2011 International Conference on Management Science & Engineering (18th)
September 13-15, 2011 Rome, Italy

Prediction Based on Support Vector Machine for Travel Choice of High-Speed Railway Passenger in China

KANG Shu[1]，LI Jing[1]，LIU Mei[1]，ZHU Xin[2]
1 School of Economics and Management, Beijing Jiaotong University, P.R.China,100044
2 China United Network Communications Group Co.,Ltd, P.R.China, 100032

Abstract: High-speed railway is a very important part of transportation industry in China, and travel choice has key effect on the development of high-speed railway. Therefore, research on travel behavior of passengers and prediction their travel choice, will offer valuable suggestion for high-speed railway running. In this paper, support vector machine (SVM) is the main method being used to predict. Support vector machine is based on the structural risk minimization principle, and it improves the generalization ability of learning machine to the maximum extent. When solving the limited-sample and nonlinear problems, support vector machine has advantages in predicting. In this research, we get six most important factors, which affect travel choice by the means of questionnaire survey, then use libsvm tool to build prediction model and optimize the train parameters of support vector machine. Finally the prediction accuracy is as high as 91.44%, which shows that support vector machine is good at predicting.

Keywords: prediction, travel choice, support vector machine, high-speed railway

1 Introduction

Railway plays a very important role in the Chinese transportation system; its burden of passenger accounts a high share of the transportation markets, along with the rapid development of national economy and railway technology. Making passenger transport faster is the trend of railway development for many countries, so people start to pay attention on the field of high-speed railway. In 1964, Japan built the world first high-speed railway—from Tokyo to Osaka with a speed of 21km/h, thus caused a worldwide revolution in transportation, and profoundly changed the pattern of transportation [1]. China started the research of high-speed railway in 1980s. Compared with other travel choice, high-speed railway has the characters of big transportation capacity, safety, low energy consumption, low pollution, small footprint, all-day work [2], which not only solve the problem of huge number and mobility of intercity passengers, but also offer a fast and comfortable choice for passengers to travel, so many passengers prefer high-speed railway now [3].

However, based on the special condition of China, high-speed railway can't satisfy all the demands, such as price, time, traveling environment, speed. Therefore, utilizing and developing various travel modes is the current situation and future trend[4].

While various travel modes develop coordinately, the competition among them is inevitable. So analyzing the character of travel choice using scientific method, and predicting travel choice can offer theory support for high-speed railway running and improve the level of service and make passenger more satisfied [5]. In the past, to study the characters of travel choice usually based on the experience of specialists or simple logical decision model. However, travel choice is very complicated, which belongs to nonlinear problems. Traditional method for prediction such as exponential smoothing, recession analysis, moving average has limits in solving nonlinear problem[6]. Support Vector Machine (SVM) is a method of pattern recognition based on learning theory, which has special advantages on solving small sample, nonlinear and high- pattern recognition problems. This paper aims to build the prediction model of travel choice after brief introduction of basic principle for support vector machine, and make prediction analysis of travel choice among high-speed railway passengers using support vector machine.

2 Reviews

2.1 The current study on travel behavior of railway passenger

Currently, there are many researches about travel behavior:

(1) Studies on travel choice decision. By anglicizing of travel demands, and the factors that have influence in passengers to make travel choice, then use certain method to build decision model, such as chance constrained programming model with fuzzy parameters[7].

(2) Researches on the value of travel time. Studies

978-1-4577-1888-5/11/$26.00 ©2011 IEEE

the value of travel time and make relation to the travel, then conclusion are made that different value of travel time will affect the passengers to make travel choice. Based on the theory of consumption, expand the original work-leisure time model, and proposed the travel choice mechanism. [8] Some other research even built the model to calculate value of travel time [9].

(3) Researches on the utility of travel for predicting travel choice. By departing the utility into certain utility and random utility and making assumption that random utility subject to certain probability distribution, get the probability of each travel choice [10].

(4) Some other research use three-parameter data loggers or disaggregate travel demand models to do the research of travel behavior.

Usually, there are two ways to make analysis of travel choice: one is qualitative analysis based on conducting a survey of passengers, this method is relatively realistic, but is hard to explain the inner mechanism. The other one is to build model using statistical theory, although those methods are accurate, sometimes they are not applicable to problems like travel choice which are affected by many factors.

2.2 The application of support vector machine in the field of high-speed railway

At present, it is not common that apply support vector machine in researches on high-speed railway. But there still some cases that give examples of application of support vector machine in railway.

(1) Using support vector machine to predict the railway passenger volume. Compare the prediction results of support vector machine prediction model and BP neutral network prediction model, the former one is better, which proves that support vector machine has advantage in predicting in the situation of limited sample over BP neutral network. So it is an advanced method [11].

(2) Evaluating the investment of railway construction. Furthermore, when we expand the scale into the field of whole transportation, we will find more support vector machine application. Such as real-time intelligent recognition of chaos in traffic flow [12] and travel time prediction on urban networks [13].

In this paper, we make a creativity innovation of applying support vector machine in the prediction the travel choice of high-speed railway, we collected the data through survey and extract key factors that affect travel choice most, then we build the prediction model based on those data. The research aim is to combine qualitative analysis with quantitative analysis to make prediction more accurate

3 Methodologies

3.1 Basic theory

The basic idea of support vector machine is make description about multidimensional and complex variables using black-box model, depart the variables into two parts, input variables and output variables, by making optimization and adjustment with support regression machine, find the optimal function that approach the relation between input variables and output variables, thus solving complex problems is transformed into solving the optimal function based on sample data. On the basis of structural risk minimization principal, it improves the generalization ability of learning machine to the maximum extent [14]. Generally speaking, support vector machine's main advantages are reflected as the following areas:

(1) It is specific to the situation of limited sample, and it aims to get the optimal solution according to the current information, but the optimal solution when the number of sample approaches infinity, thus the phenomenon of overearnings can be avoided effectively;

(2) Algorithm finally transformed into quadratic optimization problem, in theory, what we get is the global optimum. Thus the problem of local maxima in neural network can be solved effectively;

(3) With the introduction of kernel function, practical problems will transformed into features space, in which linear discriminant function will be built, in order to achieve nonlinear discriminant function in input space, meanwhile it solve the dimension problem

Skillfully, which make the complexity of algorithm has nothing to do with the dimension [15].

For travel choice, select several key travel factors $\{x_1, x_2, x_3, ...x_i\}$ as train data. Thus several groups of train data constitute a set of area in n-dimensional space, and different travel choice is corresponding to different area. Therefore to forecast travel choice is equal to find the boundaries of these areas. The identification of boundaries depends on train data. However, in many cases, travel factors are nonlinear, even inseparable. Support vector machine offers a simple method to solve this problem, which is making dimension higher. In higher dimension we can have deeper data mining, in this way, the nonlinear problem in low dimension is transformed into linear problem in higher dimension [16]. The procedure of dimension transformed is shown as Fig1.

3.2 The selection of kernel function

The main idea of support vector machine is to mapping nonlinear data into the futures space through a transformation, then we can build the optimal separating hyper plane H (equation (1)) through solving the constrained optimization problem (equation (2)):

$$f(x) = \omega * \phi(x_i) + b \tag{1}$$

$$\begin{cases} \min(\omega, \xi) = 0.5|\omega|^2 + \sum_{i=1}^{l} \xi_i \\ y_i[\omega^* + b] \geq 1 - \xi_i \end{cases} \tag{2}$$

In the above equation, ω is the separating hyper plane of feature space. b is the threshold value of

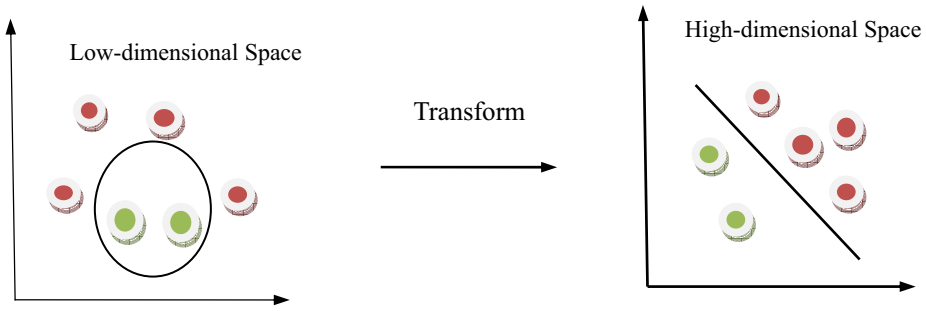

Fig.1 Dimension transformed

Tab.1 Travel choice based on six most important factors

Next travel choice:	Price	Train time	Speed	Environment	Safety	Overall satisfaction:
High-speed railway	Very Expensive	Passable	Passable	Passable	Passable	Passable
Plane	Passable	Dissatisfied	Very dissatisfied	Good	Safe	Dissatisfied
Train	Very Expensive	Very dissatisfied	Passable	Good	Safe	Dissatisfied
High-speed railway	A little Expensive	A little dissatisfied	Very dissatisfied	Good	Very unsafe	Passable
High-speed railway	Passable	A little dissatisfied	Satisfied	Very good	Very safe	Satisfied
Train	Very Expensive	Satisfied	Passable	Very good	Safe	Dissatisfied
Long-distance bus	Cheap	Passable	Satisfied	Good	Safe	Satisfied
Train	A little Expensive	Satisfied	Dissatisfied	Very good	Passable	Passable
….	…	…	…	…	…	…

separating plane, $\xi = (\xi_1, \xi_2, \ldots \xi_l)$ is a relaxation factor in consideration of separating error. c is the penalty factor for error.

As the dimension of feature space is usually very high, direct calculation will cause "dimensional disaster". However, because of the calculation of support vector machine is dot product operation, according to the introduction of kernel function we can transform the dot product operation in high-dimensional space into kernel function in low-dimensional space, and what we need do is to choose the kernel function which is satisfied with Mercer (nonlinear or linear) condition even we don't know how sample information is mapped from original space to feature space. Thus direct calculation in the high-dimensional space can be avoided, and "dimensional disaster" is solved .The main kernel function of support vector machine includes Polynomial Kernel, Radial Basis Kernel and Sigmoid Kernel.

4 Prediction based on support vector machine for travel choice

4.1 Data preparation

From the perspective of passengers, making travel choice is affected by many factors, such as income, level of education, age, the aim of travel, public expense or at their own expense, the preference of travel. These factors limit travel choice to some extent. From the prospective of station and train, factors such as price, environment, speed, and distance of travel affect travel choice also. To satisfy the demand of research, we conducted a survey of high-speed railway passenger. We distributed 2000 questionnaires in total, and 1376 were recovered, including 1232 effective questionnaires, the rate of effectiveness is 89.5%. By making reduction for 21 factors using rough set in basis of quantified data in questionnaires, we get the six most important factors: price, train times, speed, environment, safety and overall satisfaction (Tab 1.)

4.2 Build prediction mode

With the basic principle of support vector machine, we build travel choice prediction model using the separating method of support vector machine. The following are five steps to build the prediction model:

Step 1: The questionnaires data will be processed to be a set of collection constituted of 6 inputs: price, train time speed, environment safety, overall satisfaction, and

1 output: next time choice. What we should do next is separating the next travel choice into two classes: high-speed railway, which is expressed by 1, and others (train, plane and bus), which is express by 2.

Step 2: In Tab 1, we can see that all the data are in the type of classification. However, when use support vector machine to train data and build prediction model,

the data required being the type of numeric, so we transform the data into numeric type. For example, “cheap, passable, a little expensive, expensive, very expensive” in price is expressed by”1, 2, 3, 4, 5” [17].

Step 3: Because kernel values usually depend on the inner products of feature vector, for example, the linear kernel and the polynomial kernel, large attribute values might cause numerical problems, so should scale the data in order to avoid the attributes in greater numeric ranges dominate those in smaller numeric ranges , numerical difficulties during the calculation and improve the prediction’s convergence and accuracy. As Tab 2.shows the data of six attributes are scaled to the range [-1, +1].

Step 4: Kernel function selection. Select radial basis kernel, which is the widely used kernel function to build the prediction function. Because of its wide domain of convergence, radial basis kernel is applicable in spite of low-dimension, high-dimension, small sample or big sample [18][19].

Step 5: Input the train data to initially determine the C value and gamma value. C value is penalty parameter and gamma value is kernel parameter. The toolbox used is libsvm developed by professor Chi-Len Lin, then conduct the first prediction after inputting test data [21].

Step 6: Adjusting C value and gamma value continuously until we get the optima prediction result. From the Fig 2, we can see the accuracy is increased, and lines of different color represent different accuracy.

4.3 Prediction result

After a succession of calculation, the final accuracy is as high as 91.44%, Tab 3 gives the main parameters of the prediction model

Tab.2 Data scaled

Next travel choice	Price	Train time	Speed	Environment	Safety	Overall satisfaction:
1	-0.2	-0.333	-0.667	-0.714	0.333	-0.333
1	0.2	0.667	0.667	0.429	0.333	0.333
2	-0.2	0.333	0.333	0.428	0.333	0.333
2	0.2	-1	-0.667	-0.714	0.333	-0.667
2	0.2	0	0.667	0.429	-0.667	0.333
….	…	…	…	…	…	…

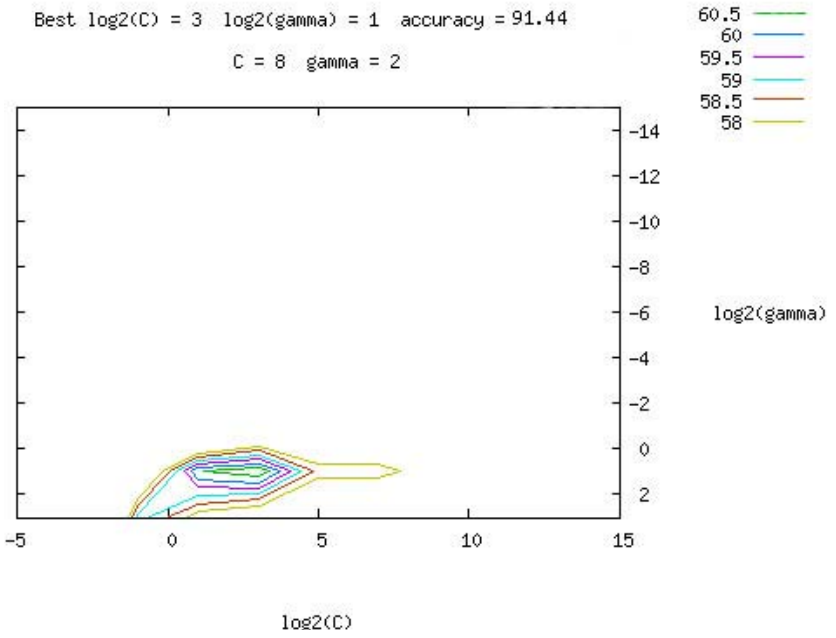

Fig.2 Adjusting C value and gamma value

Tab.3 Main parameters of the prediction model

The type of support vector machine	c_svc
Kernel function	Radial basis kernel
Parameter gamma of Radial basis kernel	2.0
The number of support machine	160
Bias of decision function *b*	0.424
Penalty parameter C value	8.0
Kernel parameter gamma value	2.0

5 Conclusions and discussion

Passengers are the main source of profit when running high-speed railway. It's necessary to pay attention the demands of passengers, which help high-speed railway improve their competitiveness in the transportation market. Only in this way can high-speed railway enlarge their market share.

In the research, we analysis the travel behavior of passengers, we find that to passengers, what affect their travel choices most are six factors: price, speed, train time, environment, safety and overall satisfaction. Those six factors are in priority to other factors when high-speed railway develops. Measures, such as appropriate reduction of price, development of new technology, or improving the professionalism of servers should be taken into consideration. The research is based on support vector machine, which is applied well in the situation of small sample. Using support vector machine to build a prediction model, we finally get a satisfying result, whose accuracy is 91.44%.

When we make prediction, it is inevitable to have deviation. Although the prediction result is relatively satisfying, there are still some restrictions in this research, which reduce the accuracy of prediction.

(1) The survey was conducted during spring festival. As all we known, spring festival is the time of passenger flow peak. Under such a special situation, people will pay more attention on the travel time and speed compared than usual time, and less sensitive to the environment, price and seat level, for they have the urgency of gong back home. So the real demands may be weaken. For example, a person who just buys the ticket of seat in second level may buy a ticket of seat in first level with much more money than usual, just for going back home.

(2) Research in this paper use support vector machine to build a prediction model. The prediction result of the exact rate is 91.44%. In this paper, we just classify the travel choice to two- class: high-speed railway and other choices, including plane, train and bus. If we transform the two–class prediction to multiclass prediction, it will offer more valuable suggestion for the field of transportation. But to make multiclass prediction is not easy, we should try to construct some two-class machines and introduce voting mechanism into the prediction model. So multiclass is worthy of being studied in the future.

Supported by "the Fundamental Research Funds for the Central Universities".

References

[1]The development of the world's high-speed railway. Railway survey and designer, 2006, 1: 54-56. (in Chinese)

[2]GONG Xin, YU Xin-an, OUYANG Wei-hong. Review of developing high-speed railway[J]. Study Paper of Railway 1994, 16(1): 124-128. (in Chinese)

[3]MAO Bao-hua, LI Zhen, ZHOU Lei-shan, XIE Hia-hong. The analysis of high-speed railway passengers' demands[J]. Great Technology, 1996,2: 13-16.(in Chinese)

[4]CHEN Zhang-ming, JI Xiao-feng. The study on railway passengers' travel activities[J]. Transportation and Economic of Railway, 2008, 30(11):23-25. (in Chinese)

[5]WU Qun-qi, XU Xing. The study on the mechanism of travel choice[J]. Study Paper of Changan University (social and scientific version), 2007,9(2):13-16.(in Chinese)

[6]QIANG Li-xia, YAN Ying. The study on deficiency between transportation at home and abroad[J]. Transportation and Economy of Railway, 2006, 28(9): 18-21. (in Chinese)

[7]LIU Wei-guo, Hu Si-ji. Chance Constrained programming model with fuzzy parameters for passenger traffic mode choice behavior[J]. Journal of Northern Jiaotong University, 2002(2). (in Chinese)

[8]XU Xing. The principle of transportation method choosing of passengers when they go out[D]. Changan University, 2008.(in Chinese)

[9]ZHU Da. Research on the value of travel time based on travel decision-making[D]. Beijing: Beijing Jiaotong University, 2008.(in Chinese)

[10]SHI Feng, DENG Lian-bo, HUO Liang. Travel choice of railway passenger and its utility[J]. China Railway Science, 2007, 28(6).(in Chinese)

[11]Harata Noboru, Morikawa Takayuki, Yai Tetsuo**.** Review and perspective of travel behavior analysis focusing on disaggregate travel demand models[J]. Proceedings of the Japan Society of Civil Engineers, 1993, 470:97-104.

[12]PANG Ming-bao, HE Guo-guang. Real-time intelligent recognition of chaos in traffic flow using reduced support vector machine[C]//International Conference on Wireless Communication, Networking and Mobile Computing, 2007: 5667-5670.

[13]CHEN Yao, YANG Qing-peng, van Zuylen H J. Travel time prediction on urban networks based on combining rough set with support vector machine[C]//Conference on Logistics System and Intelligent Management, 2010:586-589.

[14]ZHANG Hai. Discussion about data mining and tools of mining- support vector machine[J]. Industrial Technology, 2009,38(2):50-51.(in Chinese)
[15]BING Zhang-xian, XIAO Hai-bo. Support vector machine and its application[J]. Fujian Computing, 2007, 4: 110.(in Chinese)
[16]Nello Cristianini, John Shawe-Taylor. An introduction to support vector machines and other kernel-based learning methods[M]. House of Electronic Industry, in press, 2004.
[17]ZHANG Gen-ming, XIANG Xiao-yi. Financial forecasting model based on support vector machine [J]. Science and Technology Management Research, 2007, 4:235-242.(in Chinese)
[18]LI Peng-lin, MENG Yan. Residential consumption forecasting model based on support vector machine [J]. Statistic and Information Forum, 2009, 44(12): 89-91. (in Chinese)
[19]YE Wei, WANG Shi-long. Predicting model of cutting tool wear based on squares support vector machine and MATLAB simulation[J]. Tool Technology, 2009, 43(10):43-45. (in Chinese)
[20]YANG Dong-yun, LI Shu-han. The method of building support vector machine function[J]. Scientific Study Magazine, 2010, 30(2):25-27. (in Chinese)
[21]WU Wei, SENG De-wen. The application of support vector machine in corporation credit rating[J]. J. Zhejiang Wat. Cons and Hydr. College, 2009, 21(4):55-57.(in Chinese)
[22]SHI Guang-ren. The application of support vector machine in multi-geological factors[J]. ACTA Petrolei Sinica, 2008, 29(2). (in Chinese)