of historical data will help drivers to improve search efficiency. When the passenger distribution affects the traffic conditions, the taxi needs to trade off between the two factors: at the low passenger density the optimal search strategy for the driver is moving towards areas with more passengers, and at high passenger density the optimal search strategy for the driver is moving towards areas with better traffic conditions. The coupling relationship between passenger distribution and traffic conditions in actual traffic is more complex and needs to be analyzed further in the future. The results of this paper can help taxi companies to better guide taxis to operate more efficiently based on real-time forecasting of passenger distribution and traffic conditions.

Acknowledgments:
This work is supported by the National Natural Science Foundation of China (Grant 71771009) and the Fundamental Research Funds for the Central Universities.

References
[1] Daqing, L., Yinan, J., Rui, K., & Havlin, S. Spatial correlation analysis of cascading failures: congestions and blackouts[J]. Scientific reports, 2014, 4.
[2] Li, D., Fu, B., Wang, Y., Lu, G., Berezin, Y., Stanley, H. E., & Havlin, S. Percolation transition in dynamical traffic network with evolving critical bottlenecks[J]. Proceedings of the National Academy of Sciences, 2015, 112(3): 669-672.
[3] Liu, M., Li, D., Qin, P., Liu, C., Wang, H., & Wang, F. Epidemics in Interconnected Small-World Networks[J]. Plos One, 2015, 10(3):e0120701.
[4] Liu, C., Li, D., Zio, E., & Kang, R. A Modeling Framework for System Restoration from Cascading Failures[J]. Plos One, 2014, 9(12):e112363.
[5] Liu, C., Li, D., Fu, B., Yang, S., Wang, Y., & Lu, G. Modeling of self-healing against cascading...
overload failures in complex networks [J]. Epl, 2014, 107(6):68003.
[6] Li, D., Zhang, Q., Zio, E., Havlin, S., & Kang, R. Network reliability analysis based on percolation theory [J]. Reliability Engineering & System Safety, 2015, 142: 556-562.
[7] Zhao, J., Li, D., Sanhedrai, H., Cohen, R., & Havlin, S. Spatio-temporal propagation of cascading overload failures in spatially embedded networks [J]. Nature communications, 2016, 7: 10094.
[8] Manski C F, Wright J D. Nature of equilibrium in the market for taxi services [J]. Transportation Research Record, 1967.
[9] Douglas, G. W. Price regulation and optimal service standards: The taxicab industry [J]. Journal of Transport Economics and Policy, 1972: 116-127.
[10] Shreiber, C. The economic reasons for price and entry regulation of taxicabs: a rejoinder [J]. Journal of Transport Economics and Policy, 1981, 15(1): 81-83.
[11] Beesley, M. E., & Glaister, S. Information for regulating: the case of taxis [J]. The economic journal, 1983, 93(371): 594-615.
[12] Cairns, R. D., & Liston-Heyes, C. Competition and regulation in the taxi industry [J]. Journal of Public Economics, 1996, 59(1): 1-15.
[13] Yang, H., & Wong, S. C. A network model of urban taxi services [J]. Transportation Research Part B: Methodological, 1998, 32(4): 235-246.
[14] Wong, K. I., Wong, S. C., & Yang, H. Modeling urban taxi services in congested road networks with elastic demand [J]. Transportation Research Part B: Methodological, 2001, 35(9): 819-842.
[15] Yang, H., Wong, S. C., & Wong, K. I. Demand–supply equilibrium of taxi services in a network under competition and regulation [J]. Transportation Research Part B: Methodological, 2002, 36(9): 799-819.
[16] Yang, H., Ye, M., Tang, W. H. C., & Wong, S. C. A multiperiod dynamic model of taxi services with endogenous service intensity [J]. Operations research, 2005, 53(3): 501-515.
[17] Mu, C., Xuan, H. Y., & Zhang, F. Simulation Study on Impacts of Vacant Taxi Behavior on System Performance [J]. Journal of Systems Engineering, 2008, 23(5): 554-562.
[18] Mu, C., & Xuan, H. Y. A Multiple Searching Strategies Based Simulation Model for Taxi Services [J]. Journal of Systems & Management, 2009, 18(3): 355-360.
[19] Mu, C, & Zhao, X. M. Multi-Agent Simulation Model for Vacant Taxi Searching Behavior [J]. Natural Science Journal of Xiangtan University, 2012, 1: 025.
[20] Wong, R. C. P., Szeto, W. Y., Wong, S. C., & Yang, H. Modelling multi-period customer-searching behaviour of taxi drivers [J]. Transportmetrica B: Transport Dynamics, 2014, 2(1): 40-59.