Competitiveness Analysis of Urban Public Transport Based on SVM

Hui Sun[1], Yuchun Li*[2], Zhiqing Fan[3], Ye Shi[4]

1. School of Management, Tianjin University, Tianjin, 300072
E-mail: bighui2000@sina.com

2. School of Management, Tianjin University, Tianjin, 300072
*E-mail: lyc-1182@163.com

3. School of Management, Tianjin University, Tianjin, 300072
E-mail: fanzhiqingip@yahoo.com.cn

4. School of Management, Tianjin University, Tianjin, 300072
E-mail: shiyetiger@126.com

Abstract: The comprehensive analysis of urban public transport has an important role in promoting the development of urban public transport. Based on the characteristics of urban public transport, following the principles of scientific, integrity, independence and operational, this paper built a comprehensive analysis index system for urban public transport, then the comprehensive evaluation method based on the support vector machine (SVM) is brought out trying to analyze the urban public transport of comprehensive after the index system is established. And then, a case study is executed by taking 20 cities in Hebei province in China as an example. The research results reveal that the method can efficiently evaluate comprehensively to the urban public transport of comprehensive. Thus it could be useful for the government to make relevant policies to improve the operational performance for public transport. In the end, some policy advices are given.

Key Words: Public Transport, Competitiveness Analysis, SVM

I. Introduction

As one kind of important urban infrastructures, the operational state of urban public transport closely interrelated to people's daily life and production. In nowadays, idea of urban development led by transport is quite prevalent, thus it is urgent and important to enhance the research in urban public transport system and evaluate the competitiveness of public transport. However, there still remain some problems such as insufficient investment, insufficient policy support in the developmental process of urban public transport, and to solve these problems, the basic way is to analyze and quantize the competitiveness of urban public transport in each city and find out the key factors which limit its development. Then the result can be applied as theoretical guidance for policy making of urban public transport development. Based on the above, the paper proposes a new method to quantize and forecast competitiveness of urban public transport by support vector machine and conducts a empirical research taking Hebei Province as a example. The results reveal that SVM could exactly fit the development level of urban public transport of Hebei Province, and at the same time the results provide theoretical according for policy making and investment decision of public transport.

II. Support Vector Machine

SVM was developed by Vapnik and his partner and then it was introduced in to machine learning realm on the conference of calculation theory learning in 1992. After that, SVM was widely noticed by scholars and was deeply developed in late 90's of 20th century. In today, SVM has been a standard tool in the realms of machine learning and data mining. SVM, which is a successful realization of SLT, following the standard of Structural Risk Minimization (SRM), improves the generalization (forecasting) capability of Learning Machine and properly solves the practical problems such as small sample, nonlinearity, higher-dimension and local minimum. In summary, the basic thought and principle of SVM is that alternating the input space to a higher dimensional space through nonlinear converting, and a unique globally optimal solution could be acquired when solving a constrained convex quadratic programming problem [1].Because of the above characteristics of SVM, it becomes an excellent learning algorithm which has been successfully applied in the realms of pattern recognition and function fitting.

A. Regression Principle of SVM

The foundation of regression algorithm of SVM are ξ insensitive function and kernel function algorithm. If using the fitting mathematical model to express certain curve in hyperspace, then according to the results from ξ insensitive function, the graph will show the curve itself and ξ pipeline formed by training points. Of all the sample points, only those distributed on the "pipe wall" could determine the location of pipeline and these training samples are called "support vector" [2]. To suit the nonlinearity of training sample sets, the traditional fitting methods usually add some high order terms in the last of linear equation, but this action always causes some adjustable parameters which will increase risk of over fitting. SVM regression algorithm perfectly avoids this contradiction, which replaces the linear terms by kernel function to make former linearized algorithm "nonlinearity". Meanwhile,

This work is supported by National Nature Science Foundation under Grant 70772057

Elearnica
Downloaded from http://www.elearnica.ir

bringing in the kernel function achieves the objective of “increasing dimension” and the situation can still be under control if the adjustable parameter added is over fitting [2,3].

B. Selection of Kernel Function

Kernel functions are mainly divided into four categories, which are linearity, multi-classification, radial basis function, sigmoid kernel function. When dealing with the problems as classifying and regression forecasting, radial basis function is always selected, as follows:

$$K（x, x_i) = \exp[-g^*(x - x_i)^2] \quad (1)$$

There are several reasons for selecting radial basis function:

(1)Compared to linear kernel function, radial basis kernel function could resolve the linearly nonseparable case, which maps low dimentional input pace to higher dimensional feature apace through non-linear transformation. Moreover, linear kernel function can be seen as a special kind of radial basis kernel function, both of them play the same roles when solving with linear cases [4]. Besides, Sigmoid kernel function performs the same as radial basis function when defining parameters.

(2)Multinomial kernel function has more hyper-parameters, which will make the model much more complex.

(3)Radial basis kernel function has less numeral troubles than other three kinds of kernel functions.

III. Apply SVM in Competitiveness Evaluation of Public Transport

A. Principle of Evaluation Model of Public Transport Competitiveness Based on SVM

The evaluation model of public transport competitiveness based on SVM, which uses series of evaluation indices, takes advantages of expert’s knowledge and experience, excludes the effect of subjective factors through learning and training by SVM and provides support for decision makers. Generally, the principle and process of SVM forecasting algorithm of public transport competitiveness can be described as follows, firstly transform the input space to a higher space through non-linear transformation defined by integral operator kernel functions and then seek the Optimal Hyperplane in this space [1,5]. Suppose there are m (M>2) competitiveness factors of public transport, and the number of competitiveness indices needed is n (n>0), so a map exists from m dimention to n dimention. If define m as the input number and n as the output number of of SVM, then in the m dimentional space R, m has a bounded subset A, and there is a map from n dimentional space Rn to a bounded subset F(A) [6], as follows:

$$F : A \in R_{\mathrm{m}} \to R_{\mathrm{n}}，\ Y = F(X) \quad (2)$$

In regards to training set A={X,Y}, a optimized approximate map G could be found out through learning, making

$$Y_j = G(X_i)(I = 1,2,\ldots，\ m) \quad (3)$$

This example can be seen as the application of SVM in regression problems.

B. Establish the Evaluation Index System of Public Transport Competitiveness

In the study of competitiveness factors of public transportation, selection of indexes is quite critical. The selection of indexes should comply with the reality and data acquirable principles, From the Figure Ⅰ, we can find that the indexes of public transportation competitiveness factor could be selected form the following four aspects:

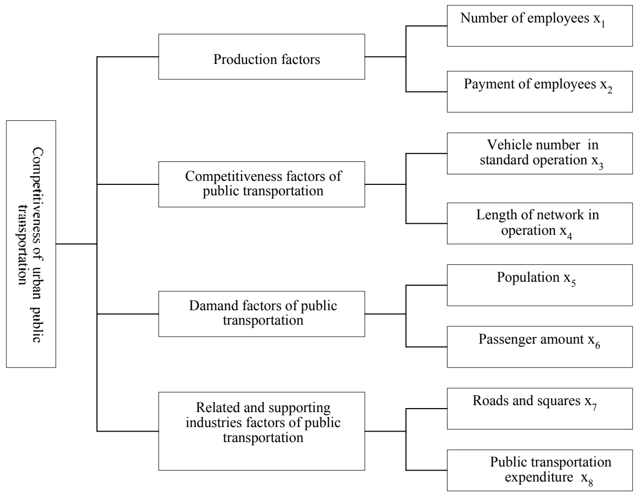

FIGURE Ⅰ INDEXES OF URBAN PUBLIC TRANSPORTATION COMPETITIVENESS

(1) Indexes of production factors (PF)

In this aspect, number of employees (NOE) x_1 and payment (PT) of employees x_2 are selected. The number and payment of employees can reflect the competitiveness of an industry horizontally compared to other industries. People’s effect is the most significant in all the production factors, thus number and payment of employees can fully reflect this standard.

(2) Indexes of public transportation competitiveness factors (CF)

Vehicle number (VN) in standard operation x_3 and length of network (LON) in operation x_4 are selected as the indexes, as these two indexes are the standards of facility index and infrastructure index, which could best reflect the competitiveness of public transportation.

(3) Indexes of demand factors (DF) of public transportation

Here the population (POP) x_5 and total passenger amount (PA) x_6 are selected, as the residents are the main service objects of public transportation and also the potential customers of public transport, so population can directly reflect this index. Meanwhile, total passenger amount reflects the actual requirement of public transport.

(4) Indexes of related and supporting industries (RSI)

As the most direct indexes of related and supporting industries to the public transportation competitiveness, roads and squares (RAS) x_7 and public transportation expenditure (PTE) x_8 are selected.

IV. CASE STUDY

A. Data Select

According to the indices defined above in the above of the paper, data in this paper is all from Annual construction statistical report (2006) of cities in Hebei province. There are twenty sets of public transport data from the twenty major cities in Hebei. In this paper, the twenty sets of data is divided in to training samples and testing samples (the first 15 sets are training samples and the others are testing samples). Target values in Table Ⅰ are the scores given by experts to mark the magnitude the competitiveness indices of urban public transport, ranking from 0~1, closer to 1 means the city's public transport competitiveness is stronger.

Table Ⅰ Annual statistical report of urban construction in Hebei

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	Target Value
Shijiazhuang	1575	1100	207	23000	13	2100	20411	29128	0.9005
Xinle	25	60	12	37	1	210	240	19056	0.1220
Tangshan	1802	422	183	12549	15	3780	12019	28970	0.7805
Qian'an	13	56	15	68	1	205	344	21045	0.1370
Qinghuangdao	662	258	76	8707	8	1731	5808	26173	0.4555
Handan	1161	526	139	11491	10	3341	10447	28680	0.6855
Xingtai	408	340	56	4644	4	2105	5684	27316	0.3975
Shahe	66	90	10	300	1	100	530	22316	0.1470
Baoding	572	320	99	6960	9.	500	6760	28756	0.4410
Zhuozhou	79	90	21	186	2	391	516	21316	0.1645
Dingzhou	68	65	21	160	1	255	625	23005	0.1635
Zhangjiakou	581	338	86	4333	8	2728	5645	28219	0.4845
Chengde	417	188	46	6727	2	1200	4936	27065	0.3365
Cangzhou	210	199	49	2254	2	360	3858	26901	0.2650
Renqiu	116	70	31	840	5	530	1262	25918	0.2270
Huanghua	49	60	8	85	1	176	1300	23458	0.1535
Langfang	159	270	8	2031	6	15700	3780	30646	0.3285
Bazhou	108	50	12	90	2	1500	883	24356	0.2170
Sanhe	82	82	11	52	3	292	220	19005	0.1460
Hengshui	66	124	32	805	5	300	1840	23045	0.2135

B. Evaluate the Competitiveness of Urban Public Transport by Libsvm Software

When dealing with the data of urban public transport competitiveness, this paper takes Libsvm software as the evaluation tool of urban public transport competitiveness. The basic steps include:

(1) Prepare data sets according to the form required by Libsvm software package

The file format of training data and testing data applied by this software can be seen as follows:

<label> <index1>:<value1> <index2>:<value2> ...

In the format, <label> is the target value of training data sets, as to classification, it stands for the integer which marks certain category (supporting multi categories). As to regression, it can be any real number. <index> is an integer beginning from 1, and it could be incontinuous. <value> is a real number, which is the independent variable we normally called. Label in the testing data file is just used for calculating accuracy and error, if it is unknown, a random data can be used to fill this blank, or just let it empty.

(2)Simply scaling the data set

It is necessary to scale the data by svm-scale, as it is helpful for selecting parameters and accelerating the speed of processing data through scaling. Svm-scale can scale the value, and the scopes are usually [0,1] or [-1,1]. Besides, it should be noticed that testing data and training data need to be scaled together.

(3)Select the kernel function

Select radial kernel function in regression of SVM

$$K(x, x_i) = \exp[-g^*(x - x_i)^2] \quad (4)$$

(4)Select optimal parameter C and g by Cross-Validation

In the software of Libsvm, there is a program gridregression.py, which is used to select the optimal parameters C and g in this paper. After the training of relevant command is finished, the value of C and g can be get from the last column of the file of "gridregression_data.parameter".

(5)Apply the optimal parameter C and g in to the whole training sample and ge the SVM model.

Input the value of parameters into the command and training the training samples, then get the SVM model. In the options, selecting -s=3 means regression option, selecting -t=2 means radial basis kernel function. After the training finished, a "data.txt.model" file is acquired.

(6)Testing and forecasting by the model

Combined with the model acquired, input the command by svm-predict program, test and predict the testing samples, the results can be seen in TableⅡ.

In the above table, we respectively list the expert assessment value and SVM evaluation value of testing samples. From the results, we can find that the value of SVM evaluation of the five testing samples are very close to expert assessment and the error is very small. Last of all, from the evaluation values of the testing

sample, define expert assessment value as the real evaluation value, set following formula as the standard of evaluation quality.

$$\delta = \sqrt{\frac{1}{\mathrm{k}} \sum_{i=1}^{k} (v_i - e_i)^2} \quad (5)$$

In the formula, δ is the mean square error, which assesses the quality of testing results. v_i and e_i respectively stands for the evaluation result of the i-th testing sample by SVM and expert, k=5.

Table Ⅱ Comparison between expert assessment and SVM

Testing Samples	Expert Assessment	SVM evaluation	Absolute Error	Ralative Error	Mean Square Error
Huanghua	0. 1535	0. 15339	−0. 00011	−0. 0717%	2. 56e−3
Langfang	0. 3285	0. 333352	0. 004852	1. 4770%	
Bazhou	0. 2170	0. 214088	−0. 002912	−1. 3419%	
Sanhe	0. 1460	0. 146121	0. 000121	0. 0829%	
Hengshui	0. 2135	0. 214392	0. 000892	0. 4178%	

Above definition indicates that smaller the mean square error, closer between the SVM evaluation result and expert assessment and better of this evaluation model. Otherwise, bigger the error, the evaluation of SVM is worse. In table Ⅱ, we calculate out the mean square error of the five testing samples is 2.56e-3, which means the SVM is quite valid in competitiveness evaluation of urban public transport.

V. CONCLUSION

Competitiveness analysis of urban public transport is the basis for input decision of public transport system and reforming transport development. When dealing with this kind of problems, previous evaluation material of public transport competitiveness is needed and subjective factors should be excluded. SVM, which is a rising artificial intelligent science, could take place of human to some extent and reduce the subjectivity of competitiveness predicting. Therefore, SVM is chosen as the evaluation model of urban public transport competitiveness in this paper and empirical research is conducted to verify this model.

There are three advantages of SVM mentioned in this paper.

First, SVM is based on the statistics theory, thus it can deal with the problems of small samples.

Second, SVM resolves the problems of over-learning and local extremum, it has excellent fitting accuracy and generalization capability and it doesn’t need mathematical modeling but realized by computer.

Third, SVM utilizes previous predicting and analysis material of urban public transport, so it establishes a model without manual intervention, which reduces the human subjective and realizes the automation of predicting and evaluation.

REFERENCES

[1] Wun-Hwa Chen, Jen-Ying Shih, A study of Taiwan’s issuer credit rating systems using support vector machines, Expert Systems with Applications, Vol.30, No.3, pp.427-435, 2006.

[2] Cheng-Lung Huang, Mu-Chen Chen, Chieh-Jen Wang, Credit scoring with a data mining approach based on support vector machines, Expert Systems with Applications, Vol.33, No.4, pp.847-856, 2007.

[3] Jian-tao Xia, Ming-yi He, High dimensional multi-spectral image classification by SVM and its characteristic analysis, Computer Engineering, Vol.29, No.13, pp.27-28, 2003.

[4] David Martens, Bart Baesens, Tony Van Gestel, Jan Vanthienen, Comprehensible credit scoring models using rule extraction from support vector machines, European Journal of Operational Research, Vol.183, No.3, pp.1466-1476, 2007.

[5] Bin Yang, You Lu, Classification method of support vector machine based on statistical learning theory, Computer Technology and Development, Vol.16, No.11, pp.56-58, 2006.

[6] Jin-mu Zhang, Bo Wu, Yu Chen, Jian Lin, Study on the method of projection risk evaluation based on support vector machine, Optimization of Capital Consruction, Vol.27, No.6, pp.77-94, 2006.