

Urban Taxi Resource Optimization Using Probability Model and Cellular Automata

Shaocong Mo, Xiaotian Jin⁺, Tao Zheng[#], Jia Shuai, Yuntao Yang

School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070 China

Abstract. This study explores the supply-demand relationship in urban taxi service and provide the means to optimize taxi distribution. Flaws can be observed in current configurations of taxi resource, hindering the efficiency of public transport. We use principal component evaluation to summarize supply-demand relationship on a macro scale. A probability model based on cellular automaton is established to study the effects of existing subsidy programs, with regard to which we propose a new bounty-based subsidy scheme as the tool for optimization. The results of this study are consistent and can adapt to solve taxi distribution problems under different configurations.

Keywords: probability model; cellular automaton; taxi resource distribution

1. Introduction

Taxis are an important part of urban traffic. The difficulty of catching taxis is often caused by the poor optimization of resource distribution instead of resource shortage. Scholars all over the world have conducted insightful research regarding this issue. For example, Microsoft Asia Research Institute [1] presented a framework of regional discovery, in which passenger mobility and points of interest are used to determine the distribution and flow of taxi resource; Zhang et al [2] analysed and classified the flow patterns of taxis, leading to their discovery of taxis locational characteristics and that of passenger mobility; Liu Yu [3] analysed the relationship between pick-up and drop-off spots and locale in the city by collecting trajectory data of trips; Miwa T et al [4] proposed a method of predicting future taxi conditions using taxis and passengers relationship flow; a team [5] in Peking University assembled a probe system to obtain more reliable traffic information for taxis; and there are other methods of traffic data collection involving embedded systems like GPS associating themselves with taxis[6]. Currently, taxis are a decent complement to the core components of urban passenger transport, together with which they can complete the network of a comprehensive public transportation system. A proportion model was made by Shanghai Jiao-tong University [7], suggesting the optimal weight for each type of public transport in regard to historic data. This helps solving the problem of taxi resource distribution, but is not a perfect solution. To address this issue, Du Wei [8] proposed a simulation model of taxi market operation based on taxi service software, and analysed the distribution status when theses software services are present. Hu Yucheng et al [9,10] introduced cellular automaton into their model of taxi traffic flow. The innovation of our work is:(1) Proposal of a probability model related to the assessment indicators of taxi resource distribution;(2) Use of cellular automaton, combined with a probability model to simulate taxi behaviour and evaluate the effectiveness of subsidy schemes.

We seek to make improvements by introducing a new subsidy scheme for internet taxi service platforms.

⁺ Corresponding author. Tel.: +86 15172355097.

E-mail address: jinxiaotian2016@163.com.

[#] The original affiliation of the author is School of Computer Science and Technology, Wuhan University of Technology, now his affiliation is School of Foreign Languages, Wuhan University of Technology.

2. Analysis of Taxi Macroscopic Supply-Demand Matching Rate

This paper uses data obtained from two sources. The 2015 economic database of China Economic Information Network (CEI) [11], and the online taxi service platform of Didi & Kuaidi[12].

Dimensional reduction is achieved through PCA. In total, 27 chosen cities are analysed and are scored. Considering regional differences, we run regression analysis for score values and indicators to obtain a city's overall traffic intensity in relation to its taxi supply-demand matching rate.

1. Comprehensive evaluation method based on principal components: Using raw data from the 27 cities and 10 indicators, a matrix $A = (a_{ij})_{27 \times 10}$ is constructed, where a_{ij} is the value of indicator j in city i .

(1) The raw values are standardized as follows $\tilde{a}_{ij} = (a_{ij} - a_j) / s_j, i = 1, \dots, 27, j = 1, \dots, 10$, a_j represents the average value of indicator j in the cities, $a_j = (\sum_{i=1}^n a_i) / n, (n = 10)$, and s_j is the sample's standard deviation for indicator $j, s_j = \sqrt{\sum_{i=1}^n (a_{ij} - a_j)^2 / (n - 1)}$. The corresponding \tilde{x}_j is considered a standardized indicator variable, hence $\tilde{x}_j = (x_j - a_j) / s_j, j = 1, 2, \dots, 10$, where refers to the indicator. Correlation coefficient matrix $R = (r_{ij})_{10 \times 10}$ is then calculated $r_{ij} = \sum_{k=1}^{27} \tilde{a}_{ki} \cdot \tilde{a}_{kj} / (n - 1), i, j = 1, 2, \dots, 10$.

(2) Eigenvalues and eigenvectors comes next, where $\lambda_1, \lambda_2, \dots, \lambda_{10}$ are the eigen-values of the correlation coefficient matrices, while u_1, u_2, \dots, u_{10} are the matrices' eigenvectors. Hence $u_j = [u_{1j}, u_{2j}, \dots, u_{10j}]^T$, and 10 index factors formed by the eigen-vectors can be obtained

$$\begin{cases} y_1 = u_{11}\tilde{x}_1 + u_{21}\tilde{x}_2 + \dots + u_{m1}\tilde{x}_m \\ y_2 = u_{12}\tilde{x}_1 + u_{22}\tilde{x}_2 + \dots + u_{m2}\tilde{x}_m \\ \vdots \\ y_m = u_{1m}\tilde{x}_1 + u_{2m}\tilde{x}_2 + \dots + u_{mm}\tilde{x}_m \end{cases}, m = 10 \quad (1)$$

(3) Principle components are then chosen, based on which the overall score shall be calculated. $d_j (j = 1, 2, \dots, 10)$ is used to represent the contribution rate of an eigenvalue λ_j , where h_p is the accumulated contribution rate of the previous p number of indicators. Thus $d_j = \lambda_j / \sum_{k=1}^{10} \lambda_k, j = 1, 2, \dots, 10, h_p = \sum_{k=1}^p \lambda_k / \sum_{k=1}^{10} \lambda_k$, the value of h_p approaches 1, the accumulated contribution rate of the principle component grows accordingly. The first number of principle components are selected to replace the 10 original indicator variables. Finally, evaluation based on these principle components are conducted to calculate the overall score $Z = \sum_{j=1}^p d_j y_j$.

2. City evaluation results: 4 principle components are obtained by the end of evaluation, and used eigenvectors to construct the mathematical expression for it. This expression of our evaluation model is weighed according to the principle components' contribution rate $Z = 0.5500y_1 + 0.1661y_2 + 0.1075y_3 + 0.1021y_4$. When the yearly values for these for components are input into the above formula, the results of the comprehensive evaluation for the year can be found in Table.1.

Table.1: Comprehensive Evaluation Results and Ranks

City	Beijing	Shanghai	Guangzhou	Shenzhen	Tianjin	Shijiazhuang	Hohhot	Chengdu	Hangzhou
value	4.905	3.131	1.444	0.719	0.314	0.152	0.111	0.002	-0.099
City	Nanjing	Wuhan	Xi'an	Harbin	Shenyang	Jinan	Changsha	Zhengzhou	Nanning
value	-0.140	-0.152	-0.210	-0.239	-0.278	-0.280	-0.436	-0.516	-0.595
City	Kunming	Guiyang	Fuzhou	Taiyuan	Yinchuan	Changchun	Lanzhou	Hefei	Nanchang
value	-0.611	-0.723	-0.741	-0.742	-0.890	-0.968	-0.984	-1.071	-1.104

3. Relationship between regional difference and supply-demand matching rate: The differing results shows how regional difference can impact the evaluation process [13]. For each city, multiple linear regression analysis is conducted for the score value Z , the ownership rate p and the mileage utilization rate q . (Data for the above two indicators are also from CEI's database.) Valid regression analyses for 11 cities are completed for the following results $Z = 0.7776p - 0.2843q$. Where $R^2 = 0.8622, p = 0.0190 < 0.05$. A city's score is positively correlated with taxi ownership rate (the coefficient being 0.7776) and negatively correlated with mileage utilization rate (the coefficient being -0.2843). This is because taxi ownership rate should rise when a city has more complex traffic conditions, in which case the extra taxis can dilute the market and lower the

utilization efficiency. This is a typical case where taxi supply and demand are not well matched. Hence, matching rate should fall as a city's traffic intensity grows.

3. Robabilistic Selection Model

After analysing taxi drivers' behavioural patterns, two main factors influencing their decisions are identified: Distance between them and the passenger, and how demanding the destination is for more taxis.

(1) In terms of distance, drivers are more likely to pick up closer passengers for convenience. If we use the grid vertices as the origins of coordinate systems to describe taxi locations, use x, y to mark the abscissa and ordinate values respectively, and use α_k, β_k to represent the coordinate of the k^{th} location, then S_k , the distanced travel by a taxi to reach the k^{th} location can be calculated $S_k = \sqrt{(\alpha_k - x)^2 + (\beta_k - y)^2}$, Let φ_k be the probability for a taxi driver to answer a call on the k^{th} location $\varphi_k = S_k^{-1} / \sum_{i=1}^q S_i^{-1}$, q is the total number of locations where passengers are calling in for taxis.

(2) In terms of destination's demand, taxi drivers tend to go to passenger-rich destinations, which can be translated into extra business opportunity. We include the number of calls into calculation. Let be the probability for a taxi driver to answer a call on the location, then we have

$$\gamma_k = A(\alpha_k, \beta_k) / \sum_{i=1}^q A(\alpha_i, \beta_i) \quad (2)$$

$A(\alpha_k, \beta_k)$ is the total population of all k number of locations, and q is the number of locations with passengers sending in calls. After combining the two factors above, we get the elaborate probability function that shows the possibility for drivers to reach location k , not accounting the subsidies provided by service platforms:

$$\zeta = \omega_1 \varphi_k + \omega_2 \gamma_k \quad (3)$$

The Cellular Automaton (CA) model is commonly used for traffic analysis [14,15]. During the generation process, we use "1" to mark a full cell, and "0" to mark an empty cell. We run the automaton using the following rules based on probability model. A square region is constructed and divided into 10×10 grids. The grids, each representing a cell, are labelled from 1 to 100. A random number between 1 and 100, is generated to be the total number of calling passengers, then the same number of locations are generated. Each location would make a number of demanding calls according to its population. Fig.1 depicts how the automaton works.

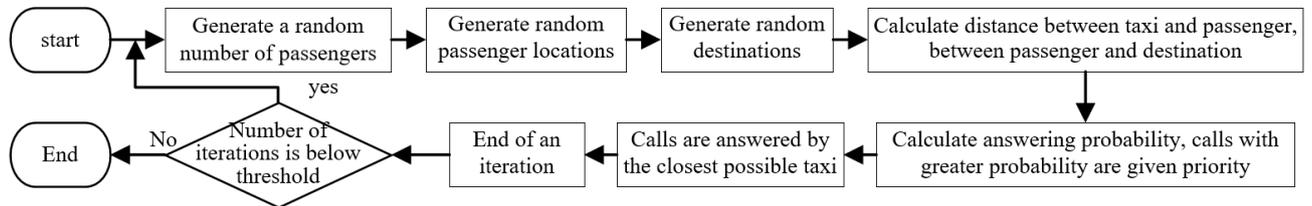


Fig.1: Iteration process of the cellular automaton

4. Relationship between Subsidy Schemes and Taxi Resource Distribution

4.1. Subsidy Schemes of Different Companies

We choose the schemes used by two major taxi service platforms as our research object. They are ① Didi & Kuaidi, who offers drivers a 1.5x multiplier on trip fare ② Yaoyao Dache, who gives drivers a subsidy of ¥3 per kilometre.

Analysis shows that the 1.5x multiplier is more rewarding over long distances, thus encouraging drivers to run longer trips and lowering the rate of service denial. While the ¥3 per kilometre subsidy allows drivers to fetch passengers from afar without wasting mileage, thereby making taxi distribution in the city more even, and eases the difficulty of catching a taxi at otherwise obsolete areas.

4.2. Calculation of weights under different schemes

The Coefficient of Variation (CV) method is used to determine the weights by inputting the values of different indicators into the model. Let u_i be the variation coefficient of different indicators, σ_i be the standard deviation of the i number of indicators, and \bar{x} be the average value of the number of indicators, then ,

The weight of each indicator can thus be confirmed: $\omega_i = u_i / \sum_{i=1}^m u_i$, ω_i is the weight of indicator i , while m is the number of indicators. Using the CV method, we find the weight of each indicator as shown in Table.2.

4.3. Proposed Design for Subsidy Scheme

With the help of survey results, we find that when passengers are willing to pay 50% more for their taxi fares, no less than 93% of taxi drivers are willing to enter the centre areas, even at the rush hour. Based on these results, we propose a bounty-style subsidy scheme to lower the difficulty of catching taxis by rewarding drivers who operates at demanding regions. For the drivers, there may be multiple areas of interest that possess bounties. Hence the probability selection function for the drivers is

$$\psi = \left| \min(T(\alpha_k, \beta_k) - A(\alpha_k, \beta_k), 0) \right| / \sum_{j=1}^q \left| \min(T(\alpha_j, \beta_j) - A(\alpha_j, \beta_j), 0) \right| \quad (4)$$

Hence the final selection function for taxi drivers is $\zeta = \omega_1 \varphi_k + \omega_2 \gamma_k + \omega_3 \psi_k$. For this function, the same CV method is used to determine weights, and the results are $\omega_1 = 0.3, \omega_2 = 0.1, \omega_3 = 0.6$.

4.4. Simulated results for different schemes

Beijing's taxi distribution without subsidy schemes is simulated first. The probability model based on cellular automaton can be used to find the difference between supply and demand. If the value is negative, then a case of insufficient supply exists, and vice versa. The first image (up left) in Fig.2 shows supply-demand relationship without subsidies, taken after an average period of simulation time. The higher the grid value, the more asymmetrical the supply and demand is. Without the subsidies, it is apparent that central areas have more supply than demand, while marginal areas have more demand than supply. The explanation for this is that population is scarce around the border, which may also be inconvenient to reach.

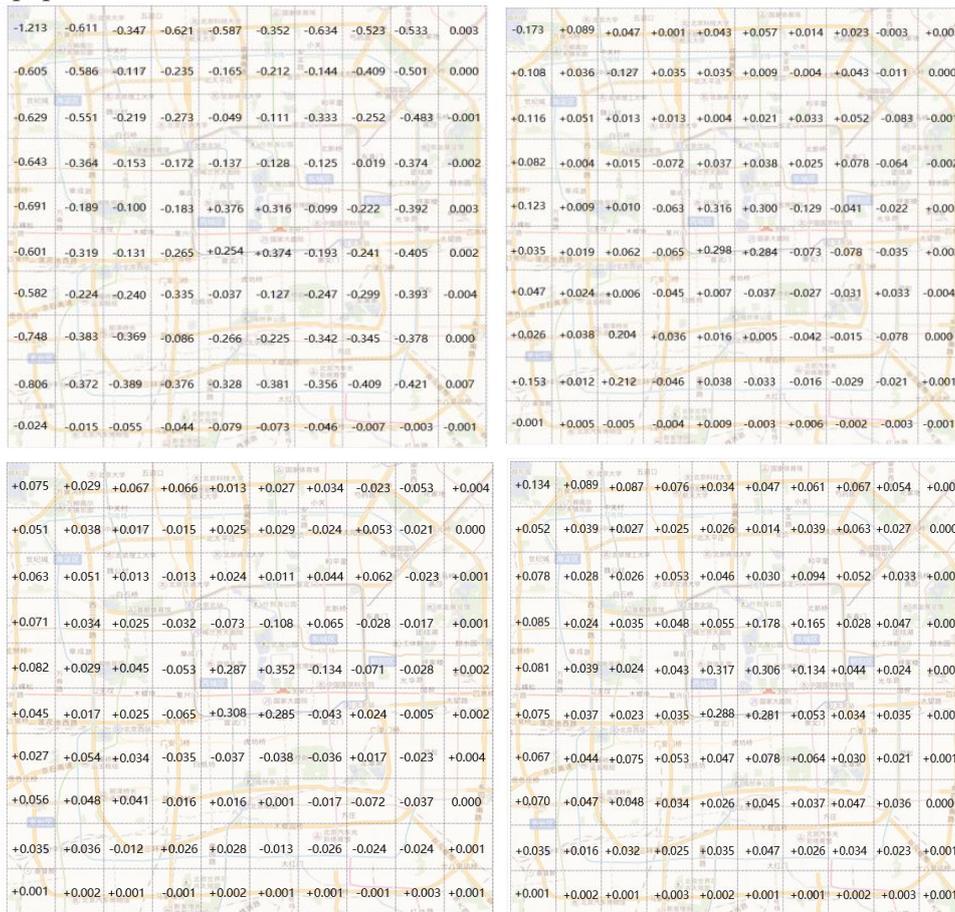


Fig.2: Supply-demand relationship with and without subsidies in effect

With the “1.5x multiplier” scheme in effect, central areas are still over-supplied, yet the shortage in most parts of the city are fixed to a certain extent. Results for this scheme can be found in the second image (up right) in Fig.2. Certain previously undersupplied locations have more than enough supply now. Around the borders, the changes are significant, the supply-demand relationship turned around and oversupplying is now very common. It is inferred that the scheme encouraged drivers to run long-distance routes and lower the rate of passengers getting rejected. With the ¥3/km scheme implemented, central areas are well supplied, and

certain marginal areas are also abundantly supplied. Results are displayed in the third image (down left) in Fig.2. On the whole, very few locations are under-supplied, showing how taxi distribution has been optimized.

The results are obtained from simulation after an average time of simulation. The algorithm assumes that all taxis begin their day from the city center, which contributes to the fact that the center is always oversupplied. However, that does not hinder the conclusion that both subsidy schemes can ease the difficulty of catching taxis, but they are not perfect, for they cannot solve the problem of oversupply at the city center. A comparison of the 3 cases are shown in the top 3 rows in Table.3.

As mentioned in the section above, we design a new bounty-based subsidy scheme to be added into the probability model and repeat the simulation. With the new scheme implemented, most parts of the city are now oversupplied, and the distribution is more even. Calculations show that the average supply-demand difference is now 600 per 10,000 passengers, while the variance of supply-demand difference is 0.07. Compared with other schemes, taxi resource distribution is now better optimized. The results are in the bottom row of Table.3 in comparison with previous cases.

All schemes are capable of easing the difficulty of catching taxis, and can redistribute taxis more evenly. However, the bounty-based scheme works the best. Contrary to the two existing schemes, when the new scheme is in effect, most cells have positive value, indicating better distribution of taxi resource.

Table 2: Indicator weight under different schemes

Scheme	ω_1	ω_2	ω_3
1.5x multiplier	0.35	0.3	0.35
¥3 / km	0.3	0.4	0.3

Table 3: Subsidy schemes' impact on taxi catching difficulty

Scheme	Average Supply-demand difference	Variance of Supply-demand difference
None	-0.20	0.23
1.5x multiplier	0.12	0.08
¥3 / km	0.08	0.08
New scheme	0.06	0.07

5. Conclusion

Comprehensive analysis was conducted on the supply-demand matching rate of taxi resource in 27 cities and distribution patterns are summarized. Macroscopically, the matching rate of major metropolises are more centralized than that of smaller cities. When a city's traffic intensity increases, the supply-demand matching rate would fall. Specifically, using Beijing as an example, we find that regions with a high matching rate tends to be the functionally condensed parts, which are affected greatly by rush hours. The effect of existing subsidy schemes of online taxi service is analysed and confirm to be capable of scattering overcrowded taxi concentration around city centres and even out overall taxi distribution. We introduce a new bounty-based subsidy scheme to solve problems and to further optimize the distribution of taxis.

The proposed probability model based on cellular automaton considers factors such as trip distance and passengers' demand. It can be of significance for its analytical use, and is applicable to the analysis of taxi resource distribution under the influence of various policies.

6. Acknowledgements

We thank Dr. Pengfei Duan for valuable discussion. This paper is supported by the Fundamental Research Funds for the Central Universities (#2017-JSJ-A1-01, #2017-JSJ-B1-05).

7. References

- [1] Yuan, Jing, Y. Zheng, and X. Xie. "Discovering regions of different functions in a city using human mobility and POIs." *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ACM*, 2012:186-194.
- [2] Zhang, Wangsheng, S. Li, and G. Pan. "Mining the semantics of origin-destination flows using taxi traces." *ACM Conference on Ubiquitous Computing ACM*, 2012:943-949.

- [3] Liu, Yu, et al. "Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai." *Landscape & Urban Planning*, 106.1(2012):73-87.
- [4] Miwa, Tomio, et al. "Allocation planning for probe taxi devices based on information reliability." *Transportation Research Part C Emerging Technologies*, 34.34(2013):55-69.
- [5] Moreira-Matias, Luis, et al. "Predicting Taxi–Passenger Demand Using Streaming Data." *IEEE Transactions on Intelligent Transportation Systems*, 14.3(2013):1393-1402.
- [6] Zhang, Wei, X. U. Jian-Min, and Z. Zhang. "The Dynamic Traffic Information Collection and Processing Methods of ITS Common Information Platform Based on GPS Equipped Taxi Data." *Journal of Highway & Transportation Research & Development* (2007).
- [7] Lu, Linjun, et al. "An Optimal Allocation Model of Public Transit Mode Proportion for the Low-Carbon Transportation." *Mathematical Problems in Engineering*, 2015(2015):1-8.
- [8] Wei, D. U., et al. "Simulation Model for Taxi Service Market Equipped with Taxi-calling Apps." *Journal of Transportation Systems Engineering & Information Technology* (2016).
- [9] Hu, Yucheng, et al. "City traffic forecasting using taxi GPS data: A coarse-grained cellular automata model." (2016).
- [10] Xie, Siqi, S. C. Wong, and W. H. K. Lam. "Microscopic decision model for pedestrian route choice at signalized crosswalks." *Journal of Advanced Transportation*, 50.6(2016):1181-1192.
- [11] CEInet Statistics Database [DB/OL] <http://db.cei.gov.cn/page/Default.aspx>
- [12] Han, Zhonggeng, and D. U. Jianping. "Comments on Taxi Allocation in the "Internet+" Era." *Mathematical Modeling & Its Applications* (2015).
- [13] Poeschl, M, S. Ward, and P. Owende. "Environmental impacts of biogas deployment - Part I: Life cycle inventory for evaluation of production process emissions to air. " *Journal of Cleaner Production*, 24.3(2012):168-183.
- [14] Kai, Nagel, and M. Schreckenberg. "A cellular automaton model for freeway traffic." *Journal De Physique I*, 2.12(1992):2221-2229.
- [15] Leng, Biao, et al. "Analysis of Taxi Drivers' Behaviors Within a Battle Between Two Taxi Apps." *IEEE Transactions on Intelligent Transportation Systems*, 17.1(2015):296-300.