Prediction of E-Commerce Credit Rating based on PCA-SVR

Zhuoxi Yu, Faculty of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun, China
Huansen Zhang, Faculty of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun, China
Zhiwen Zhao, College of Mathematics, Jilin Normal University, Siping, China
Limin Wang, Faculty of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun, China

ABSTRACT

Using the principal component analysis (PCA) and support vector regression machine (SVR) in predicting the credit rating of online stores. Collects 14 variables, including 982 observations of dress shops in Taobao. Firstly, the authors use the method of PCA to filter and reduce the dimension of the data, and obtain five factors, namely, the evaluation factor, the traffic factor, the price factor, the preferential policies factor and the reliability factor. Secondly, they use the PCA’s output as the input of the SVR for the credit rating prediction. Thirdly, they extract 300 as the training samples and 150 as the test samples from the data, and utilize GA algorithm for parameter optimization in order to improve the prediction accuracy of SVR. Finally, carry on an empirical test. The result shows that this combination method is accurate and effective in prediction rate than the consequences of the traditional SVR, and it is valid and feasible.

KEYWORDS

Credit Rating, E-Commerce, PCA, SVR

INTRODUCTION

The credit is referred to the accumulation of the bilateral or multi-agent participation in long-term trust and honesty originally. It is a kind of behavior to transfer part of the right, and to get the appreciation of the other part of the ownership of the property in the future. It’s usually characterized by the transacting parties in the current transfer of value, and to obtain cash flow for the future (28:72-74). With the vigorous development of e-commerce, the e-commerce credit rating has come into being. It means the credit relations between the multiplicity interaction of buyers, the sellers, e-commerce platform providers and e-commerce service providers of each link in the transactions, to maintain and expand the ability of the trust relationship by satisfying the reasonable expectations among the transaction. It can also be understood as the degree of compliance of the market agreements and commitments during the e-commerce activities of all concerned parties.

In the virtual electronic marketplace, the goods cannot be touched, and most traders are anonymous. For the buyers, it’s hard to know the sellers’ true identity and the quality of the goods (05:49-52), so the buyers would like to choose the shop with a higher credit. The current credit evaluation algorithm is based on the transaction rather than the amount, so some businesses use a large number of small amounts to improve the credit as soon as possible. In this way, they can enhance the credit rating in a short time significantly, so that buyers are notoriously difficult to judge which shop
has a real higher credit. In this respect, the existing credit evaluation system still exists an obvious deficiency. Some areas should be stepped up and perfected, to impel e-commerce sustained, rapid and healthy growth. Therefore, it’s an important topic to explore the e-commerce credit rating issue, not only has an essential value of theory-developing but also has a significant influence on reality.

It has been more than twenty years since the world’s first network electronic trading completed. During this period, domestic and overseas scholars have done a mass of research in e-commerce. However, there are a few achievements on the research of the e-commerce credit problem. Most studies have focused on the formation mechanism and the mechanism of e-commerce credit risk. Currently, domestic methods for the e-commerce credit problem are mainly including comprehensive evaluation, multivariate statistical and artificial intelligence (04:70-74, 05:1882-1885, 03:508-514). Zhang (2014) used the cumulative weighted scoring method to measure the e-commerce system (03:49-53), and Zhang (2007) classified the e-commerce credit risk by using multivariate statistical analysis (07:79-81). Now, domestic has never had a complete evaluation system, but Wei (2014) established a set of an index system from the perspective of online trading (02:254-260). Foreign researchers focus on the application of e-commerce credit risk. They are using mathematical modelling to credit assessment, Chenga&Chiangb established the personal credit evaluation model by using DEA method, Wang (2008) used the Markov Chain to analyze and evaluate the e-commerce credit risk.

Nowadays, at home and abroad, the commonly used evaluation methods of e-commerce credit risk including the comprehensive evaluation method, the improved fuzzy mathematics and the statistics analysis method, but there are a few studies focus on the artificial intelligence. While, artificial intelligence algorithm has been widely applied in the credit evaluation since the 1980’s, for the advantage of the learning mechanism, adjust automatically and robustness (04:311-317, 09:6233-6239, 07:1114-1123). Since the 1990’s, the researchers found the validity of the support vector machine (SVM) as a new classification of credit scoring (01:127-133, 07:4902-4909, 05:5325-5331). The e-commerce data indexes are multifarious, and the dimensions are numerous, so using principal component analysis (PCA) to reduce the dimension of the raw data. It is very important to choose the parameters when using SVR to solve the nonlinear regression problems. There can be a huge difference between different parameters. For now, to solve the problem of parameter selection, genetic algorithm and particle swarm algorithm are wildly used for automatic parameter optimization (09:1430-1433, 06:33-39, 06:1309-1313, 05:54-62, 04:593-598), to make the model getting a higher accuracy and to make the prediction result more accurate and reliable.

METHODS INTRODUCE

Principal Component Analysis (PCA)

The principal component analysis method is mainly using the ideas of dimension reduction. Dimensions of the feature space are significantly reduced without excessive information loss.

Usually, we call the transformed comprehensive indexes as main components, each main component can be expressible in terms of a linear combination of the original signals. Each main component are uncorrelated, and have some superior performance than the original variables (Jolliffe I, 2014; He XQ, 2010; DiamantarasKI, 1996; BaiXS, 1995; Lin HM, 2007).

The basic steps can be summarized as:

1. Make the Z-Score normalization in order to eliminate the differences in magnitude;
2. Calculate the correlation matrix;
3. Compute eigenvalues, eigenvectors and variance contributions of the correlation matrix;
Impact of Digital Revolution on the Structure of Nigerian Banks
www.igi-global.com/article/impact-digital-revolution-structure-nigerian/37398?camid=4v1a