Review of Application Research of Support Vector Machines and Rough Sets in Credit Risk Evaluation

Li DONG\textsuperscript{1,2,3,a} and Mu ZHANG\textsuperscript{1,b,*}

\textsuperscript{1}School of Finance, \textsuperscript{2}Guizhou Institute for Urban Economics and Development, \textsuperscript{3}Guizhou Institution for Technology Innovation & Entrepreneurship Investment, Guizhou University of Finance and Economics, Guiyang Guizhou 550025, China
\textsuperscript{a}954388937@qq.com, \textsuperscript{b}rim_007@163.com
\textsuperscript{*}Corresponding author

Keywords: Credit risk evaluation, Support vector machines, Rough sets, Artificial intelligence.

Abstract. Theoretical and empirical research on credit risk evaluation plays an important role for enterprises in reducing transaction costs, enhancing competition ability and creating a good competition environment. At present, in view of the application of support vector machines and rough sets in artificial intelligence method in enterprise credit risk evaluation, research abroad mainly focuses on the innovative application of support vector machines and rough sets; and the domestic research mainly focuses on the effectiveness, the advantages and the feasibility of support vector machines and rough sets. In this paper, the research status of support vector machines and rough sets at home and abroad is discussed in detail so that the development of artificial intelligence theory and practice is promoted.

Introduction

The method of credit risk assessment can be traced back to 1930s, and it gradually became a hot spot after 1960s. The method of credit risk evaluation, which was widely adopted by commercial banks, is the expert analysis method. Since 1980s, international commercial banks and investors are facing unprecedented challenges of credit risk. All kinds of new credit risk evaluation methods emerge in endlessly, and credit risk management theory and technology are improving day by day [1]. Generally speaking, the methods of credit risk assessment have undergone a gradual process from simple to complex, from qualitative analysis to quantitative calculation and from credit risk evaluation of single asset to credit risk evaluation of portfolio [2]. The traditional credit risk evaluation method, multivariate statistical analysis method, artificial intelligence methods and models of credit risk measurement appeared successively.

Artificial Intelligence (AI) is a new technological science that studies, extends and develops theories, methods, techniques and application systems. Artificial Intelligence is a branch of computer science that tries to understand the nature of intelligence and produces a new intelligent machine that responds in a similar way to human intelligence. Since the birth of artificial intelligence, the theory and technology have become more and more mature, and the field of application has also been expanding. Now, it has been gradually applied to credit risk evaluation. In this paper, the application status at home and abroad of expert system and neural network in credit risk evaluation is discussed in detail so that the application of artificial intelligence methods in credit risk assessment is promoted.

Support Vector Machines

Support vector machines (SVM) is a new general learning method, which is developed on the basis of statistical learning theory. It has both strict theoretical foundation and good solution to some practical problems such as small sample and nonlinear. And it has already shown a lot of performances better than existing methods. The basic idea of support vector machines (SVM) used in pattern recognition is that a hyper plane is used as the decision plane, which not only can divide the
two classes of samples correctly (the training error rate is 0), but also can make the classification interval maximum. It means that it transforms the optimization problem into a convex two programming problem, which is solved by the standard Lagrange multiplier method, and the result is the global optimum and has high precision [3].


Rough Sets

Rough Set (RS) was a data analysis theory proposed by Z. Pawlak, a scholar of Poland in 1982. It is a mathematical tool to describe incompleteness and uncertainty. It can effectively analyze and deal with incomplete information such as inaccuracy, inconsistency and incompleteness. The most notable feature of RS theory is that it does not need to deal with any prior information outside the relational table. It has been successfully applied to machine learning, decision analysis, process control, pattern recognition and data mining and so on [28]. There are some application steps of rough set in credit risk evaluation: data preprocessing, attribute reduction, decision rule extraction and classification discrimination. In recent years, the application research of rough sets in credit risk evaluation is becoming increasingly prosperous. The research of foreign scholars such as Dimitras et
al (1999) [29], Malcolm and Michael (2001) [30], Tay and Shen (2002) [31] and etc has a certain representativeness. At home, Konglin Ke (2006) [32] introduced the rough set model which has nonparametric test characteristics and adaptability to noise data to commercial bank personal credit assessment. Konglin Ke and Zongxian Feng (2008) [33] established a enterprise loan default discriminant model of rough sets and genetic algorithms. Xiao Dongrong and Li Li (2006) [34] used the attribute association degree of measuring knowledge set with rough set method to provide reliable basis for personal credit scoring index. Ping Kang and Junyi Shen (2008) [35] proposed an evaluation method based on extended rough set theory from the basis of summarizing the traditional credit risk rating model. Through principal component analysis and attribute reduction function of rough set, Jingjing Liu and Yinhua Tian (2009) [36] established an index system combining quantitative analysis and qualitative analysis. Junhua Guo and Bangyi Li (2009) [37] proposed a model of credit risk evaluation combining by fuzzy clustering (FC) and variable precision rough set theory (VPRS). Referring to the variable precision rough set (VPRS) model, Jing Gu and Zongfang Zhou (2010) [38] proposed a new class of enterprise credit risk identification method based on VPRS. Aiming at the problem of supplier credit evaluation, Meng Wang and Yue He (2011) [39] proposed an evaluation method based on rough set and grey theory. Ting Du (2012) [40] introduced the rough set support vector machine (SVM) combing by the rough set and support vector machine (SVM) into the practice of personal credit evaluation. Jianping Wu (2016) [41] built an e-commerce credit risk classification model (RS-GA-SVM) based on rough sets (RS), genetic algorithms (GA) and support vector machines (SVM).

**Brief Comments**

Support Vector Machine (SVM) is a kind of supervised learning methods which can be widely used in statistical classification and regression analysis. And it show many unique advantages in solving small sample, nonlinear and high-dimensional pattern recognition, and it can be extended to other machine learning problems such as function fitting. The empirical research shows that the accuracy of SVM not only is better than other credit risk evaluation models and methods but also has higher prediction accuracy. SVM has very prominent advantages in theory, but compared with the theoretical study of SVM, the application research is relatively backward; the existing experimental research reports are simulation and experiments. So the application of SVM is a promising research direction.

Rough set theory is another mathematical tool for dealing with uncertainty after probability theory, fuzzy set theory and evidence theory. As a new software calculation method, more and more attention has been paid to rough sets in recent years, and its effectiveness has been confirmed in many successful applications in science and engineering. And it is one of the hot topics in the field of artificial intelligence theory and its application. Rough sets can not only deal with the incompleteness, imprecision and ambiguity of data, but also deal with data with many variables. The application of rough set method in credit risk assessment has many unique advantages, but there are also some shortcomings. For example, the index variables of the new sample may not match the decision rules extracted by the rough set, and it need to be solved by other methods at this time.

**Acknowledgement**

This work was financially supported by National Natural Science Foundation of China (71263011). And this work also was supported by Guizhou University of Finance and Economics’ Student Science Research Project—Research on the coupling and coordinated development of big data industry, ecological civilization and financial industry in Guizhou.

**References**


