

Research on Earning Warning of Railway Freight Based on Support Vector Regression (SVR)

Lingyan Cheng^{1, a}, Shiwei He^{1, b} and Weichuan Yin^{1, c}

¹ School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China.

^azjlycheng@163.com, ^bshiweihe@bjtu.edu.cn, ^c13120824@bjtu.edu.cn

Keywords: Earning Warning, Railway Freight, SVR

Abstract. The traditional railway freight early warning adopted independent indicator thresholds discrimination. But it would lead to inaccurate results warnings, it may alarms when abnormal events occur instead of early warning. This paper proposes a railway freight early warning model, which apply support vector regression (SVR) to rail freight early warning. this model could find their variation, sum up the formation of abnormal alarm mod by analyzing historical indicators, including of number of booking railway freight vehicles, loading vehicles, freight volume and freight turnover quantity. Early warning signals will be given when an exception occurs at the beginning of freight. Early analysis and processing potential risk are as to stabilize railway freight market. Experimental result shows a high prediction accuracy of this model for the railway freight market.

1 Introduction

Currently, the traditional railway freight early warning adopted independent indicator thresholds discrimination, which lacks a comprehensive analysis of internal market rules. To ensure a smooth development of the railway freight market, there is a need to establish early warning system. Support vector machine (SVM) is pattern classification method based on the minimization principle of statistical learning theory structure risk, which was proposed by Vapnik et al in 1995 [1]. Support vector regression (SVR) is an application of SVM in regression estimation problem. Compared with SVM, SVR also has the merit of complete mathematical theory, global optimization and minimization principle of structure risk. Meanwhile, fitting and complexity of training samples are taken into account to achieve good promotion performance. Thus this paper proposed a model railroad freight early warning model based on support vector regression, which would alarm upcoming abnormal situations and has high feasibility and accuracy.

2 Support vector regression (SVR) Overview

Support vector machine is a small-sample machine learning method, which is more suitable for nonlinear regression problems. Through a variety full description of typical space, SVR characterizes the relationship between factor group and prediction objects. It is suitable for handling non-linear nature problem since it can identify and extract key information automatically from massive information.

In 2007, Mingbao use least square support vector regression model of least squares to predict the traffic flow based on phase space reconstruction fuzzy rules. The simulation results show that this model is feasible, and the prediction accuracy is more accurate than other models [2]. Wenwu He predicted power load through support vector regression model with global optimization, which shows a higher model prediction accuracy in 2008 [3]. And in 2009, support vector regression model with chaos ant colony optimization algorithm is applied by Wei-Chiang Hong to forecast the power load, the experimental results show that this model is superior to single support vector regression model [4]. In the same year, Wei-Chiang Hong adopted support vector regression model with immune algorithm and achieved better lithium load forecasting results than SVM and neural network model [5].

In a nutshell, input space can be converted to a high-dimensional space by SVR through a nonlinear transformation defined by the inner function. And SVM is the learning process to solve

regression function in this high-dimensional space. SVR is similar to the neural network in form. Output is a linear combination of intermediate nodes, each intermediate node corresponds to a support vector, whose weight is the corresponding Lagrange multiplier. Support vector regression structure shows in Fig.1.

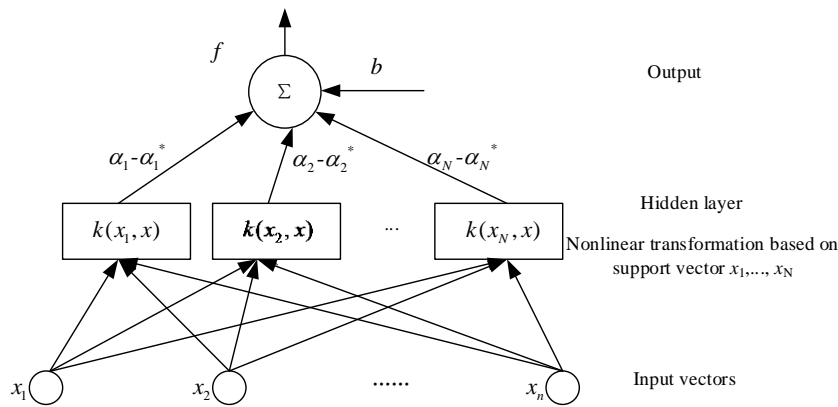


Fig.1 Support vector regression structure

3 Modeling and solution

3.1 Feasibility Analysis

(1) As there are many factors that affect the railway freight. Meanwhile, the relationship between factors is complicated. The relation between input and output is nonlinear when early warning modeling, so it's difficult to achieve a satisfied result by general mathematical analysis methods. As a machine-learning technique with strong adaptability, the support vector machine could handle complex nonlinear problems, dig out the information and rules implied in the sample through the learning of limited learning samples under the criterion of minimum structural risk. So this paper use SVM to early warn the freight volume.

(2) The biggest advantage of support vector machine is learning algorithm for limited sample, especially the problem of small sample. As the optimal method of regression analysis and classification of small samples, SVM could effectively solve the puzzle of early-warning of railway freight, which is the difficulty to obtain sample data, and the lack of sample size.

(3) Currently, support vector machine is the latest and best method to solve regression problem based on time series analysis and classification problem based on pattern recognition. Since the early warning of railway freight can be a time-series problem, but also a process of regression, so SVR to solve this problem is a good choice.

(4) While training support vector machine, we need to convert it into a quadratic programming problem. The problem has only one solution to avoid local minima problems and ensure the stability of training results, with obvious global approximation. Change process of railway freight has global superiority. The change of one factor could lead to severe fluctuations of railway freight.

3.2 Modeling based on SVR

The process of establishing rail freight early warning model is divided into three stages: the first stage is the data preprocessing. De-noise value processing is used for existing freight history data. Then the data is normalized. After that, select the appropriate set as the test sample data; the second stage is to select the parameters of support vector regression. Determining the kernel function and penalty factor test to ensure the highest accuracy of sample prediction; the third stage is to obtain warning results with SVR on the basis of the test sample data as input space. Specific steps as follows:

1. Data preprocessing. Select the appropriate set of data as production sample data, including:

(1) Fill in missing values, smooth the data, remove noise value, and exclude abnormal situation caused by human factors.

(2) Data normalization process. Scaling attribute data to fall within a specific small range. This paper the minimum - maximum standardization:

$$v_i' = \frac{v_i - \min_A}{\max_A - \min_A} (\max_A' - \min_A') + \min_A' \quad (1)$$

\max_A and \min_A are maximum and minimum value of attribute A. Map the value of A to the interval $[\max_A', \min_A']$ by calculating. The new interval is $[-1,1]$.

2. Select the kernel function and penalty coefficient c. Key issue is the choice of kernel function in SVR [6]. Linear regression in high-dimensional space can be corresponded to nonlinear regression in low-dimensional space in the kernel function. Then non-linear regression can be solved under the same computational complexity. The most commonly used kernel functions are:

(1) Linear kernel: $K(x_i, x_j) = (x_i, x_j)$ (2)

(2) Polynomial kernel: $K(x_i, x_j) = [(x_i, x_j) + 1]^n$ $n = 1, 2, \dots$ (3)

(3) Radial basis function: $K_\gamma(x_i, x_j) = \exp\{-\gamma |x_i - x_j|^2\}$ (4), where γ is the parameter;

(4) Sigmoid kernel function [7]: $K(x_i, x_j) = \tanh(\gamma(x_i \cdot x_j) + c)$ (5), the formula $\gamma > 0, c > 0$.

3. Test the SVR model with samples, and finally railway freight early warning models.

4 Simulation results analysis

First, select the number of booking railway freight vehicles, loading vehicles, freight volume and freight turnover quantity as the early-warning indicators. Respectively, we acquired the data of various indicators from Jan. 2013 to February 2015. Normalization processing was made because each dimension index of indicator is not uniform. Then, expert scoring method was used in early-warning classification pretreatment of monthly data, where the early-warning types were divided into four categories [8]: none, low warning, middle warning and serious warning. Each value of warning type are as follows: none warning is (0, 0.3], the warning value can be 0.15, the low warning is (0.3, 0.6], the warning value can be taken as 0.45, the middle warning is (0.6, 0.8], the warning value can be taken as 0.70, and the serious warning is (0.8, $+\infty$], warning value can take 0.90.

K-fold cross-validation method is used to select the appropriate support vector regression model parameters with the combination of grid search algorithm. Specific steps shown in Figure 2.

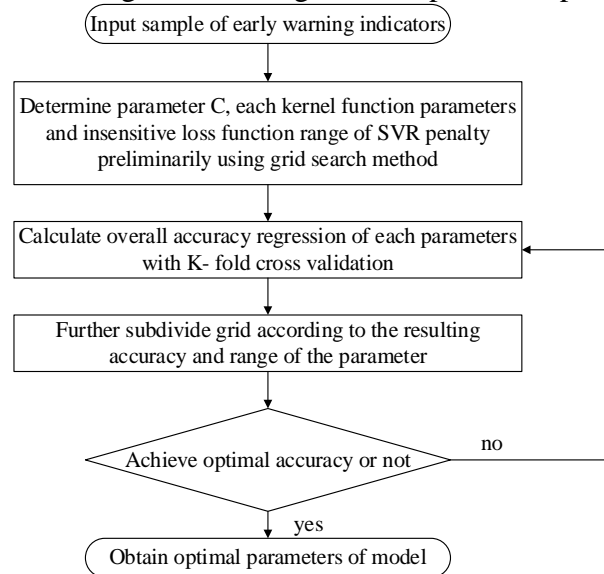


Fig.2 selection steps of early warning model using K-fold cross-validation method and grid search algorithm

By calculation, the value of parameter C of the model is 1000, insensitive loss function coefficient is 0.01. Use Eclipse6.0 software, warning model was run in the Java EE platform. Run time is less than 10s, the number of iterations is 11333, the number of support vector is 26, the results in the table

below. 24 support vector regression warning value were consistent with expert scoring situations. For the total number of situation is 26, then the early warning accuracy will be $\frac{24}{26}=92.31\%$.

Table 2-1 Normalized value of warning indicators and type of warning

| time | freight turnover quantity | reight volume | number of loading | number of booking | Warning value (actual) | Warning value (SVR) | Warning Type (actual) | Warning Type (SVR) |
|--------|---------------------------|---------------|-------------------|-------------------|------------------------|---------------------|-----------------------|--------------------|
| 201301 | 0.9920 | 0.9804 | 0.9809 | 0.9979 | 0.1500 | 0.149339422 | none | none |
| 201302 | 0.8562 | 0.8818 | 0.9681 | 0.9561 | 0.1500 | 0.25168626 | none | none |
| 201303 | 0.9818 | 0.9673 | 0.9658 | 0.9508 | 0.1500 | 0.192232424 | none | none |
| 201304 | 0.9000 | 0.8886 | 0.9243 | 0.8951 | 0.4500 | 0.354143708 | low | low |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 201404 | 0.8410 | 0.8644 | 0.8977 | 0.8512 | 0.7000 | 0.529583077 | middle | low |
| 201405 | 0.8846 | 0.9132 | 0.9108 | 0.8665 | 0.4500 | 0.49547028 | Low | low |
| 201406 | 0.8528 | 0.8850 | 0.9045 | 0.8602 | 0.4500 | 0.524298556 | Low | low |
| 201407 | 0.8630 | 0.8934 | 0.8929 | 0.8424 | 0.7000 | 0.569082098 | middle | low |
| 201408 | 0.9015 | 0.9300 | 0.9231 | 0.8755 | 0.4500 | 0.44496069 | low | low |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 201501 | 0.8918 | 0.8896 | 0.8834 | 0.8425 | 0.4500 | 0.543337122 | low | low |
| 201502 | 0.6931 | 0.7506 | 0.8193 | 0.7615 | 0.9000 | 0.908486252 | serious | serious |

Summary

Application of early warning model of railway freight based on support vector regression model show that this method has the advantages of fast convergence, high efficiency and good accuracy. It can be effective to achieve real-time early warning in railway freight, and provide the basis for the relevant department decisions and prevention.

References

- [1] Naiyang Deng, Yingjie Tian. New Method for Data Mining-- Support Vector Machines [M]. Beijing: Science Press, 2004: 51-52.
- [2] Ming-bao, PangGuo-guang, HelCAL. Traffic Flow Predicting of Chaos Time Series Using Support Vector Learning Mechanism for Fuzzy Rule-based Modeling [A]. Automation and Logistics, IEEE International Conference 2007: 666~670
- [3] Wenwu He. Forecasting electricity load with optimized local learning models [J]. Electrical Power and Energy Systems, 2008, 30(10): 603~608
- [4] Wei-Chiang Hong, Yucheng Dong, Li-Yueh Chen, Chien-Yuan Lai. Taiwanese 3G mobile phone demand forecasting by SVR with hybrid evolutionary algorithms [J]. Expert Systems with Applications, 2010, 37(6): 4452~4462
- [5] Wei-Chiang Hong. Electric load forecasting by support vector model [J]. Applied Mathematical Modeling, 2009, 33(5): 2444~2454
- [6] Zhuo Li, Bin Liu, Tienan Liu. The Research on Fault Diagnoses of Oil Pump Based On Support Vector Machine [J]. Control & Automation, 2006, 22(7):209-211.
- [7] CHIH-Wei Hsu, CHIH-Chung Chang, CHIH-jen Lin. A Practical Guide to Support Vector Classification [J]. Taiwan. Na-tional Taiwan University. 2008.
- [8] Zhi-wei Xing, Hui Zhang. Aircraft Icing Type Classification Forecast Based on Support Vector Machine [J]. Computer Technology and Development, 2012, 22(6):247-250.