Application of SVM model to environmental resource analysis in tourism development

Jing Han\(^1\)

\(^1\)Shandong Management University, Shandong, China

Corresponding author and e-mail: Jing Han, hanjing@sdmu.edu.cn

Abstract. With the improvement of economic and living standards, the demand for tourism is increasing. However, the development of tourism brings many problems to the environment, such as aggravating the environmental pollution and reducing the available environmental resources. At the same time, good environmental resources can stimulate local tourism development and economic growth. This paper uses support vector machine (SVM) model to establish the nonlinear relation model between tourism market turnover and various influencing factors, analyses the development of present Chengdu’s tourism market, and predicts the market turnover in three years to provide quantitative reference for the analysis of environmental resources in tourism development.

1. Introduction

With the rapid development of tourism, more and more people choose to go to tourist attractions. The environmental conditions of tourist sites and various environmental problems, such as air pollution, water pollution and environmental pollution, will affect the willingness of tourists to travel there, which is becoming increasingly prominent. Environmental resources are the basis and main components of tourism development. For a long time, people believed that tourism environment was used for free, until the sustainable theory was proposed [1]. Tourism resources and environmental quality of tourist destinations are always the foundation for the survival and development of tourism. In recent years, due to the favorable conditions brought by the reform and opening up policy and the rapid economic development, the tourism industry has developed fast and produced considerable economic benefits. However, people always regard the development of tourism as an economic activity, lay particular stress on the pursuit of its economic benefits, but ignore the impact of tourism on the environment.

The development of tourism and the protection of environmental resources are interdependent. Effective protection of environmental resources can better promote the development of tourism, and tourism should also do a good job in environmental protection while developing rapidly. Shi Xiaohong put forward six countermeasures for environmental protection of tourism: perfect the legal system of environmental protection of tourism, strengthen the scientific research of environmental protection of tourism, apply the aesthetic point of view, do well the environmental planning of tourism scenic spots, study the environmental capacity, ensure the sustainable use of tourism resources and increase environmental protection funds. Environmental protection is an important issue for the whole nation. We should strengthen publicity and education, and take feasible protection measures [2]. Aiming at the problems in the development of tourism resources development and environment, Wu Chucai believes that environmental resources should be fully developed and exploited, making it an important tourism
resources, and one of the core attractions of the tourist area. Also we should pay attention to the development to make the development of more beneficial to health of body and mind of resources, avoid harmful factors, and hope that the state formulates the corresponding evaluation standard, make more effective use of resources [3].

2. Principles and algorithms

In view of the limited samples in the actual situation, statistical learning theory was put forward in the 1960s. It has complete theoretical support and strict theoretical system. In 1995, Vapnik et al. introduced a new machine learning tool, the Support Vector Machine (SVM), which is based on the structural risk minimization principle of statistical learning theory and the VC dimension theory. It has the following advantages: strong ability of extrapolation, not easy to fall into the local minimum and independent of dimension.

Support vector machine is essentially a forward neural network, which transforms the input space into a high-dimensional space by using the nonlinear transformation defined by the inner product function, and finds the nonlinear relationship between the input variables and the output variables in the high-dimensional space. Support vector machines are divided into linear separable support vector machine and nonlinear support vector machine. For the linearly separable model, only the hyperplane with the maximum interval that separates the two types of samples needs to be found, as shown in Figure 1, where \( w \) is the normal vector. \( w \cdot x + b = 0 \) is the maximum interval hyperplane, and \( w \cdot x + b = \pm 1 \) is the edge hyperplane parallel to it, and the interval between two adjacent hyperplanes is \( 2/\|w\| \), also known as the classification interval. Therefore, the problem of solving the hyperplane with the maximum interval becomes to make the classification interval maximum, that is \( \|w\| \) minimum.

![Figure 1. Linear separable support vector machine.](image)

For a nonlinear support vector machine, the so-called nonlinear separability means that the hyperplane cannot directly separate two classes of samples. A kernel function must be used to map the sample to a higher dimensional space to make it separable. As shown on the left of Figure 2, hyperplane cannot be used to separate the two types of samples. Then, a nonlinear mapping, namely kernel function, can be selected to map the samples to the high-dimensional feature space, and then the maximum interval hyperplane can be found in the space, as shown on the right of Figure 2, to accurately separate the two types of samples.
3. Experimental results and discussion

3.1. Experimental design

The experimental data set is divided into two parts, the training set and the test set. This paper collected from 2008 to 2014 of Chengdu environmental resources, including the annual tourism turnover of Chengdu, annual average temperature, annual rainfall, annual average relative humidity, park area, forest area, year-end resident population, good air quality days and costs of energy conservation and environmental protection, as the training data set. Most of the sample data are collected from Sichuan statistical yearbook and environment quality bulletin; The test set is corresponding data from 2015 to 2017. The experiment is carried out in two steps. The first step is to obtain the training model with the training algorithm, which is to solve the maximum interval hyperplane mentioned above. The second step is to predict the test set of the data set and verify the effectiveness of the algorithm. The error is determined by the following formula:

\[
\text{error} = \frac{\text{true value} - \text{predicted value}}{\text{true value}} \times 100\%
\]

SVM method can solve the problem of small sample, nonlinear, high dimension, and is often used to identify and predict. This experiment is implemented in MATLAB, using the libSVM toolbox for SVM regression. The penalty factor was selected as 28.011365319071743, and the kernel function was 0.066324924195784. The experimental data set is shown in Table 1. At the same time, after all the data are normalized, eight different influencing factors and the broken line chart of the tourism turnover are established (shown in Figure 3). The horizontal axis is the tourism turnover of Chengdu (100 million yuan), and the vertical axis is the eight different influencing factors. It can be seen from the figure that some factors affected by random conditions, such as temperature, days with good air quality, rainfall and trading volume, have no obvious linear correlation, while others, such as park area, forest area and resident population, have certain positive correlation.

Table 1. Tourism turnover and environmental resources information of Chengdu from 2008 to 2017.

| Year | Tourism turnover in Chengdu (100 million YUAN) | Temperature (degree) | Rainfall (mm) | Relative humidity (%) | Park area (Hectare) | Days with good air quality | Forest area (10,000 mu) | Year-end resident population (10,000) | Energy conservation and environmental protection expenses (10,000 yuan) |
|------|-----------------------------------------------|----------------------|---------------|----------------------|---------------------|--------------------------|-------------------------|--------------------------------------|---------------------------------------------------------------|
| 2008 | 363.6                                         | 16.3                 | 1028.2        | 75                   | 2243                | 319                      | 789                     | 1270.6                              | 88179                                          |
| 2009 | 485.2                                         | 16.8                 | 724.2         | 74                   | 2080                | 315                      | 793                     | 1286.6                              | 90980                                          |
| 2010 | 584.6                                         | 16                   | 936.8         | 79                   | 2562                | 316                      | 798                     | 1404.8                              | 101364                                         |
| 2011 | 776.4                                         | 15.9                 | 1003.2        | 74                   | 2859                | 322                      | 803.7                   | 1407.08                             | 128235                                         |
3.2. Results and discussion

Through the analysis of the established SVM model, it was found that the relevant environmental data did not have a good linear relationship, indicating that the environment-related data were relatively complex, but some of the data had a great and linear correlation with the trade volume of the tourism market, such as forest area, population number and investment cost for environmental protection. Due to the influence of many factors, the prediction result is multiplied by a fitting parameter, and the value of the fitting parameter can be calculated according to the optimization algorithm. As some environmental factors are weakly correlated, such as temperature, air quality, environmental suitability and other factors, they also contribute to the forecast results, which indicates that the government's investment in tourism and environment makes tourism transactions achieve better results.

In order to verify the reliability of the algorithm, the predicted value of tourism turnover in Chengdu from 2015 to 2017 is compared with the actual value shown in Table 2. The results show that the SVM predicted values in these three years are 208.7, 262.4 and 290.3 billion yuan, and the relative errors with the actual tourism turnover published on the official website of Sichuan Statistics are -2.3%, -4.9% and 4.3% respectively, and the relative errors are all less than 10%, indicating a good prediction effect.

![Figure 3. The relationship between the influencing factors and the tourism turnover of Chengdu.](image-url)
In the first data processing, it can be found that the prediction result is not ideal, and if apply it to predict more years, the change trend may be lower than the actual data. This is because training data is too small, leads to too few neural networks, prediction accuracy is not well, only three dimensions of each data, and the correlation between the input data is not high enough, few years of official data was published. So it is difficult to find a large amount of data for training. Therefore, the method of data expansion is considered to improve the original method. In the process of data expansion, the expansion quantity and output quantity can be set separately. This experimental design adopts data expansion for ten times and outputs three points of data (namely, three years of data volume). The actual output results are as follows.

Table 2. Comparison of actual tourism turnover in Chengdu from 2015 to 2017 with SVM predicted value.

| Year | Actual tourism turnover in Chengdu (100 million yuan) | Predicted tourism turnover in Chengdu (100 million yuan) | Relative error (%) |
|------|------------------------------------------------------|----------------------------------------------------------|--------------------|
| 2015 | 2040                                                 | 2087                                                    | -2.3               |
| 2016 | 2502                                                 | 2624                                                    | -4.9               |
| 2017 | 3033                                                 | 2903                                                    | 4.3                |

In the first data processing, it can be found that the prediction result is not ideal, and if apply it to predict more years, the change trend may be lower than the actual data. This is because training data is too small, leads to too few neural networks, prediction accuracy is not well, only three dimensions of each data, and the correlation between the input data is not high enough, few years of official data was published. So it is difficult to find a large amount of data for training. Therefore, the method of data expansion is considered to improve the original method. In the process of data expansion, the expansion quantity and output quantity can be set separately. This experimental design adopts data expansion for ten times and outputs three points of data (namely, three years of data volume). The actual output results are as follows.
Figure 5. Comparison of actual tourism turnover in Chengdu from 2015 to 2017 with the improved SVM predicted value.

Table 3. Comparison of actual tourism turnover in Chengdu from 2015 to 2017 with improved SVM predicted value.

| Year | Actual tourism turnover in Chengdu (100 million yuan) | Predicted tourism turnover in Chengdu (100 million yuan) | Relative error (%) |
|------|------------------------------------------------------|--------------------------------------------------------|--------------------|
| 2015 | 1011                                                 | 948.9                                                  | 6.1                |
| 2016 | 1330                                                 | 1267                                                   | 4.7                |
| 2017 | 1617                                                 | 1554                                                   | 3.9                |

Among them, due to the small amount of data, data expansion leads to the transformation of results. The final prediction result and the actual result are different from the first time. It can be seen from the data that the predicted trend is almost consistent with the actual result, indicating that this method can predict more years and have good effect.

4. Conclusions
Based on the tourism turnover of Chengdu from 2008 to 2014 and the data of eight additional factors related to environmental resources, this paper uses an improved SVM method to predict and analyze the tourism turnover from 2015 to 2017. A general SVM analysis and an improved SVM analysis were performed. The result error of both times is less than 10%, which verifies the reliable prediction performance of SVM method. At the same time, in the improved SVM prediction analysis, data expansion is introduced to increase the data volume of the training set. In this case, the predicted model for three consecutive years is almost consistent with the change trend of the actual data, and the error is well controlled within 10%. Since most of the environmental resource variables are uncertain and the linear relationship with the tourism turnover is not obvious, the SVM method should be improved according to the actual situation in the prediction and training.

References
[1] Li yuxin. 2012. Resource and environmental economics analysis of environmental Problems in tourism development. China Tourism Review.
[2] Shi xiaohong. 2009. Brief analysis of environmental problems and countermeasures in tourism development. Statistics and Consultation, (02):14-15.
[3] Wu chucai, Wu zhangwen. 2008. Environmental resources are important tourism resources. Summit Forum on Ecological Tourism and Ecological Civilization.
[4] Wang li. 2018. Quality and predictive analysis of big data in e-commerce. Master's thesis, Nanjing University of Posts and Telecommunications.