Analysis and research on financial risk prevention based on artificial intelligence algorithms

Zhiming Xu*

{jimmy.zhixu@gmail.com}

Imperial College Londons

Abstract. With the advent of the era of big data, the development of big data has brought opportunities as well as challenges to various industries. Effectively utilizing and analyzing the vast and complex big data has become a focal point across industries. Support Vector Machines (SVM), as a novel technology in data mining, possess unique advantages in addressing nonlinearity and high-dimensional problems. It demonstrates high accuracy in regression prediction and is capable of handling massive amounts of data.This article primarily elucidates the risk types faced by grassroots tax bureaus within the context of big data. By integrating ESG risk management, specific application scenarios, existing issues, and their underlying causes, the present state of risk management in grassroots tax bureaus in the era of big data is analyzed, providing insights into the dynamic field of risk management and offering potential avenues for future research.

Keywords: Artificial intelligence; Risk management; Online learning algorithms; Big data sample selection

1 Introduction

Currently, data and information network security currently faces significant challenges. Despite the rapid development of big data technology, there are inherent deficiencies in infrastructure security, network and information security management, as well as in the development, deployment, analysis, and utilization of application systems [1]. The existing laws and regulations in our country have not fully played their intended role and require comprehensive integration and adjustment from multiple levels [2]. As the comprehensive governance platform continues to improve, the integration of various taxpayer information becomes crucial, and the security of tax information emerges as a critical concern in tax risk management [3]. Grassroots tax authorities face ESG risks related to fairness and social responsibility because the application of big data can result in information asymmetry and unfair treatment of certain enterprises. Therefore, when handling the collection of extensive data, tax authorities should avoid disclosing relevant information without the permission. Additionally, by considering the local conditions, they should establish and refine tax risk mechanisms, facilitating the effective implementation of ESG risk management in later stages [4]. In this context, Support Vector Machines (SVMs) can serve as a valuable tool within the new tax risk mechanism.

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China

DOI 10.4108/eai.19-5-2023.2334410

2 The current status of application of big data in tax risk management in grassroots tax bureaus

(1) Risk Identification: Constructing a Vast Data Mart to Improve Risk Identification Efficiency [5] [6].

(2) Risk Assessment: Leveraging Big Data Technology for Enhanced Risk Assessment [7].

(3) Risk Response: Establishing an Intelligent Risk Analysis and Monitoring Platform [8].

2.1 Types of tax risks in grassroots tax offices in the context of big data

(1) Tax enforcement risks

China's economy and society have experienced rapid development, leading to the advent of the big data era. As a result, taxpayers' production and operation activities have become increasingly complex, accompanied by diversified accounting methods. The scale of taxation has expanded, and the amount of tax-related information has grown [9].

(2) Tax system risk

One of the key challenges is the inadequate protection of data information within the legal framework. Presently, China's laws and regulations regarding the security of third-party information are not yet comprehensive, creating potential issues in cases of information leakage that can impact the litigation process in China [10].

(3) Tax data security risks

The security of data and information networks poses a significant concern. While big data technology has been advancing rapidly, there are still certain shortcomings in the information management system. These deficiencies span from infrastructure security to network and information security management, and even extend to the development, release, utilization, and analysis of application systems [11].

3 Statistical theory

3.1 Machine Learning Problem Description

Machine learning is an important data analysis technique in big data analytics. Its basic model can be seen in Figure 1.

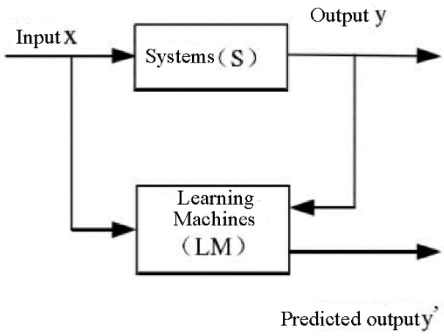

Fig. 1. Basic model of machine learning

In the machine learning system shown in Figure 1, the machine learning problem can be represented as follows: there is a certain unknown joint probability distribution function $F(x, y)$ in a set of l known samples $\{(x_1, y_1), (x_2, y_2), \cdots, (x_l, y_l)$, that are known to have the same probability distribution, and the goal of the machine learning model is to find an optimal function $\{f(x, a_0)\}$ among a set $\{f(x, a_0)\}$ of functions by estimating the output y of the system S in this function, where $\{f(x, a_0)\}$ is called the set of learning functions and can represent any set of functions with the expected risk expressed as follows:

$$R(a) = \int L(y, f(x, a)) dF(x, y) \quad (1)$$

In equation (1), a is denoted as the generalised parameter of the function. The loss function is: for $L(y, f(x, a)$ given input x, the expectation of loss between the output y of the system S and the $f(x, a_0)$ given by the learning machine LM.

3.2 Minimising the risk of experience

From equation (1), it is clear that to calculate the expected risk, one needs to know the joint probability distribution function $F(x, y)$. However, in practical machine learning problems, we do not know the joint probability distribution function $F(x, y)$ and therefore have no way of calculating the expected risk directly. In order to know the expected risk, the empirical risk is usually used instead, which is the empirical risk minimisation (ERM) criterion.

$$R_{emp}(a) = \frac{1}{l}\sum_{i=1}^{l} |f(x_i) - y_i| \quad (2)$$

Eq. (2) shows that the empirical risk is the error rate of the sample, and when the sample size l tends to infinity, the empirical risk approximates the expected risk infinitely.

3.3 The boundary between VC dimension and promotion

To understand the VC dimension theory more intuitively, the following diagram is used to illustrate. In Figure 2 you can see that there are three hyperplanes separating the seven samples, so the VC dimension of this function set is 7. The maximum number of samples that can be broken up by the set of functions in Figure 3 is 2, so its VC dimension is equal to 2.

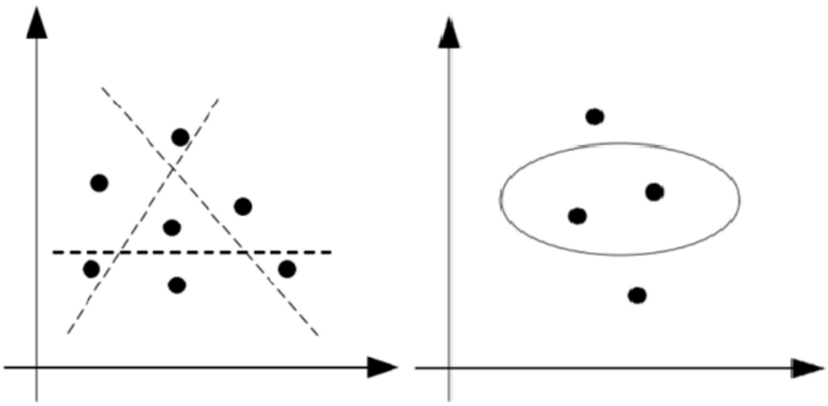

Fig. 2. Schematic representation of VC dimension equal to 7.

Fig. 3. VC dimension equals 2 schematic.

From statistical learning theory, the relationship between the expected risk $R(\alpha)$ and the empirical risk is as follows:

$$R(\alpha) \leq R_{emp}(a) + \sqrt{\frac{hln(2n/h+1) - \ln(\eta/4)}{l}} \tag{3}$$

Where: the number of samples is l and h denotes the VC dimension of the function set. From equation (3), it can be seen that the actual risk of the learning machine is related to the empirical risk $R_{emp}(a)$ and the confidence range, while the VC dimension of the function set and the number of training samples also affect the actual risk. Equation (3) can also be expressed as:

$$R(\alpha) \leq R_{emp}(a) + \Phi(\frac{h}{l}) \tag{4}$$

The boundary of extensibility can be expressed as:

$$\Delta R(\alpha) = R(\alpha) - R_{emp}(a) \leq \Phi(h/l) \tag{5}$$

The above equation reflects an upper bound on the difference between the empirical risk and the desired risk, it reflects the generalisability of the learning machine obtained according to the empirical risk minimisation criterion, hence called generalisability bound.

When the sample size is small, using the empirical risk to approximate the expected risk has a large error and poor generalisability; if the sample size is large enough, the empirical risk is closer to the expected risk.

3.4 Support vector classifier

The basic principle of SVM is to find the Optimal Separating Hyperplane based on the principle of interval maximization. The optimal classification hyperplane requires that the two classes of samples are separated without error, while maximising the classification distance between the two classes.

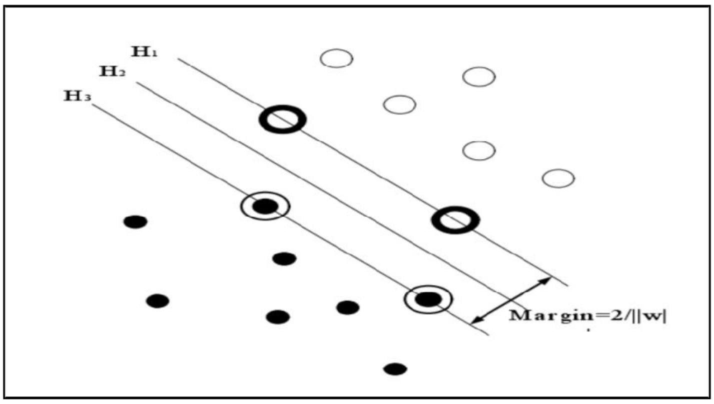

Fig. 4. Optimal classification hyperplane

In Figure 4 W is the normal vector. H_1 and H_3 are the planes over the boundary points of the two classes of samples respectively, and parallel to H_2, which is the optimal classification hyperplane to be found. The distance between H_1 and H_3 is called the interval (Margin). The classification interval is equal to$2/||w||$, and the maximum classification interval is actually minimized $||w||$.

3.5 Support vector regression machine

The loss function is a measure of the accuracy of a model's prediction, measuring the extent to which the predicted value is different from the true value, and there are differences in the support vector regression machines obtained using different loss functions.

Common loss functions in support vector machines are as follows:

(1) ε insensitive loss function

$$L_{\varepsilon}(\xi) = \begin{cases} 0, |\xi| < \varepsilon \\ |\xi| - \varepsilon, |\xi| \geq \varepsilon \end{cases} \tag{8}$$

(2) Secondary loss function

$$L_{quad}(\xi) = \frac{1}{2}\xi^2 \tag{9}$$

(3) Huber loss function

$$L_{Huber}(\xi) = \begin{cases} \frac{|\xi|^2}{2\mu}, |\xi| \leq \mu \\ |\xi| - \frac{\mu}{2}, |\xi| > \mu \end{cases} \tag{10}$$

(4) Laplace loss function

$$L_{lap}(\xi) = |\xi| \tag{11}$$

Combining vector machines (SVM) with risk management in tax administrations can provide an effective way to process and analyse large volumes of tax data and to identify and manage the various risks faced by tax administrations.

3.6 Application of vector machines (SVM) in risk management

(1) Risk classification and identification:

Using SVM algorithms, classification models can be constructed to classify and identify tax data according to different risk types. By training on known risk cases, SVM learns patterns and features of the data and classifies new tax data based on these features. This helps the tax authorities to quickly and accurately identify potential risk cases so that management measures can be taken accordingly.

(2) Risk prediction and analysis:

Support vector regression (SVR) can be applied to regression analysis of tax data for risk prediction and analysis. SVR can build regression models to predict possible future risk scenarios based on trends and correlated factors in historical tax data. This helps tax authorities to identify potential risk trends in advance and take corresponding regulatory and management measures to reduce the probability of risk occurrence.

Also, using SVM algorithms, classification models can be constructed to classify and identify tax data according to different risk types. By training on known risk cases, SVM learns patterns and features of the data and classifies new tax data based on these features. This helps the tax authorities to quickly and accurately identify potential risk cases so that management measures can be taken accordingly.

(3) Decision support and optimisation:

Using SVM's classification and regression models, decision support and optimisation recommendations can be provided to the tax authorities. Through the analysis and mining of a large amount of tax data, problems in tax administration can be identified. Based on the results of these analyses, tax authorities can formulate corresponding strategies and measures to optimise the risk management process and improve the efficiency and quality of tax collection and administration.

4 Conclusion

This article focuses on the manifestation of tax risks in the grassroots tax office, including tax data securtity risks, tax system risks and tax enforcement risks. Some theories of support vector machine (SVM) technique were also introduced, including minimising the risk of experience and the boundary between VC dimension and promotion. At present, grass-roots tax bureaus have made some achievements in using big data for tax risk management, but there are still some problems, mainly in the following aspects: lagging risk management concept, inefficient utilization of big data assets, insufficient quality of risk response, etc. Therefore, tax bureaus need to establish a sound data governance mechanism to ensure the accuracy, integrity and reliability of data. In addition, grassroots tax bureaus should actively fulfil their social responsibilities to ensure that the tax collection and administration process is fair and meets the expectations of society, thereby avoiding related financial and ESG risks.

References

[1] Lu Minfeng. Characteristics, types and management measures of digital financial risks[J]. Journal of Modern Finance, 2022(12):29-32.

[2] Lu Minfeng. Research on the characteristics and types of new financial risks and countermeasures for managing new financial risks[J]. Journal of Baoji College of Arts and Science (Social Science Edition),2022,42(06):62-70.DOI:10.13467/j.cnki.jbuss.2022.06.009.

[3] Sun Yuxi. Risk analysis and prevention strategies in the era of financial technology [J]. Tianjin Economy, 2022(10):34-38.

[4] Liang Zhan. Research on the legal system of financial supervision from the perspective of artificial intelligence [D]. Hebei University of Economics and Business, 2022. DOI:10.27106/d.cnki.ghbju.2022.000304.

[5] Zhou Daoxu, Zhang Yifei, Li Enyin. Understanding and thinking about fintech security[J]. Tsinghua Financial Review,2021(09):107-112. doi:10.19409/j.cnki.thf-review.2021.09.027.

[6] Lin Nan,Wang Yinqi. Research on the mechanism of leapfrogging the technology track of China's financial industry in the context of artificial intelligence[J]. Taxation and Economy,2021(04):80-86.

[7] Xia Shiyuan. Research on the advantages, risks and regulation of artificial intelligence in the financial market[J]. Financial Theory and Teaching, 2021(01):8-13. DOI:10.13298/j.cnki.ftat.2021.01.002.

[8] Li Yan. The use and regulation of artificial intelligence in the financial field [J]. Hebei Enterprise,2020(10):97-98. doi:10.19885/j.cnki.hbqy.2020.10.044.

[9] Chen Mancai. Building a technology innovation engine to empower business value creation[J]. China Financial Computer, 2020(08):10-13.

[10] Huang Shun. An analysis of the application of financial technology in financial risk management[J]. Mall Modernization,2020(10):94-96. doi:10.14013/j.cnki.scxdh.2020.10.041.

[11] Shen, Su-Peng, Wang, Ying-Zi, Yang, Chen. Research on the development and application of financial technology in the era of artificial intelligence[J]. National circulation economy, 2020(08):156-157. doi:10.16834/j.cnki.issn1009-5292.2020.08.071.