Non-Linear Analysis of Influencing Factors of Commercial Vitality Within the Radiation Range of Metro Stations

Mingchao Liang[1], Mei Liu[2]

[1] Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai, China
[2] China Institute of FTZ Supply Chain, Shanghai Maritime University, Shanghai, China

202130510208@stu.shmtu.edu.cn (Mingchao Liang)
meiliu@shmtu.edu.cn (Mei Liu)

Abstract. Relying on the strong radiation ability of urban rail transit, the areas near the stations have become the agglomeration area that attracts commercial facilities and consumption behaviors. In this paper, five metro stations in the main urban area of Shanghai are selected as research stations. The walking radiation range of each station is defined by utilizing the AMap API (Application Programming Interface) and the commercial vitality is quantified from the dual perspectives of physical space and pedestrian flow. Then, the influencing factor system of commercial vitality is established. Finally, a gradient boosting decision tree model based on Bayesian optimization is constructed to explore the main influencing factors of commercial vitality. The research results suggest that commercial function richness is the most important influence on commercial vitality, followed by public facilities density, and the importance of the two is more than half of the overall importance of all the factors. The presence of diverse commercial POIs types, sufficient public service facilities, and convenient metro accessibility to the station are crucial factors in ensuring commercial vitality within the vicinity of metro stations. Additionally, the study reveals non-linear relationships between the influencing factors and commercial vitality, with the presence of threshold effects for certain factors.

Keywords: metro stations; commercial vitality; gradient boosting decision tree; non-linear; spatial distribution

1 Introduction

Under the background of transit metropolis, urban rail transit has become the skeleton of urban public transportation system. In 2022, the share ratio of urban rail transit passenger volume to the total public transport passenger volume in China is 45.82%, and the share ratio in Shanghai has exceeded 70%[1]. Urban rail transit not only ensures people's green and convenient travel, but also promotes the integration of people's travel

T. Yao et al. (eds.), *Proceedings of the 2024 3rd International Conference on Engineering Management and Information Science (EMIS 2024)*, Advances in Computer Science Research 111,
https://doi.org/10.2991/978-94-6463-447-1_34

habits into their daily lifestyles, adding vitality to the city's social activities such as commerce and life.

The relationship between urban rail transit and commerce along the route has always been the focus of relevant research. Wang et al. [2] analyzed the correlation among accessibility, passenger flow and commercial spatial structure in areas near 15 metro stations in Guangzhou. Zheng et al. [3] found that new metro stations have positive contributions to the quantity, diversity and consumer demand of catering services in the surrounding areas. Sonnen-schein et al. [4] concluded that the expansion of urban subway has a significant positive impact on the density and versatility of facilities around stations. In addition to considering the impact of the station on the commercial development of the surrounding area, researchers also pay attention to the impact of urban rail transit on commercial agglomeration along the line[5-6] and the development of underground commercial space of the station[7-8]. However, studies that specifically investigate the relationship between commerce and construction factors in the vicinity of the station are limited. Urban rail transit's strong radiation capacity attracts a significant influx of people, goods, and information to the station area, leading to the aggregation of commercial elements in the vicinity. Therefore, it is crucial to explore the relationship between commercial vitality and building factors in the region.

2 Study Area and Data

2.1 Study Area

Shanghai, located at the estuary of the Yangtze River, functioning as China's center of economy, finance and trade. In 2022, Shanghai's annual GDP reached 4.47 trillion yuan, and the total retail sales of consumer goods amounted to 1.64 trillion yuan. There are 20 urban rail transit lines with 936.17 km of operation, and a total of 486 stations are distributed along these lines[1]. **Fig. 1** illustrates five metro stations located in the main urban area of Shanghai that serve as the research stations for this paper.

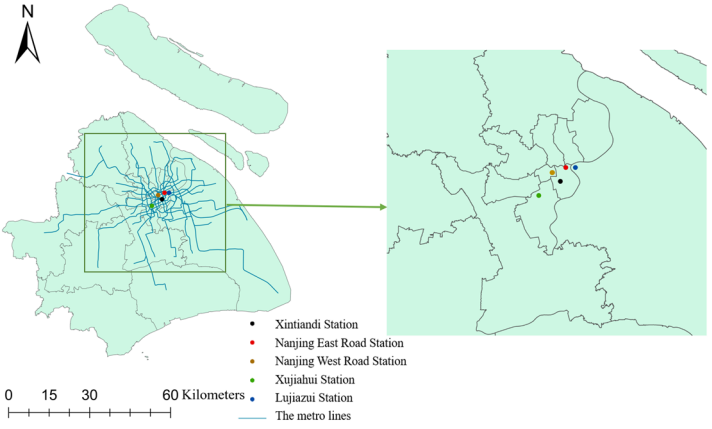


Fig. 1. Distribution of the target metro stations[1]

2.2 Data

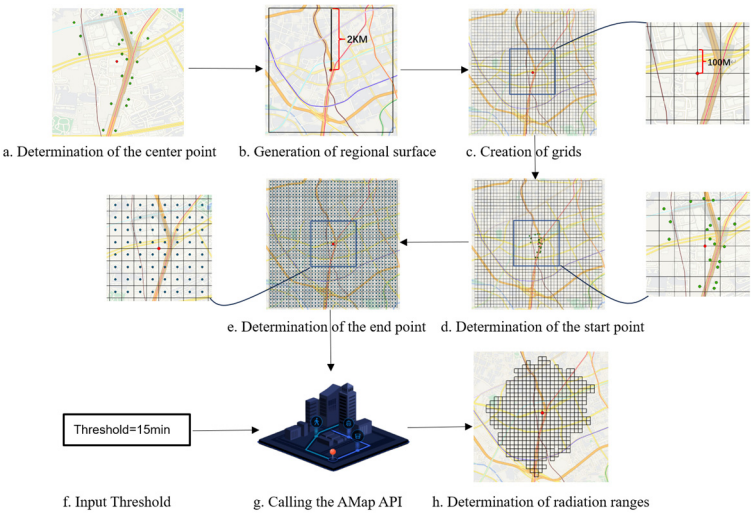
Table 1. Relevant information about the research data[2]

Name of Data	Source	Version-in-time
Administrative boundaries	National catalogue service system for geographic information	2017
Metro stations and routes	AMap API	July 2023
Planning of walking routes	AMap API	December 2022
POIs	AMap API	December 2022
Heat map	Baidu map API	December 2020
Bus stops	AMap API	December 2023
Green spaces	AMap API	December 2023
Building footprint	National Tibetan plateau data center	2020

Table 1 displays the relevant information of the research data. With reference to the categorization of POIs in AMap, the data of commercial POIs including catering, shopping, life, leisure, and accommodation, as well as the data of public facilities, are captured. The walking route planning data is utilized to determine the radiation range of each research station, while the commercial POIs and heat map data are used to evaluate commercial vitality.

3 Methodologies

3.1 Threshold-Based Point Distribution Method



[1] Figure source: Author's self-made
[2] Table source: Author's self-made

Fig. 2. The process for defining the radiation range for Xujiahui Station as an example[3]

After determining the research metro stations, the next fundamental step is to identify the specific study boundaries, i.e. the radiation range of each metro station. The threshold-based point distribution method is applied to determine each study station's radiation range, as shown in **Fig. 2**.

3.2 Hill Number

Table 2. Definition and sample statistics of influence factors[4]

Factor	Data	Description	Minimum	Maximum	Mean
Bus accessibility	Bus stops	Distance from grid cell to the nearest bus stop	2	452	136
Metro accessibility	Walking route planning	Walking time from station to grid cell	1	900	540
Business function richness	Commercial POIs	Hill number	0	5	2.69
Business function mixedness	Commercial POIs	Hill number	1	4.86	2.4
Building density	Building footprint	Ratio of building area to grid cell area	0	0.77	0.25
Plot ratio	Building footprint	Ratio of total building area to grid cell area	0	72.71	2.16
Public facilities density	Public facilities POIs	Number of public facilities POIs in grid cell	0	44	0.72
Green space accessibility	Green spaces	Distance from grid cell to the nearest green space	0	954	300

As shown in **Table 2**, the influencing factors of this paper include eight aspects: bus accessibility, metro accessibility, commercial function richness, commercial function mixedness, building density, plot ratio, public facilities density and green space accessibility. Wherein, the commercial function richness and mixedness are calculated by Hill number, the formula is as follows:

$$\begin{cases} {}^0D = \sum_{i=1}^{s} p_i^0 \\ {}^1D = \exp\left(-\sum_{i=1}^{s} p_i \ln p_i\right) \end{cases} \quad (1)$$

Where: p_i represents the proportion of class i commercial POI to the total number of commercial POI in grid cells, s is the number of commercial POI categories; ${}^0D$ represents the commercial function richness of grid cells; ${}^1D$ represents the commercial function mixedness of grid cells.

[3] Figure source: Author's self-made

[4] Table source: Author's self-made

3.3 Gradient Boosting Decision Tree Based on Bayesian Optimization

Based on the principles of statistical and machine learning methods, the gradient boosting decision tree algorithm combines decision trees with gradient boosting. By constructing multiple weak learners and iteratively combining them, a strong learner is formed. Each iteration improves upon the previous result, reducing the prediction error of the model, and constructing a new model in the direction of the gradient descent of the previously established model loss function. The formula is as follows:

$$f_M(x) = \sum_{m=1}^{M} T(x:\theta_m) = \sum_{m=1}^{M} \sum_{j=1}^{J} c_{mj} I(x \epsilon R_{mj}) \tag{2}$$

Where: $T(x:\theta_m)$ represents the decision tree; m is the number of iterations, and M is the maximum number of iterations; j is the number of leaf nodes, and J is the total number of leaf nodes; c_{mj} is the predicted value of the output of iterative leaf node j in the m round; I is the indicator function, which takes 1 if x belongs to R_{mj} and 0 otherwise; R_{mj} is the eigenvalue of the training sample of iterative leaf node j in the m round.

The gradient boosting decision tree regression model can provide the ranking of independent variables based on their contribution degree. Additionally, the model can generate partial dependence plots to illustrate the relationship between each independent variable and the dependent variable. Some studies have demonstrated that the gradient boosting decision tree regression model offers advantages in terms of efficiency, accuracy, and interpretability compared to linear regression models[9-10].

However, the gradient boosting decision tree has many parameters, and whether the parameters can be efficiently adjusted is very important to the performance of the model. In this paper, Bayesian optimization is employed to adjust the parameters of the gradient lifting decision tree model, and the hyperopt library in the Python programming language is utilized to implement Bayesian optimization. The surrogate model in Hyperopt library is Tree-structured Parzen Estimator (TPE) with the following formula:

$$p(x|y) = \begin{cases} \ell(x) & if\ y < y^* \\ g(x) & if\ y \geq y^* \end{cases} \tag{3}$$

Where: $\ell(x)$ is the probability distribution formed using the observation space $\{x^{(i)}\}$; $g(x)$ is the probability distribution formed using the remaining observation space; y is the true loss value; y^* represents the loss threshold.

After constructing the surrogate model, it is necessary to evaluate the sample data through the acquisition function and select the best x^* to pass to model. The acquisition function of Hyperopt is Expected improvement (EI), and its formula is as follows:

$$\begin{cases} EI_{y^*}(x) = \int_{-\infty}^{y^*} (y^* - y)p(y|x)dy = \int_{-\infty}^{y^*} (y^* - y)\dfrac{p(x|y)p(y)}{p(x)}dy \\ \int_{-\infty}^{y^*} (y^* - y)p(x|y)p(y)dy = \ell(x)\int_{-\infty}^{y^*} (y^* - y)p(y)dy = \gamma y^*\ell(x) - \ell(x)\int_{-\infty}^{y^*} p(y)dy \\ p(x) = \int p(x|y)p(y)dy = \gamma\ell(x) + (1-\gamma)g(x) \\ EI_{y^*}(x) = \dfrac{\gamma y^*\ell(x) - \ell(x)\int_{-\infty}^{y^*} p(y)dy}{\gamma\ell(x) + (1-\gamma)g(x)} \end{cases} \tag{4}$$

It can be seen that $EI_{y^*}(x) \propto (\gamma + \frac{g(x)}{\ell(x)}(1-\gamma))^{-1} \propto \frac{\ell(x)}{g(x)}$, in order to maximize $EI_{y^*}(x)$, it is necessary to find the parameter point x^* with the largest $\ell(x)/g(x)$.

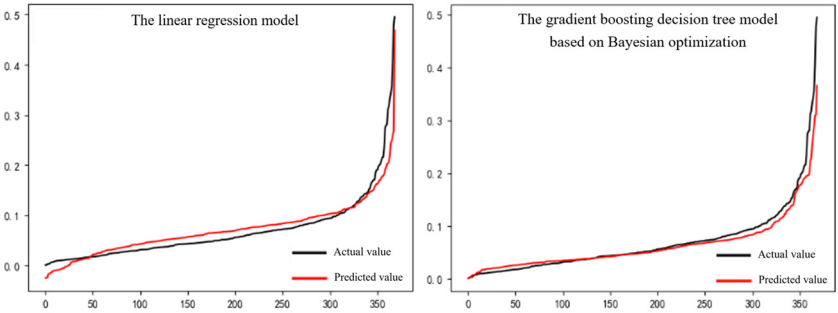


Fig. 3. Comparison of the fitting results of each model[5]

As shown in **Fig. 3**, the fitting effect of the gradient boosting decision tree model based on Bayesian optimization is significantly better than that of the linear regression model. The key parameters of the model in this paper are set as follows: the maximum number of iterations is 150, the shrinkage parameter is 0.05, the maximum tree depth is 5, the loss function is selected as the Huber loss, and the branching method is selected as the mean squared error. Finally, the R-squared score of the model on the test set is 0.68, and the relative importance of the independent variables, as well as the partial dependence graph, are generated for further analysis and discussion.

[5] Figure source: Author's self-made

4 Analysis Results

4.1 Radiation Range of Metro Stations

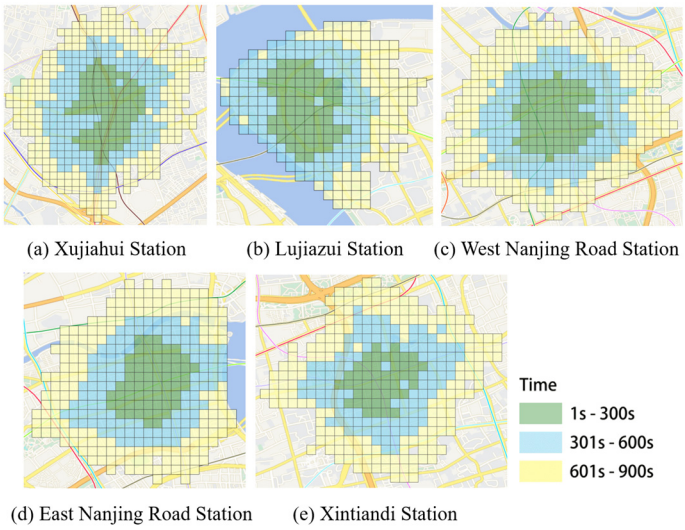


Fig. 4. The spatial distribution of radiation range of the research stations[6]

It is evident from **Fig. 4** that that the radiation ranges of urban metro stations do not conform to a regular circle. Due to the differences in the number of station entrances and the layout of the surrounding road network, the size and spatial distribution of the radiation range of different stations are also diverse. The radiation range of Xujiahui and Nanjing West Road station is significantly higher than other stations, and the overall spatial distribution of Nanjing West Road station is uniform, while the distribution uniformity of Xujiahui station is not well within 0-5 minutes. The area of Nanjing East Road and Xintiandi station is similar, but the overall spatial distribution uniformity of Xintiandi station is poor. The radiation range of Lujiazui station is greatly affected by the water area, and its range is the smallest among all research stations.

4.2 Importance Analysis of Influencing Factors

For the gradient boosting decision tree regression model, different influencing factors have different levels of importance for commercial vitality. **Fig. 5** gives the importance of all influencing factors for commercial vitality. The importance of commercial function richness is the highest, reaching 0.293, and the importance of public facility density is 0.210, second only to commercial function richness. Metro accessibility is 0.125, and commercial function mixedness is 0.102. The importance of other influencing factors

[6] Figure source: Author's self-made

is not less than 0.100, in descending order of importance, are: green space accessibility, bus accessibility, plot ratio and building density.

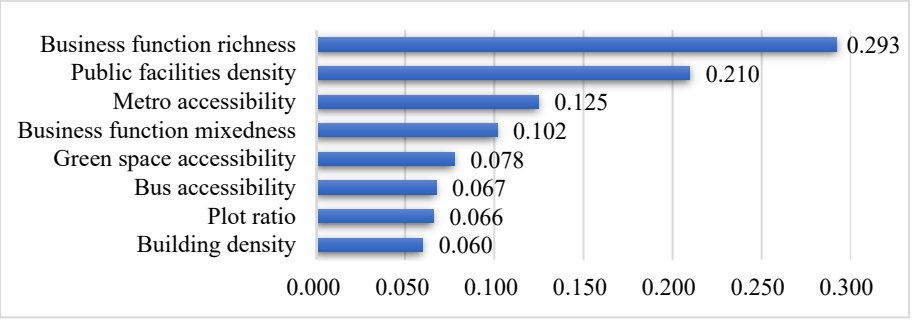


Fig. 5. Importance of influencing factors[7]

4.3 Non-Linear Analysis Between Commercial Vitality and Influencing Factors

By holding other features constant and observing the impact of changes in one feature on the model, the partial dependence graph can reflect how the feature affects the model prediction. This study utilizes partial dependence plots to illustrate the relationship between the influencing factors and the predicted commercial vitality value.

Within the range of 0-30 meters, there is a sharp decline in commercial vitality as the distance between the grid cell and the nearest bus stop increases. Once the distance exceeds 30 meters, commercial vitality fluctuates slightly (**Fig. 6** (a)). Similarly, within the range of 0-300 seconds, there is a significant drop in commercial vitality as the walking time increases. However, once the walking time exceeds 300 seconds, the change in commercial vitality becomes less noticeable (**Fig. 6** (b)). These findings indicate that the vicinity of bus stops and metro stations is conducive to the concentration of commercial facilities and human traffic, creating a vibrant commercial space. Commercial vitality increases significantly when commercial function richness exceeds 2 (**Fig. 6** (c)). This indicates that with the increase of commercial POIs types in grid cells, commercial vitality is significantly improved. The mixedness of commercial function shows a negative correlation with commercial vitality across the entire range (**Fig. 6** (d)). This indicates that more disorderly distribution of commercial POIs in the main urban area is less favorable for promoting commercial vitality. Increasing building density is highly effective in promoting commercial vitality within the range of 0.3-0.5 (**Fig. 6** (e)). Similarly, within the range of 0-10, increasing plot ratio has a significant positive impact on commercial vitality (**Fig. 6** (f)). These findings indicates that the promotion effect of building density and plot ratio on commercial vitality is limited to specific ranges. When the density of public facilities ranges from 0 to 25, commercial vitality increases rapidly. Once the density of public facilities exceeds 25, commercial vitality shows a declining trend (**Fig. 6** (g)). This suggests that a higher density of public

[7] Figure source: Author's self-made

facilities contributes to the improvement of commercial vitality, but blindly increases in density have little impact on its development. Commercial vitality continues to increase as the distance between grid units and the nearest green space increases (**Fig. 6** (h)). This indicates that the high-vitality commercial area around metro stations in the main urban area generally has a long distance from the green space.

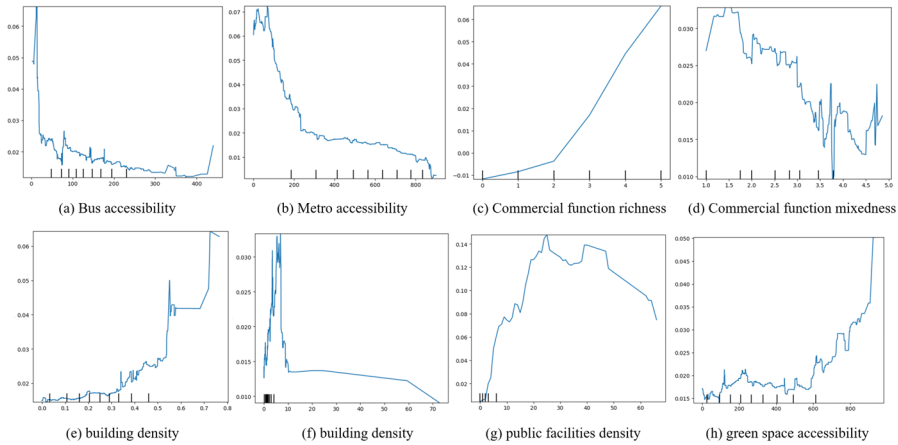


Fig. 6. Partial dependence plots of each influencing factor[8]

5 Conclusions

This paper explores the nonlinear relationship between influencing factors and commercial vitality through a gradient boosting decision tree model based on Bayesian optimization. The main conclusions are as follows:

(1) The radiation ranges of the target metro stations are delimited by the AMap API proved that the radiation ranges of urban metro stations do not necessarily conform to regular circles, which is also observed in the previous studies[11-12]. The size and spatial distribution of the radiation range of different stations are different, mainly depending on the distribution characteristics of the built environment around the stations.

(2) The results of the gradient boosting decision tree regression model indicate that commercial function richness is the most influential factor on commercial vitality, followed by public facility density. The combined importance of these two factors to commercial vitality accounts for more than half of the overall importance of all considered factors. The remaining factors, in descending order of importance, are: metro accessibility, commercial function mixedness, green space accessibility, bus accessibility, plot ratio and building density.

(3) The partial dependence plots provide evidence of nonlinear relationships between the influencing factors and commercial vitality, with the identification of threshold effects for certain factors. The presence of diverse commercial POIs types,

[8] Figure source: Author's self-made

sufficient public service facilities, and convenient metro accessibility to the station are crucial factors in ensuring commercial vitality within the vicinity of metro stations.

This paper defines and quantifies the radiation range, commercial vitality, and influencing factors of the research stations. Based on this, the paper focuses on the nonlinear effect of the influencing factors on commercial activity within the radiation range of each station. The research results are helpful in understanding the impact of various construction factors on the commercial areas surrounding the station. In future research, we aim to expand the scope by including more metro stations from diverse urban centers in Shanghai. This will enable more in-depth analysis of the nonlinear effects of influencing factors on commercial vitality within the radiation range of the stations.

References

1. China Association of Metros. (2023) Urban rail transit 2022 annual statistics and analyses report. https://www.camet.org.cn/tjxx/11944.
2. Wang, Y., Chen, Z., Qin, S. (2015) Study on the relationship between metro station accessibility, passenger flow and retail business structure in the station business district: Taking Guangzhou as an example. Human Geography, 30(04): 66-71. http://doi.org/10.13959/j.issn.1003-2398.2015.04.011.
3. Zheng, S., Hu, X., Wang, J., Wang, R. (2016) Subways near the subway: Rail transit and neighborhood catering businesses in Beijing. Transport Policy, 51: 81-92. http://doi.org/10.1016/j.tranpol.2016.03.008.
4. Sonnenschein, T. S., Scheider, S., Zheng, S. (2022) The rebirth of urban subcenters: How subway expansion impacts the spatial structure and mix of amenities in European cities. Environment and Planning B: Urban Analytics and City Science, 49(4): 1266-1282. http://doi.org/10.1177/23998083211056955.
5. Han D., Wu S. (2023) The capitalization and urbanization effect of subway stations: A network centrality perspective. Transportation Research Part A: Policy and Practice, 176, Article 103815. http://doi.org/10.1016/j.tra.2023.103815.
6. Mejia-Dorantes L., Paez A., Vassallo J. (2011) Transportation infrastructure impacts on firm location: the effect of a new metro line in the suburbs of Madrid. Journal of Transport Geography, 22: 236-250. http://doi.org/10.1016/j.jtrangeo.2011.09.006.
7. Ma C., Peng F., Qiao Y., Li H. (2022) Evaluation of spatial performance of metro-led urban underground public space: A case study in Shanghai. Tunnelling and Underground Space Technology, 124, Article 104484. http://doi.org/10.1016/j.tust.2022.104484.
8. Xu Y., Chen X. (2023) Uncovering the relationship among spatial vitality, perception, and environment of urban underground space in the metro zone. Underground Space, 12: 167-182. https://doi.org/10.1016/j.undsp.2023.02.010.
9. Gao L., Chong H., Zhang W., Li Z. (2023) Nonlinear effects of public transport accessibility on urban development: A case study of mountainous city. Cities, 138, Article 104340. https://doi.org/10.1016/j.cities.2023.104340.
10. Yang J., Cao J., Zhou Y. (2021) Elaborating non-linear associations and synergies of subway access and land uses with urban vitality in Shenzhen. Transportation Research Part A: Policy and Practice, 144:74-88. https://doi.org/10.1016/j.tra.2020.11.014.
11. Tan, P., Mai, K., Zhang, Y., Tu, W. (2021) Delineation of metro station attraction range using multi-source urban data. Journal of Geographical Information Science, 23(04)**:** 593-603. http://doi.org/10.12082/dqxxkx.2021.200183.

12. Wang, J., Wan, F., Dong, C., Shao, C. (2023) Modelling of attraction range and intensity of urban metro stations. Journal of Jilin University, 53(02): 439-447. http://doi.org/10.13229/j.cnki.jdxbgxb20210667.