Evaluation of Super Large Airport Cooperative Operation State Based on Decision Tree Support Vector Machine

Menglu Li *, Ruicong Gao[a], Gan Chai[b]
[*] Corresponding author: 924015507@qq.com, [a]81507278@qq.com, [b]chai.gan@163.com

Southeast University, School of Transportation SEU, Nanjing, China

Abstract: In order to accurately evaluate the real-time cooperative operation state of super large airports, the evaluation index system of cooperative operation is designed from four aspects: external distribution efficiency, internal transfer efficiency, key traffic corridor state and information interaction efficiency. Based on the established index system, a decision tree SVM cooperative operation state evaluation model is established, and the model parameters are optimized using Bayesian optimization method and 5-fold cross validation method. A case study of Beijing Capital International Airport is carried out. The results show that the established evaluation model has high accuracy and good classification performance, and can solve the problem of cooperative operation state evaluation of high-dimensional and nonlinear large-scale airport systems. The proposed evaluation method reduces the impact of subjective factors on the reliability of the evaluation results, and is conducive to the objective evaluation of the real-time cooperative operation of airports and the optimal scheduling of transport capacity resources.

Keywords: Super Large Airport, cooperative operation, evaluation index system, decision tree support vector machine

1 INTRODUCTION

Air transport can serve the travel demand of large volume and long distance, and is one of the main travel modes among countries at present. In order to improve the service quality of air transport, many large airport hubs in China are building and developing into airport hub demonstration areas. Accurate evaluation of airport operation state is of great practical significance to the construction and operation of airport demonstration areas.

At present, there are many evaluation methods for small and medium-sized urban comprehensive transport hubs at home and abroad, and the evaluation index systems have their own advantages and characteristics, but there is a lack of evaluation index systems and methods for international airports such as Daxing Airport. The passenger throughput and construction scale of super large airports are very large. Therefore, a correct understanding of their characteristics and needs, a reasonable evaluation index system and appropriate evaluation methods will help to obtain accurate evaluation results, and thus provide scientific and reasonable technical support for the formulation of airport management and control measures.

The evaluation of urban integrated passenger transport hub in China mainly focuses on the transport function, including transfer organization efficiency, operation efficiency and information service level. The common methods are analytic hierarchy process (AHP) and fuzzy

ICICA 2022, December 02-04, Chongqing, People's Republic of China

DOI 10.4108/eai.2-12-2022.2327958

comprehensive evaluation[1]-[4], but the evaluation results of these methods are vulnerable to human subjective factors. Some scholars also used other methods to conduct relevant research, such as data envelopment analysis (DEA) used by Qin Yu et al.[5] to evaluate the transfer efficiency of rail hubs; Fang Xiaohong et al.[6] re-integrated the DEA method and supplemented the decision-making unit (DUM) ranking method on the basis of the traditional CCR model; Hu Lingli[7] used the improved AHP based on the cloud model scale judgment matrix to calculate the weight of evaluation index; Dang Ruirui[8] combined analytic hierarchy model (AHM) with entropy weight method to determine index weights, and took the five element connection number method as an evaluation method to construct a multi-modal cooperative operation assessment model for the ground transportation center of the airport hub. In a word, whether it is the determination of index weight or the selection of evaluation methods, it is necessary to find more objective, effective and reasonable methods.

Support Vector Machine (SVM) is a new machine learning method based on statistical theory. Based on the principle of structural risk minimization, it can seek the best compromise between model complexity and learning ability according to limited sample information in order to obtain the best generalization ability[9]. Compared with traditional machine learning methods, SVM requires less samples, has simple modeling, convenient calculation, short training time, strong universality, and has good regression and classification modeling capabilities. In the field of transportation, SVM is widely used in data analysis and mining, including regression prediction of passenger flow and traffic congestion, and classification prediction of traffic status and traffic events[10]. In the field of transportation, SVM can solve the modeling problem of small sample, nonlinear and high-dimensional pattern recognition, and better balance the contradiction between "under learning" and "over learning". However, few studies have been applied to the evaluation of urban comprehensive passenger transport hub. Therefore, this paper proposes a real-time evaluation method based on SVM for the cooperative operation of multiple transportation modes of super large airports, The operation data of Capital Airport are used to verify the results.

2 EVALUATION INDEX SYSTEM OF COOPERATIVE OPERATION

2.1 Determination of evaluation index system

Selecting appropriate and targeted evaluation indexes and constructing a scientific evaluation index system are the basis for accurate evaluation of airport cooperative operation. The large demonstration airport hub in the city should not only undertake the unified scheduling of air transport, but also cooperative the transfer of various transport modes, maximize the use of transport capacity resources, and improve the cooperative operation efficiency of the entire airport system. Its core function is to provide convenient and comfortable transport services for a large number of passengers. As most airport hubs are located at the edge of cities, the transport demand is mainly passenger transport, which requires higher travel time, and the functions are mainly distributed, supplemented by transfer. The above characteristics determine that the airport hub operation evaluation index system is different from the general comprehensive transport hub. The design of its operation status indexes should be based on the demand characteristics and existing problems. The establishment of the index system should reflect the cooperative operation efficiency of multiple transport capacities.

The paper establishes the tower index system of cooperative operation state from three levels. Firstly, the target layer is established, that is, the evaluation of the cooperative operation of the super large airport with multiple transportation modes; Secondly, the criteria layer is established, including the external collection and distribution efficiency, internal transfer efficiency, key traffic channel status and information interaction efficiency of the airport; Finally, 15 specific indexes are determined as the index layer. The established operation status evaluation index system is shown in Figure 1, and the indexes are described in Table 1.

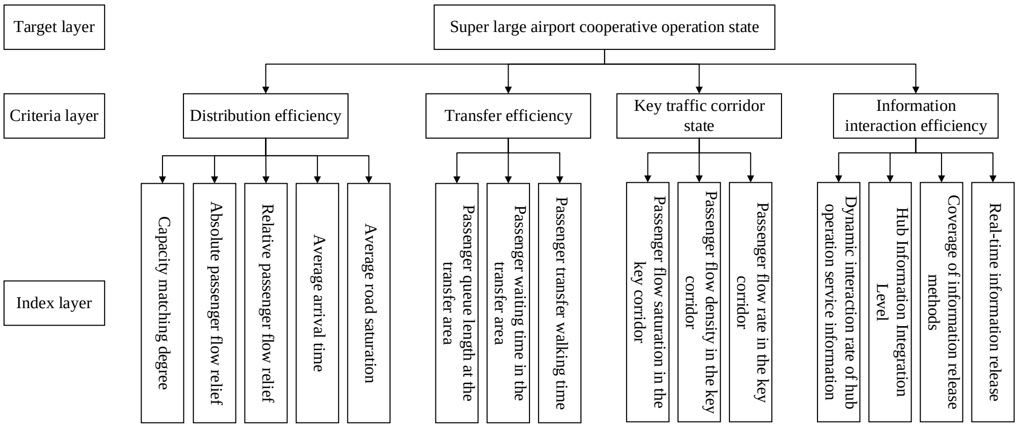


Figure 1. Evaluation index system for cooperative operation of super large airports

Table 1. Index description

Criterion Level	Index Level	Description
Distribution efficiency	Capacity matching degree (X_1)	The matching degree of passenger flow and capacity flow resources
	Absolute passenger flow relief (X_2)	The scale of evacuated passenger flow per unit time
	Relative passenger flow relief (X_3)	The ability to relieve passenger flow per unit time
	Average arrival time (X_4)	The accessibility from urban hotspots to hubs
	Average road saturation (X_5)	The space-time coordination of passenger flow and vehicle flow in the surrounding road network
Transfer efficiency	Passenger queue length at the transfer area (X_6)	The matching degree between a certain capacity resource and the passenger flow choosing this mode

Criterion Level	Index Level	Description
	Passenger waiting time in the transfer area (X_7)	of transportation
	Passenger transfer walking time (X_8)	The efficiency of passengers transferring to each mode of transportation
Key traffic corridor state	Passenger flow saturation in the key corridor (X_9)	The load situation of the passenger flow in key corridors within the hub
	Passenger flow density in the key corridor (X_{10})	The congestion levels in key corridors within the hub
	Passenger flow rate in the key corridor (X_{11})	The passenger flow relief efficiency in key corridors within the hub
Information interaction efficiency	Dynamic interaction rate of hub operation service information (X_{12})	The comprehensiveness of hub management's access to various dynamic information
	Hub Information Integration Level (X_{13})	The convenience for passengers to obtain real-time traffic information through various means
	Coverage of information release methods (X_{14})	The continuity of traffic information received by passengers while walking inside the hub
	Real-time information release (X_{15})	The time delay from the information access platform to the passenger receiving the information

2.2 Determination of evaluation grade

In view of the advance and particularity of the super large airport demonstration hub, this paper proposes new evaluation indexes, such as the state indexes of key corridors, which will be determined through expert experience feedback and comprehensive division of actual data analysis of airport cooperative operation; Other indicators are determined selectively based on actual operation data. The airport cooperative operation status is divided into five levels, namely excellent, good, normal, poor, and bad. The proposed evaluation level and the classification of each index value are shown in Table 2. The classification value of the index level is based on the time granularity of hours.

Table 2. Classification of evaluation index

Evaluation index	Single-index evaluation grade standard				
	Excellent	Good	Normal	Poor	Bad
Capacity matching degree	(0.8,0.95)	(0.95,1.1)	(1.1,1.25)	(1.25,1.4)	(1.4,1.55)
Absolute passenger flow relief (10,000	(≥ 1)	(0.6,0.8)	(0.4,0.6)	(0.2,0.4)	(0,0.2)
Relative passenger flow relief	(≥1)	(0.9,1)	(0.8,0.9)	(0.7,0.8)	(≤0.7)
Average arrival time (min)	(0,20)	(20,40)	(40,60)	(60,80)	(≥80)
Average road saturation	(0,0.6)	(0.6,0.7)	(0.7,0.8)	(0.8,0.9)	(≥0.9)
Passenger queue length at the transfer area	(0,5)	(5,10)	(10,15)	(15,20)	(≥20)
Passenger waiting time in the transfer area	(0,3)	(3,6)	(6,9)	(9,12)	(≥12)
Passenger transfer walking time (min)	(0,3)	(3,6)	(6,9)	(9,12)	(≥12)
Passenger flow saturation in the key corridor	(0,0.5)	(0.5,0.6)	(0.6,0.7)	(0.7,0.8)	(≥0.8)
Passenger flow density in the key corridor (p/m^2)	(0,0.4)	(0.4,0.7)	(0.7,0.9)	(0.9,1.1)	(≥1.1)
Passenger flow rate in the key corridor	(0,360)	(360,460)	(460,625)	(625,735)	(≥735)
Dynamic interaction rate of hub operation	(≥0.9)	(0.7,0.9)	(0.5,0.7)	(0.3,0.5)	(0,0.3)
Hub Information Integration Level	(0.8075,1)	(0.8075,0.7575)	(0.7575,0.7075)	(0.7075,0.6575)	(0,0.6575)
Coverage of information release methods	(0.9,1)	(0.8,0.9)	(0.7,0.8)	(0.5,0.7)	(0,0.5)
Real-time information release (s)	(0,1)	(1,3)	(3,5)	(5,7)	(7,9)

3 CLASSIFICATION AND EVALUATION METHOD BASED ON DECISION TREE SVMPrinciple of SVM classification

According to the sample type, SVM divides the sample data into two cases: linearly separable and linearly nonseparable. Because there are many evaluation indexes and the samples are high-dimensional data, the nonlinear method is adopted in this paper. When the sample data is linearly inseparable, the kernel function can be introduced to map the linearly inseparable sample data in the low-dimensional space to the high-dimensional space, which can be transformed into a linearly separable sample classification problem in the high-dimensional space, and then the linear classification method can be used to solve it. In order to allow the existence of outliers and measure the loss caused by outliers, a slack variable ξ_i is introduced. At the same time, a penalty factor C is also introduced to measure the degree of penalty for loss. Then the nonlinear SVM classifier is described by mathematical model as shown in formula (1):

$$\begin{cases} min\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{k}\xi_i\right) \\ s.t.\ y_i\left(w\cdot x_i + b\right) \ge 1-\xi_i, \xi_i \ge 0\ \ i=1,2,\cdots,k \end{cases} \tag{1}$$

3.2 Selection of kernel function and analysis of main parameters

Kernel function can transform nonlinear classification problems in low dimensional space into linear classification problems in high-dimensional space, which is the key to evaluate the transformation of classification problems. At present, the most widely used kernel function is the radial basis kernel function (RBF)[11]. The SVM using the radial basis kernel function for calculation has less computation, stronger learning ability, and better generalization ability. Therefore, the SVM model in this paper selects the radial basis kernel function, whose expression is:

$$K\left(x_i, x_j\right) = exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) = exp\left(-\gamma\left\|x_i - x_j\right\|^2\right) \tag{2}$$

In addition to choosing a suitable kernel function, the generalization ability of SVM also depends on a set of good parameters, among which the penalty factor and kernel parameters are the most important performance parameters. The two goals pursued by SVM classification are contradictory to each other, that is, to maximize the classification interval and minimize the training errors, so penalty factor C are needed to reconcile them. The larger the value of C, the greater the penalty for the empirical error, the smaller the classification interval, the fewer the number of support vectors, the higher the complexity of the model, and the more prone to "over learning"; Vice versa.

The parameters of the kernel function are also important to the accuracy of the model. The change of the kernel parameter can implicitly change the mapping function, thus changing the complexity of determining the subspace distribution of the sample data with the largest dimension, which also determines the minimum empirical error that the classification hyperplane can achieve. It can be seen that finding a suitable set of parameters is of great importance to the accuracy and generalization ability of SVM model.

Matlab software version 2021 is used for example verification in this paper. Two parameters to be optimized are penalty factor and kernel parameter. The corresponding parameters in the new version of Matlab are BoxConstraint and KernelScale. The selected optimization methods are Bayesian optimization method and 5-fold cross validation method. Compared with grid search method, Bayesian optimization method has the advantages of fewer iterations and faster convergence speed, so its parameter optimization efficiency is higher.

3.3 Decision tree SVM classification

In essence, SVM classifier is a two class classifier. In this paper, there are five levels (excellent, good, normal, poor and bad) of cooperative operation state in the evaluation. Therefore, it is necessary to build a multi classifier suitable for evaluating cooperative operation state by using the binary classification property of SVM. SVM can be broadly divided into two categories to solve multi classification recognition problems: the "whole method" and the "decomposition method"[12]. The latter is not only simpler to solve, but also has advantages in classification accuracy, so it is more commonly used in practical applications.

Decomposition methods mainly include one-to-one classification, one-to-remainder classification and decision tree classification[13]. Among them, the decision number classification method is widely used. It combines the two classification characteristics of SVM, recombines the categories of multi category classification, establishes multiple sub classifiers as root nodes and each category as leaf nodes, and constructs a decision tree based on the decision of SVM sub classifiers, which can accurately identify all categories. For the evaluation of the cooperative operation state of the airport, four sub-classifiers need to be constructed, each of which can identify a state. In the evaluation, the decision function only needs to be calculated from the root node, and the next node is determined according to the positive or negative value until it reaches a certain leaf node.

4 CASE ANALYSIS

4.1 Data preprocessing

Capital Airport is located in the northeast of Beijing, 25.35 kilometers away from Tiananmen Square. It has three terminals, two 4E-level runways and one 4F-level runway, as well as passenger and cargo handling facilities. Currently, the transportation resources it covers include airlines, high-speed railways, airport buses, taxis, online car hailing, subways and intercity railways.

The data source of this paper is the historical operation data and simulation data of Capital Airport, and the historical data collection frequency is 15min. The operating state level of each sample data can be calibrated according to the evaluation level classification criteria described above. A total of 1200 groups of data are selected from the calibrated historical operating data and simulation data according to the state, of which 240 samples are included in each operating state, 200 samples are taken for each state as training data, and 40 samples are taken as test data.

Part of the sample data is shown in Table 3, in which X_i represents each index in the evaluation index system, and Y represents the calibrated state level.

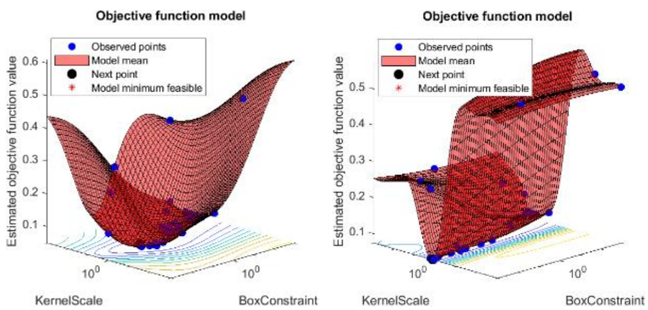
Table 3. Sample data (excerpt)

No.	X															Y
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	
1	0.963	1.247	0.978	3.2	0.275	7	1.50	3.05	0.222	0.164	205.205	0.925	0.938	0.990	0.2	1
2	0.849	1.390	0.993	4.3	0.426	5	4.70	0.33	0.544	0.401	213.852	0.942	0.834	0.983	0.0	1
3	0.986	1.358	0.964	5.0	0.470	2	4.90	2.65	0.049	0.500	36.968	0.937	0.973	0.981	0.5	1
4	1.041	0.997	0.928	43.1	0.673	9	7.65	3.44	0.529	0.437	499.870	0.857	0.765	0.775	3.2	2
5	0.906	0.831	0.992	37.4	0.623	5	5.00	4.56	0.551	0.585	474.489	0.856	0.774	0.869	2.1	2
6	0.955	1.001	0.946	47.8	0.717	11	3.24	5.36	0.640	0.404	483.967	0.793	0.773	0.894	2.9	2
7	1.168	0.453	0.790	68.4	0.714	17	6.11	8.03	0.686	0.890	667.146	0.533	0.711	0.770	4.2	3
8	1.051	0.727	0.865	42.5	0.701	15	8.80	8.34	0.722	0.887	478.469	0.674	0.750	0.732	3.3	3
9	1.163	0.647	0.778	43.2	0.763	17	7.50	9.05	0.661	0.798	567.607	0.575	0.741	0.759	3.4	3
10	1.208	0.316	0.683	73.9	0.903	16	11.79	9.01	0.786	1.037	630.778	0.323	0.659	0.517	7.0	4
11	1.299	0.513	0.666	82.9	0.844	21	11.99	9.76	0.725	1.085	723.895	0.439	0.679	0.595	6.4	4
12	1.116	0.142	0.666	84.5	0.880	20	9.74	9.57	0.722	1.180	715.743	0.390	0.675	0.671	5.4	4
13	1.392	0.117	0.617	87.2	1.034	21	12.62	12.42	0.966	1.799	844.315	0.226	0.196	0.450	8.6	5
14	1.201	0.249	0.605	81.1	1.020	20	14.70	13.80	0.810	1.799	821.347	0.270	0.222	0.109	8.4	5
15	1.355	0.241	0.665	90.4	1.081	22	14.29	14.70	0.816	2.126	777.883	0.219	0.565	0.038	7.9	5

During training, the data of one state is regarded as a positive class, and the data of the other states is regarded as a negative class. However, at this time, the number of positive class samples is 200, and the number of negative class samples is as many as 800. The number of positive and negative class samples is extremely uneven, that is, there is a problem of data skew, which will increase the training error of the model and reduce the accuracy. Therefore, in order to solve the problem of data skew, 200 samples are randomly selected from the remaining state samples as negative samples during training, and the positive samples remain unchanged, so that the number of positive and negative samples remains equal.

4.2 Parameter optimization

In this paper, function fitcsvm in MATLAB 2021 is used for modeling and training. Before training, the sample data is standardized, and the kernel function uses RBF kernel function. Bayesian optimization method and 5-fold cross validation method are used to train the data, with minimizing the classification error as the objective function, the search range of BoxConstraint and KernelScale parameters are both [0.001,1000], and the number of training iterations is 30. The processed positive and negative sample data matrix is input into each sub classifier model for training, and the training results of each sub classifier parameter are shown in Figure 2.



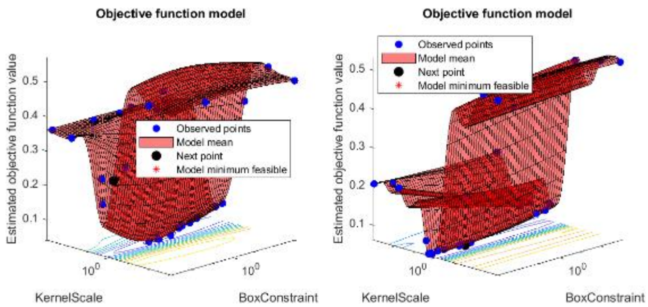


Figure 2. Parameter optimization results of SVM sub classifiers

Within 30 training iterations, the training objective function values of the four sub classifiers all tend to be stable, so it can be considered as convergence. According to the objective function model diagram, the optimal parameter combinations corresponding to each sub classifier can be obtained as follows:

BoxConstraint = 15.354, KernelScale = 43.214;

BoxConstraint = 0.0010087, KernelScale = 1.2212;

BoxConstraint = 2.3349, KernelScale = 2.4058;

BoxConstraint = 0.0011019, KernelScale = 2.3983.

The training accuracy of each sub classifier model is shown in Table 4.

Table 4. Training results of SVM sub classifier

Sub classifiers	BoxConstraint	KernelScale	Training accuracy (%)
SVM sub classifier 1	15.354000	43.214	94.9542
SVM sub classifier 2	0.0010087	1.2212	92.3283
SVM sub classifier 3	2.3349000	2.4058	95.3598
SVM sub classifier 4	0.0011019	2.3983	93.3769

4.3 Model Testing

After obtaining the optimal parameter combination of each sub classifier, the decision tree SVM evaluation model can be determined, and the test data set is input into the trained classification model to classify the state level of each sample. The test results show that among the 200 sample points, there are 9 sample points that are misclassified, including 2 misclassified by SVM sub classifier 1, 3 misclassified by SVM sub classifiers 2 and 4, and 1 misclassified by SVM sub classifier 3. The accuracy of each sub classifier model and decision tree SVM evaluation model on the test set is shown in Table 5.

Table 5. Accuracy of sub classifiers and decision tree SVM evaluation model

Sub classifiers	Number of test samples misclassified	Testing accuracy (%)
SVM sub classifier 1	2	95.0
SVM sub classifier 2	3	92.5

SVM sub classifier 3	1	97.5
SVM sub classifier 4	3	92.5
Accuracy of decision tree SVM evaluation model	95.5	

According to the training and test results, the classification accuracy of each sub classifier and decision tree SVM evaluation model on the training set and test set is more than 90%, with high accuracy and good generalization ability. The fuzzy evaluation method has strong subjectivity in the evaluation process of determining the index weight and calculating the fuzzy relation matrix, while the decision tree SVM evaluation method reduces the intervention of human subjective factors in the evaluation process, and can be objectively applied to the evaluation of large-scale airport cooperative operation state.

5 SUMMARY

Based on the characteristics and requirements of super large airport cooperative operation, the evaluation index system is designed, and the decision tree SVM evaluation model is constructed. Bayesian optimization method and 5-fold cross validation method are used to optimize the parameters. In the evaluation of large-scale airport cooperative operation with many influencing factors, the SVM model is suitable for small sample, high-dimensional and nonlinear evaluation methods. Compared with the analytic hierarchy process and fuzzy evaluation method, this method reduces the impact of subjective factors and uncertainties in many aspects of the evaluation, improves the reliability of the results, and helps airport operation managers to accurately grasp the real-time cooperative operation state and timely dispatch the transport capacity resources.

ACKNOWLEDGEMENT

National Key Research and Development Program of China (2018YFB1601200)

REFERENCES

[1] Ge, P., “Fuzzy evaluation of urban rail transit hub transfer connection based on analytic hierarchy process (AHP),” Automotive Practical Technology, 206-207+220 (2019).
[2] Zhu, X. S., Pen, G. X. and Zhou, W. H., “Fuzzy evaluation of public transport hub transfer connection based on analytic hierarchy process (AHP),” Journal of Inner Mongolia University of Technology (Natural Science Edition), 311-315 (2003).
[3] Zhang, H. Z., “Fuzzy evaluation of hub traffic operation situation based on analytic hierarchy process,” China Municipal Engineering, 1-3+86 (2011).
[4] Sun, Q. P. and Cheng, D. X., “An empirical study on fuzzy comprehensive evaluation of integrated transportation hub transfer connection,” Technology and Innovation Management, 164-166 (2010).
[5] Qin, Y. and Yan, K. F., “DEA non-uniform evaluation model of transfer efficiency of rail transit hub,” Journal of Chang'an University (Natural Science Edition), 48-51+54 (2002).

[6] Fang, X. H., Zhou, L. S. and Zhou, B., “Research on evaluation method of urban comprehensive transportation hub based on DEA,” Transportation Systems Engineering and Information, 2008.
[7] Hu, L. L., “Research on the evaluating index and method of city comprehensive traffic hub,” Wuhan University of Technology, 2008.
[8] Dang, R. R., “Research on multi-mode cooperative operation evaluation of ground transportation center in airport hub,” Beijing Jiaotong University, 2021.
[9] Yu, R., Wang, G. X., Zheng, H. Y. and Wang, H. Y., “Research on pattern recognition of urban road traffic state based on support vector machine,” Transportation Systems Engineering and Information, 130-136 (2013).
[10] Lin, H., Li, L. X. and Wang, H., “Research and application of Support Vector Machine in Intelligent transportation system,” Computer Science and Exploration, 901-917 (2020).
[11] Zhu, F., Zhao, Q. P. and Wang, J. T., “Application research of SVM classification based on RBF kernel,” Journal of Tangshan University, 13-17+39 (2017).
[12] Liu, Z. G., Li, D. R., Qin, Q. Q. and Shi, W. Z., “Generalization of Support vector Machine for multi-class classification problems,” Computer Engineering and Application, 10-13+65 (2004).
[13] Sun, D. S., “Research on classification and regression method of Support Vector Machine,” Central South University, 2004.