Advances in Social Science, Education and Humanities Research, volume 286
3rd International Seminar on Education Innovation and Economic Management (SEIEM 2018)

Review of Domestic Application Research of Big Data Mining Technology-SVM in Credit Risk Evaluation

Mu Zhang*
School of Finance
Guizhou University of Finance and Economics, GUFE
Guiyang, China
*Corresponding author

Lu-jing Pang[1,2,3]
[1.] School of Finance
[2.] Guizhou Institution for Technology Innovation & Entrepreneurship Investment
[3].Guizhou Institute for Urban Economics and Development
Guizhou University of Finance and Economics, GUFE
Guiyang, China

***Abstract*—As a classification model in large data mining technology, support vector machine (SVM) has been developing and improving continuously, it has been applied to the field of credit risk more and more widely. The effective evaluation of credit risk by support vector machine is beneficial to the development of banks and enterprises. This paper mainly combs the domestic literature from three aspects: data preprocessing, application and improvement, and integrated combination discrimination of support vector machine in credit risk assessment. Finally, a brief review based on the domestic literature is made. Through the collation of journals reviewed, we can better understand the specific application status of support vector machine in the field of credit risk and lay the foundation for the follow-up research work.**

Keywords—big data mining technology; support vector machine; credit risk; credit risk Evaluation; Journals reviewed

I. INTRODUCTION

Credit risk has gradually become an important factor in bankruptcy. Effective evaluation of credit risk is conducive to accurate decision-making, reducing bank losses and improving the effective utilization of assets. For enterprises, the effective evaluation of credit risk can improve the availability of their loans and solve the financing problem. The evaluation and analysis of credit risk mainly follow three stages: proportional analysis, statistical analysis and artificial intelligence. Since the 1980s, we have entered the stage of artificial intelligence analysis, which overcomes the disadvantages of traditional analysis such as poor at using scientific methods in problem solving, lack of overall generalization and low accuracy of quantitative evaluation results, especially the support vector machine have stood out. Vapnik (1999)[1] first proposed the concept of support vector machine (SVM). SVM is a classifier based on the principle of structural risk minimization. It is suitable for small sample, non-linear, high-dimensional data sample analysis; it has good robustness and generalization ability. Subsequently, Chang et al. (2000)[2] applied SVM to the field of credit risk assessment and established a credit risk assessment model. SVM performs well in large data mining, lit risk assessment. This paper intends to review the domestic literature on the application of large data mining technology-support vector machine model in credit risk assessment so as to better carry out risk assessment and promote the development of SVM.

This work was financially supported by National Natural Science Foundation of China (71861003).

II. DATA PREPROCESSING OF SVM

In order to improve the discriminant ability of SVM, scholars study some problems brought by the use of SVM to classify the pre-data such as dimensionality reduction, sampling, cleaning and transformation.

A. Dimensionality Reduction

Some scholars can accurately estimate the correlation between variables by dimension reduction of data. For example, Zhao Xiao-cui and Wang Lai-sheng (2007)[3] proposed a pattern classification method based on projection pursuit (PP) and SVM, which converts high-dimensional data into low-dimensional subspace, then classifies and recognizes feature vectors in low-dimensional space using SVM. Another way of dimensionality reduction is feature extraction and selection: considering the non-linearity and high-dimensionality of financial data characteristics, Li Fei-ya and Deng Xiang (2013)[4] adopt isometric feature mapping algorithm to extract financial indicators to reduce data redundancy. Xiong Zhi-bin (2016)[5] proposed a non-linear correlation feature selection method -- based on Gebelein maximum correlation feature selection method combined with SVM technology, an evaluation model GCFS-SVM model was built. Chen Jun-fei and Zhang Qiang (2016)[6] adopted a new machine learning method, namely random forest, to select the characteristics of the rating index system. They established a credit rating model of corporate debt subject based on random forest-support vector machine. In addition, the knowledge of rough set theory is also used for data reduction. For example, Wu Jian-ping (2016)[7] established the classification model of e-commerce credit risk based on rough set, genetic algorithm and SVM.

B. Sampling

In order to overcome the influence of sample imbalance on the performance of the model, a few scholars have carried out sample pretreatment research which makes the improved SVM model more stable and predictive. For example, Meng Jie et al. (2018)[8] combined step-by-step optimal decreasing under-sampling method and boundary adaptive composite sampling method with SVM to construct ODR-BADASYN-SVM model to evaluate credit risk of SMEs. Shao Liang-shan and Zhou Yu (2018)[9] constructed a prediction model of class unbalanced credit score based on Fisher-SDSMOTE-ESBoost SVM.

C. Cleaning

In order to reduce the influence of noise, outliers and improve the classification performance of SVM, Yao Xiao and Yu An Le (2012)[10] introduced a fuzzy membership degree based on the approximate support vector machine (that is to assign a membership degree to each sample point and to reduce the impact on the model by assigning a smaller membership degree to those singular points). A fuzzy approximate support vector machine is proposed. Liu Ying et al. (2016)[11] proposed a SVM risk assessment model based on outlier elimination to solve the problem of outlier and noise in credit evaluation data.

D. Transformation

Aiming at some special sample problems, different scholars have optimized the SVM model through some specific transformations. Considering the influence degree of various indicators in the process of enterprise development, Chen Hua (2007)[12] proposed a comprehensive enterprise credit evaluation method based on K-means algorithm and SVM algorithm in data mining technology. Zhang Mu and Zhou Zong-fang (2009)[13] proposed an enterprise credit evaluation model based on multi-objective programming and SVM to solve the overlapping problem of the two types of sample enterprise credit conditions. In order to solve the problem of "sample overlap" in credit risk evaluation, Song Xiao-dong and Han Li-yan (2012)[14] established a double-membership fuzzy SVM model for credit evaluation of small and medium-sized enterprises. In addition, He Yang and Li Hong-xin (2018)[15] designed a new membership function by introducing t-test feature weighting to calculate the fuzzy membership of training points and proposed a fuzzy two-norm quadric surface SVM containing the same sample class dispersion for credit score research.

III. APPLICATION AND IMPROVEMENT OF SVM

A. Application of SVM

Domestic scholars divide the application of SVM into two major aspects: one is to apply SVM to specific problems and compare its classification effect with other algorithms; the other is to apply SVM through the selection of kernels and optimization of parameters.

1) Algorithm Comparison

Liu Yun-tao et al. (2005)[16] applied SVM to the credit risk research of commercial Banks and confirmed that this method is more effective and superior than BP neural network in the credit risk assessment of commercial Banks. Liu Min and Lin Chengde (2005)[17] established the credit risk assessment model of commercial Banks by using SVM and verified its effectiveness and superiority by comparing it with multiple discriminant analysis and neural network model. Feng Yi-ning et al. (2008)[18] believed that SVM's early warning model of group credit risk could be better applied in the early warning field of enterprise group credit risk and had better generalization ability compared with the traditional early warning model based on logistic regression algorithm. Hu Hai-qing et al. (2011)[19] established the credit risk assessment model from the perspective of supply chain finance by using SVM, through the comparative analysis of principal component analysis and Logistic evaluation model, it was proved that the prediction accuracy of the evaluation model based on SVM was better. Hu Hai-qing et al. (2012)[20] used SVM to establish credit risk assessment model and compared the empirical results with the credit risk assessment model established by BP neural network algorithm. The results show that the credit risk assessment model based on SVM under small samples is more effective and superior Kernel Function Selection and Parameter Optimization.

2) Kernel function selection and parameter optimization

In order to get a better risk measurement model, some scholars take the choice of kernel function and parameters into account and optimize the application of SVM. Hou Hui-fang and Liu Su-hua (2004)[21] introduced the linear and non-linear classification algorithm of SVM in detail. Then they applied the SVM non-linear classifier to the evaluation of bank credit risk. Finally, the experimental results of different kernel functions and parameters were analyzed and compared. Zhen Tong and Fan Yan Feng (2006)[22] applied SVM nonlinear classifier to credit risk assessment. They analyzed and compared the experimental results of selecting different kernel functions and parameters. Zeng Jiang-hong et al. (2013)[23] selected the radial basis kernel function as the kernel function of SVM and finally obtained the individual credit risk measurement model of aggregate bond financing with good classification effect through data transformation, scaling and parameter optimization Improvement of SVM.

B. Improvement of SVM

Different scholars improve SVM in different ways when dealing with different problems encountered in reality.

Some scholars choose to improve the least squares support vector machine (LS-SVM) and use it. Based on LS-SVM, Ren Ge (2013)[24] designed a credit risk assessment index system for SMEs from the perspective of supply chain finance. Xiao Bin-qing (2016)[25] established LS-SVM model based on the intrinsic characteristics of small and micro enterprises and uses the micro-data of a state-owned holding bank to prove that the model can improve the prediction accuracy and stability relatively. Li Hao et al. (2018)[26] demonstrated that the company default probability prediction model based on the LS-SVM model with indefinite kernel presented the optimal prediction performance in both the whole industry and the sub-industry, it had better robustness.

Aiming at the uncertainty of classification information of training samples, Li Yi and Xia Peng (2008)[27] combined the posterior probability of samples with traditional SVM by using Bayesian rule and obtained the SVM based on posterior probability. Shen Pei-long and Zhou Hao (2010)[28] introduced the concept of relative default distance on the basis of SVM to estimate the default probability of SMEs. On this basis, they established the mapping relationship between external rating data and default probability and obtained the default probability range corresponding to the credit level of SMEs in China. Lu Ai-guo et al. (2012)[29] proposed an improved support vector machine based on three variables and decomposed the problem into a three-variable quadratic programming subproblem, which save costs and better reflect the classification accuracy of the improved SVM. Xue Fei et al. (2013)[30] proposed a novel progressive semi-supervised support vector machine learning method based on smooth regularity and applied it to SMEs credit rating problem with insufficient label data SVM algorithm integration

IV. SVM ALGORITHM INTEGRATION

In order to solve the problem of poor generalization performance of SVM, scholars merge multiple single classifiers into an integrated model of aggregated results or combine different classifiers to classify and discriminate to form different integrated algorithm models of SVM.

A. Integration of Sub-classifiers of SVM

In order to improve the classification accuracy of the evaluation model, some scholars took a single SVM as the basic classifier and then used aggregation technology to construct the SVM integration classifier, with the new integration classifier as the credit risk assessment model. For example, Wu Chong et al. (2009)[31] established the SVM integration method based on fuzzy integration, which comprehensively considered the importance of the output of sub-support vector machines. Wu Chong and Xia Hao (2009)[32] established a SVM integration method based on five-level classification, which took into account the classification results of each sub-classifier and the importance of each sub-classifier decision to the final decision-making. A few scholars have also combined Adaboost algorithm with SVM to build an integrated classifier of SVM. For example, Li Li and Zhou Zong-fang (2013)[33] applied the algorithm to the credit risk assessment of enterprise groups of commercial Banks and concluded that the integrated classifier has a higher classification accuracy than the single SVM method. Hu Lian (2014)[34] established an AdaBoost-Integrating SVM Classifier for credit risk Evaluation of Supply Chain Finance. In addition, Chen Yun et al. (2016)[35] proposed a support vector regression integration model based on random subset to improve the decision-making ability of banks. Chen Yun et al. (2016)[36] proposed an integrated learning model of SVM based on random subset model and AdaBoost two popular strategies, which improved the classification performance.

B. Integration of SVM with Other Algorithms

Scholars evaluate credit risk by integrating SVM with other integration algorithms and make integrated forecasting. Zhang Jie and Wang Fan (2008)[37] used the mixed two-stage model of SVM-Logistic regression to evaluate the credit risk of listed companies and revised the output of SVM by logistic regression analysis, which reduced the experience risk of traditional SVM method and improved the classification accuracy. Guo Ying-jian and Wu Chong (2009)[38] established a credit risk assessment model based on BP neural network, SVM and DS evidence theory, they applied this new information fusion assessment model to the credit risk assessment of state-owned commercial banks. Zhang Qi et al. (2015)[39] introduced an early warning model based on the mixture of Logistic and SVM, which significantly improved the accuracy of credit risk prediction.

V. CONCLUSIONS AND BRIEF COMMENTS

In summary, on the basis of the early research of SVM more and more theoretical research on SVM based on big data mining technology has been carried out. It has gradually become the mainstream research to improve the generalization ability of SVM and to adapt to specific research problems since the optimization and improvement of SVM. Empirical research also shows that SVM does have more accurate classification ability and generalization ability compared with other classifiers and the pertinence of SVM is more suitable for credit data research. Although the theoretical research of SVM is increasing day by day, the empirical research is still lagging behind. In the future, the research of SVM should be expanded from two-classification problem to multi-classification problem. To solve the problem of unbalanced sample classification, we can select the appropriate kernel function and parameters more accurately and effectively. SVM not only makes it possible to distinguish default accurately, but also better deal with default probability, so it can describe the different degree of default in detail and the results obtained through SVM can realize its practical significance in practical problems. Generally speaking, with the development of modern economy SVM is constantly improving and research is increasing. There will be better prospects for development in the future. However, SVM is still in the process of development and maturity, which still needs to be developed and perfected by later scholars.

ACKNOWLEDGMENT

This work was financially supported by National Natural Science Foundation of China (71861003).

REFERENCES

[1] Vapnik V N, “The nature of statistical learning theory, ” M. New York: Spring-Verlag, 1999.

[2] Chang C C, Hue C W, Lin C J, “The Analysis of Decomposition Methods for Support Vector Machines.” J. IEEE transactions on neural networks, 2000, pp.1003-1008

[3] Zhao Xiao-cui, Wang Lai-sheng, “Pattern Classification Method Based on Projection Pursuit and Support Vector Machines, ” J. Application Research of Computers, 2007(02), pp. 86-88. In Chinese.

[4] Li Fei-ya, Deng Xiang, “The Application Analysis of SVM Model Based on Isomap in the Credit Risk Assessment of Listed Companies, ” J. Journal of Hebei University (Philosophy and Social Science), 2013, 38(01), pp. 102-107. In Chinese.

[5] Xiong Zhi-bin, “Research on feature Selection Method in Credit Evaluation, ” J. The Journal of Quantitative & Technical Economics, 2016, 33(01), pp. 142-155. In Chinese.

[6] Chen Jun-fei, Zhang Qang, “Research on Credit Rating of Enterprise Debt Subjects Based on Stochastic Forest-Support Vector Machine, ” J. Financial Theory & Practice, 2016(03), pp. 80-84. In Chinese.

[7] Wu Jian-ping, “E-business Credit Risk Classification Based on Rough Set, Genetic Algorithm and Support Vector Machine, ” J. Mathematics in Practice and Theory, 2016, 46(13), pp. 87-92. In Chinese.

[8] Meng Jie, Li Tian, Yuan Ze-ming, “Credit Risk Assessment of SMES Based on ODR-BADASYN-SVM, ” J. Jinan Finance, 2018(01), pp. 24-31. In Chinese.

[9] Shao Liang-Shan, Zhou Yu, “Application of improved oversampling algorithm in class-imbalance credit scoring, ” J/OL. Application Research of Computers, 2019, 36(06), pp.1-8 [2018-11-25]. http://58.16.80.192:9000/rwt/CNKI/http/NNYHGLUDN3WXTLUPMW4A/kcms/detail/51.1196.TP.20180408.1049.012.html. In Chinese

[10] Yao Xiao, Yu Le-an, “A fuzzy proximal support vector machine model and its application to credit risk analysis, ” J. Systems Engineering-Theory & Practice, 2012, 32(03), pp. 549-554. In Chinese.

[11] Liu Ying, Wang Li-min, Jiang Jian-hua, Zhao Cheng-li, Zhang Chi-jun, Sun Tie-zheng, “SVM Credit Risk Evaluation Method Based on Eliminating Outliers, ” J. Journal of Jilin University (Science Edition), 2016, 54(06), pp. 1395-1400. In Chinese.

[12] Chen Hua, “Research on Enterprise Credit Evaluation Based on Data Mining Technology, ” J. Science of Science and Management of S. & T., 2007(07), pp. 192-194. In Chinese.

[13] Zhang Mu, Zhou Zong-fang, “An Evaluation Model for Credit Risk of Enterprise Based on Multi-ojective Programing and support Vector Machines, ” J. Awarded Thesis Collection Chosen by the 3rd Guizhou Excellent Academic Thesis of Natural Science, 2009(04), pp. 185-190. In Chinese.

[14] Song Xiao-dong, Han Li-yan, “Credit Evaluation for Small-and-Medium-Slized Enterprises Based on fuzzy SVM with Dual Membership Values, ” J. Industrial Engineering Journal, 2012, 15(01), pp. 93-98+108. In Chinese.

[15] He yang, Li Hong-xin, “Research on Credit Scoring Based on Fuzzy Binorm Quadric Surface Support Vector Machine, ” J. Statistics & Decision, 2018, 34(05), pp. 66-70. In Chinese.

[16] Liu Yun-tao, Wu Chong, Wang Min, Qiao Mu, “Research on Credit Scoring Based on Fuzzy Binorm Quadric Surface Support Vector Machine, ” J. Forecasting, 2005(01), pp. 52-55. In Chinese.

[17] Liu Min, Lin Cheng-de, “A Model Based on Support Vector Machine for Credit Risk Assessment in Commercial Banks, ” J. Journal of Xiamen University (Natural Science), 2005(01), pp. 29-32. In Chinese.

[18] Feng Yi-ning, Shao Yuan-hai, Chen Jing, Wang Lai-sheng, Deng Nai-yang, “Research of group credit risk early-warning model based on support vector machine, ” J. Journal of China Agricultural University, 2008(02), pp. 94-98. In Chinese.

[19] Hu Hai-qing, Zhang Lang, Zhang Dao-Hong, Chen Liang, “Research on Finance Credit Risk Assessment of Supply Chain Based on SVM, ” J. soft science, 2011, 25(05), pp. 26-30+36. In Chinese.

[20] Hu Hui-qing, Zhang Lang and Zhang Dao-hong, “Research on SMEs Credit Risk Assessment from the Perspective of Supply Chain Finance—A Comparative Study on the SVM Model and BP Model, ” J. Management Review, 2012, 24(11), pp. 70-80. In Chinese.

[21] Hou Hui-fang, Liu Su-hua. “Credit Risk Assessment in Commercial Banks Based on SVM, ” J. Computer Engineering and Applications, 2004(31), pp. 176-178+192. In Chinese.

[22] Zhen Tong, Fan Yan-feng, “Research of Evaluating Credit- Risk in Enterprise Based on SVM,” J. Microelectronics and Computer, 2006(05), pp. 136-139. In Chinese.

[23] Zeng Jiang-hong, Wang Zhuang-zhi, Cui Xiao-yun, “Research on Credit Risk Assessment of SMES Assemble Bond Financier Based on Support Vector Machines, ” J. CENT. SOUTH UNIV. (SOCIAL SCIENCE), 2013, 19(02), pp. 8-11+19. In Chinese.

[24] Ren Ge, “Risk Assessment Model for SMEs in Supply Chain Finance, ” J. Statistics & Decision, 2013(17), pp. 176-179. In Chinese.

[25] Xiao Bin-qing, Bo Wei, Yao Yao, Li Xin-dan, “Small and Micro Businesses Credit Evaluation Research Based on LSSVM, ” J. Audit & Economy Research | Audit Econ Res, 2016, 31(06), pp. 102-111. In Chinese

[26] Li Hao, Liang Zhou, Huang Xun, Lin Yu, “Company default probability prediction based on indefinite kernel LS-SVM model, ” J. Finance and Accounting Monthly, 2018(14), pp. 170-176. In Chinese.

[27] Li Chong, Xia Peng, “Application of Posteriori Probability SVM in Enterprise Credit Assessment Model, ” J. Computer Simulation, 2008(05), pp. 256-258. In Chinese.

[28] Shen Pei-long Zhou Hao, “Research on the Prediction of the Small and Medium-Sized Enterprises’ Credit Risk on the Basis of SVM Theory, ” J. Studies of International Finance, 2010(08), pp. 77-85. In Chinese.

[29] Lu Ai-guo, Wang Jue, Liu Hong-wei, “An improved SVM learning algorithm and its applications to credit scorings, ” J. Systems Engineering-Theory & Practice, 2012, 32(03), pp. 515-521. In Chinese.

[30] Xue Fei, Lu li-min, “Wang Lei. Novel Smooth Regularization Based Semi-supervised SVM Approach and its Application in Credit Evaluation, ” J. Computer Science, 2013, 40(10), pp. 239-242. In Chinese.

[31] Wu Chong, Guo Ying-Jian, Xia Han, “Support Vector Machine Based on Fuzzy Integral for Credit Risk Model Evaluation of Commercial Bank, ” J. Operations Research and Management Science, 2009, 18(02), pp. 115-119. In Chinese.

[32] Wu Chong, Xia Han, “Credit Risk Assesment in Commercial Banks on Five-class Support Vector Machines Ensemble, ” J. Forecasting, 2009, 28(04), pp. 57-61. In Chinese.

[33] Li Li, Zhou Zong-fang, “Construction and Application of SVM Ensemble Classifier for Assessment of Enterprise Group′s Credit Risk, ” J. Technology Economics, 2013, 32(11), pp. 65-70. In Chinese.

[34] Hu Lian, “Evaluation of Credit Risk for Supply Chain Finance by AdaBoost-Integrating SVM Classifier,” J. Credit Reference, 2014, 32(11), pp. 19-22. In Chinese.

[35] Chen Yun, Shi Song, Pan Yan, “Hybrid ensemble approach for credit risk assessment based on SVM,” J. Computer Engineering and Applications, 2016, 52(04), pp. 115-120. In Chinese.

[36] Chen Yun, Yang Xiao-xue, Shi Song, “Enterprise credit scoring model based on SVR, ” J. Application Research of Computers, 2016, 33(11), pp. 3378-3382. In Chinese.

[37] Zhang Jie, Wang Fan, “Listed Companies′ Credit Risk Evaluation Based on Integration Model, ” J. Commercial Research, 2008(04), pp. 106-108. In Chinese.

[38] Guo Chong, Wu Ying-jian. “The Assessment Model of Credit Risk of Commercial Bank, ” J. Journal of Financial Research, 2009(01), pp. 95-106. In Chinese.

[39] Zhang Qi, Hu Lan-yi, Wang Jue, “Study on credit risk early warning based on Logit and SVM, ” J. Systems Engineering-Theory & Practice, 2015, 35(07), pp. 1784-1790. In Chinese.