Research on Risk Management of Big Data and Machine Learning Insurance Based on Internet Finance

Quanpu Liu
School of Chengyi University College, Jimei University, Xiamen 361021, China
770175004@qq.com

Abstract. Insurance risk is one of the important risks faced by the insurance industry. Effective management of insurance risk is of great significance for preventing systematic risks in the insurance industry and stabilizing the stability of Chinese financial market. Risk management of insurance companies is a technical method to identify, measure and control risks. Its purpose is to directly and effectively promote the realization of organizational goals. Machine learning has a good ability to deal with non-linear classification problems. The risk assessment model based on machine learning can effectively improve the accuracy and applicability of risk assessment. The paper mainly discusses the important role and significance of big data and machine learning as an emerging data analysis method for insurance risk management, and introduces the random forest algorithm and its application in underwriting risk management.

1. Introduction
With the acceleration of international economic integration and the deepening of Chinese reform and opening up, Chinese insurance market is gradually opening up to foreign countries. Chinese insurance industry is integrating with the international market. Risk analysis and risk management of the insurance industry are becoming more and more important, and have become the core content of insurance technology management [1]. The urgency and importance of effective financial risk management are becoming more and more prominent. The corresponding theories, technologies and methods of traditional financial risk management have been difficult to deal with the huge amount of data and information generated by the development of modern financial markets. New related technologies need further research and improvement [2]. With the advent of the era of big data, emerging discipline technologies such as data mining have begun to develop, and machine learning is one of the most representative technologies. In today's business and scientific research, how to apply these emerging technologies is a puzzle that is unsolved by corporate leaders and researchers [3]. From the data itself, the characteristics of its inherent law are mined, and the new type of fraud can be quickly identified by periodically adjusting the training samples, which greatly saves the time and labor costs of the insurance company. How to effectively control and prevent the corresponding risks while the scale of the insurance company is expanding is a problem worth studying. The risks faced by insurance companies mainly include market risks, credit risks, insurance risks, policy risks, regulatory risks and so on. In view of this, the in-depth study of insurance risk is of great practical significance for promoting the integration of Chinese insurance industry with the international insurance industry, improving Chinese insurance market, developing Chinese risk management science, improving the operational efficiency of Chinese national insurance industry and enhancing its market competitiveness.

2. Necessity of Insurance Risk Management
Risk refers to the uncertainty of future losses. The most effective and basic means to manage risks is insurance. Insurance, as a basic means of spreading risks and transferring risks, has been widely used and developed rapidly, and has become an important component in the economic fields of various countries. Insurance risks mainly include the following categories: underwriting risks, management risks, investment risks and moral risks. To study the role of insurance in operational risk management and why insurance can be used to manage operational risk, we must first understand the identification and management of operational risk. However, the methods and tools of risk management have not developed significantly in recent years, and the traditional management tools are still used for qualitative analysis [4]. With the diversification of financial assets, it has been unable to meet the needs of risk management. It is easy for staff to make subjective judgments based on past experience, which will make the assessment subjective, leading to underwriting risks; as the process of international economic integration accelerates. The insurance market in various countries is also developing internationally. Some multinational insurance groups are marching into their foreign insurance market with their strong financial strength and good management technology, and have become the overlord of the international insurance industry [5]. In recent years, machine learning technology is progressing rapidly, and it is approaching the AI target in our hearts. Image detection, speech recognition, style migration, machine translation and other technologies have been applied in our real life.

Insurance risk management can promote the improvement of insurance management level. Insurance companies often provide compensation and disaster prevention services to customers for their own interests and customers' requirements. If banks do not purchase insurance, they must undertake or purchase such services from other places. They collect credit data of the banking system itself and various data of users on the Internet, including interpersonal relationships, historical consumer behavior, identity characteristics and so on. For different customers, the relevant factors and characteristics are not unified, so the evaluation results are also highly subjective. Credit scoring method is to find the main cause of default time. Insurance management is a new discipline developed from insurance and actuarial science. After nearly half a century of development, it has become an important branch in the field of management science and has been applied to various types of enterprise management [6]. After the bank purchases the insurance, no matter whether there is a risk accident or not, the bank will have sufficient internal funds to invest in new projects. Risk management of insurance companies can also speed up Chinese social insurance system.

From a historical perspective, the understanding of risk phenomena can be traced back to a very early age. At the same time, people began to adopt various methods to deal with daily life and production activities very early or intentionally. Risk. The underwriting risk mainly comes from the adverse selection and moral hazard of the insured, and the influence of the weakening of the insurable conditions under the external environment. The pricing risk is mainly due to the basic principle of insurance pricing. If the infinite information gold mine contained in the data is utilized with advanced analysis technology and converted into extremely valuable insight, it can help financial enterprises to implement real-time risk management, become a powerful protection shield for financial enterprises, and ensure the normal operation of financial enterprises. The internal risk of an insurance company can adopt different risk control methods according to the characteristics of each link. The commonly applicable method is loss prevention and loss reduction. The core content of this method is the construction of the internal control degree of the insurance company; the use of statistics, computer, data mining And artificial intelligence and other means to identify and judge potential risks in Internet finance from the ocean of data; At the same time, in order to ensure insurance operation or profit from insurance operation, insurance providers have strengthened the hazard assessment of risk phenomena, and created and applied a variety of novel disaster prevention and loss reduction technology tools.

3. Big Data and Machine Learning Assess Risks
As an emerging technology, big data has not yet been defined with special authority. Some are defined as massive data, others as full data. Through operators' communication attribute data and user behavior data, they are deeply collected, mined and analyzed to form all-round portrait data, and through machine
learning technology and big data related technology. There may be many divisions of hyperplanes. The optimal hyperplane is the plane that makes the data points belonging to two different classes have the largest interval, because the hyperplane has the best tolerance for the local perturbation of the training samples, that is, the hyperplane the classification results are the most robust and have the strongest generalization ability for the unseen examples. A currently accepted definition in the IT world is that data sets that cannot be analyzed and utilized with traditional database software tools can be called big data within an affordable time frame.

Machine learning is a statistical analysis of a large amount of Internet data. Only the larger the amount of data learned, the higher the accuracy of prediction. In the algorithm training process, the connection weight between the input layer and the hidden layer and the bias vector of the hidden layer do not need to be iteratively adjusted repeatedly, and the only optimal solution is obtained by solving a minimum norm least squares problem. At the same time, although other methods can reduce or reduce the frequency and degree of risk to some extent in practice, it is difficult to completely eliminate it. Machine learning conducts supervised or unsupervised learning analysis on massive historical data through various algorithms, summarizes the rules, and uses the analysis results to predict future data [7]. And use the model to predict a future method, which is to transform the process of human thinking and inductive experience into the process of computing the data from the computer.

3.1. Data collection
Information management should actively coordinate risk management, collect massive data, connect ports, eliminate information lag and isolation, and establish a data center suitable for big data processing. Its core idea is to establish an optimal classification line or an optimal hyperplane to correctly classify the two types of training samples. This paper uses portfolio risk analysis technology to decompose various cash flows of insurance companies, and then uses financial risk measurement technology to analyze portfolio risk investment of insurance funds. During the modeling process, a series of user credit-related features generated from user's mobile data are formed for modeling similar products. The prediction of machine learning algorithm depends on the information associated with a similar risk event. The more information, the more accurate the prediction is. Because the large data itself has the characteristics of possibly unreal and the fuzzy characteristics of machine learning algorithm itself, the prediction results are often uncertain. The larger the amount of data learned, the higher the accuracy of prediction. Insurance risk management analysis technology uses probability analysis and trend analysis to predict the possibility of future losses by collecting past loss data. Risk assessors should provide complete and accurate historical assessment information, including not only the risks that have occurred, but also the potential accidents. All information should be reported and incorporated into the risk case database.

3.2. Machine algorithm
Machine learning includes supervised learning, unsupervised learning, semi-supervised learning, intensive learning, deep learning, migration learning, integrated learning, and dimensionality learning. Sample data has requirements for both quantity and quality. Sample quality refers to the fact that the sample is a good representation of the customer base. In the case of constant mass, the number of samples is as good as possible; different algorithms have different prediction effects for different data sets, application methods and prediction targets. This requires repeated testing to achieve better application results. Different learning methods have different algorithms. Different algorithms have different prediction results for different application modes, data sets and prediction objectives. The established model and collected data are the objects of analysis, and the results of analysis may lead to changes in any link; the results may lead to further machine learning, such events are modified to be risk-free by the model, resulting in the omission of major risk events. In general, both trend analysis and probability analysis are to find the past loss patterns and project them into the future so as to predict the distribution of future losses [8]. Using advanced modern statistical techniques can make more accurate
predictions. Most algorithms perform better on training sets than on test sets, but the difference is different.

3.3. Iterative learning

Machine learning is not a once-in-a-lifetime process, but a cyclical process that self-renews and verifies and corrects existing empirical models and logic. The performance gap between the training set and the test set is small. Business verification, using indicators commonly used in business to visually illustrate the effects of the model. This makes the prediction results more uncertain. In fact, the predictions given by many machine learning algorithms are themselves highly probable. The trend analysis mainly calculates the impact of each independent trend and each independent trend on the frequency and extent of future loss, and integrates the trend effects to conduct an overall trend analysis [9]. On the one hand, too much missing information will lead to modeling difficulties; on the other hand, even if there is a reliable model, it cannot extract customer information and cannot be used. Due to its strong personal subjectivity, the evaluation results are scattered and cannot form a unified standard. Relying on comprehensive analysis of information, quantitative evaluation and rating method is not unified in different industries, related factors, characteristics, etc. Machine learning is not a one-off process, but it can continuously iterate itself and revise existing empirical models and logics. Therefore, applying machine learning to risk assessment is a continuous process of learning, summarizing and relearning, thus realizing a virtuous circle and improving the accuracy of risk identification. It is a good prediction for negative treatment, over-commitment, over-optimism, lack of control and regulatory risks in all walks of life.

4. Application of Big Data and Machine Learning in Insurance Risk Management

In fact, traditional risk management methods have been difficult to effectively analyze, prevent and supervise various risks faced by insurance companies, including insurance risks. The traditional risk management has a relatively single line of thoughts and methods. It takes disaster prevention and mitigation as the starting point and adopts traditional, local and non-systematic safety management modes. Insurance companies can analyze risks in combination with all aspects of insurance operations. Underwriting risks mainly come from the adverse selection and moral hazard of the insured, and the weakening of insurable conditions under the influence of external environment is increasingly serious. New cognitive technologies, such as big data, machine learning and natural language processing, are replacing traditional methods of analysis to quantitatively analyze and process the ever-growing data sets generated in the insurance market, ultimately helping to find various risk indicators, thus achieving more effective risk management. To control financial risks, we need to have a sound financial early warning mechanism as a guarantee. The so-called financial early warning mechanism mainly refers to various organic forms that reflect the financial risk alert, the source of the police and the trend of change, the indicator system and the early warning method. Especially the emergence of Internet technology, massive data on human consumption, entertainment, credit and other behaviors have also arisen. At this time, traditional software tools cannot acquire, manage and analyze data collections of such scale within a certain time frame. Therefore, new processing concepts and technical tools are needed. In order to ensure insurance operations or profit from insurance operations, insurance providers have strengthened the assessment of the hazard of risk phenomena and created and applied various innovative tools for disaster prevention and depletion.

The purpose of the insurance company's risk control method is to change the risk status of the insurance company, help the insurance company to avoid risk and prevent losses, and strive to reduce the negative impact of the risk on the loss in the event of loss. Machine learning is a big data-based modeling and analysis technology. It mines useful data from a large amount of data. By continuously selecting data, building model data, verifying data, and re-adjusting models, it periodically adjusts training samples. The data itself looks for intrinsic patterns and laws that ultimately produce optimal output. Most of the insurance company's risk financing decisions are the combination of the retention risk and the transfer risk to varying degrees [10]. In the design process, quantifiable indicators should
be considered as much as possible, and certain qualitative indicators should also be set to further systemically reflect financial risks that cannot be characterized by quantitative indicators. Large enterprises build their own credit rating systems through big data mining, while small and medium-sized enterprises obtain risk assessment services through information sharing and with the help of third parties. Insurance companies will choose to transfer risks. Self-retaining decision-making requires a basic trade-off between two aspects. Different learning methods correspond to different algorithms, including BP neural network, support vector machine, random forest, clustering analysis algorithm and so on. Different algorithms will give different prediction results for different application modes, data sets and prediction objectives.

The final step of risk management for insurance companies is to monitor and evaluate risk management decisions and implement complete comprehensive risk management. Taking random forest algorithm as an example, it has application value in underwriting risk management: underwriting risk is the risk brought by the extensive management of insurance companies. In order to avoid the corresponding subjective risks, random forest algorithm can be used to analyze and screen the insured information. When the output result is not equal to the original value of the data, the error is reversely transmitted: the error is reversely transmitted layer by layer according to the output layer, the hidden layer and the input layer, and at the same time the connection weight is corrected in a manner that the error gradient decreases. A self-sampling method is used to extract a training sample set from the original data set to establish a series of decision trees. When a decision tree is established for each training sample set, n explanatory variables are randomly selected as candidate variables for the segmentation nodes from the explanatory variables such as age, height, education, consumption behavior, and line record. The process of machine learning first selects the data, then builds the model data, then validates the data, tests the data again, uses the data again, and finally tunes the data. The decision tree is established with the insurance company's insurance coverage as the objective variable and its classification accuracy as the optimization objective of the prediction model. The responsibilities of the data center include: data collection, collation, processing and storage, providing convenient and reliable data manipulation interface for the use of other levels of users. Machine learning algorithm is not only more efficient in the process of analyzing and screening policyholders' information, but also greatly eliminates the judgment bias caused by human subjective experience, so it has great application value and significance.

5. Summary

Chinese financial and insurance market is changing rapidly. While the scale of insurance companies is expanding rapidly, they are also facing the construction and optimization of a comprehensive risk management system. It is the most effective way to utilize emerging science and technology and tools. Machine learning has many applications, including pattern recognition, image recognition, intelligent decision-making and so on. Scientific prediction and evaluation of risk in advance is the first essential to prevent risk. The prediction results of big data in machine learning are based on real risk sources, and are continuously optimized after information processing through neural network, training model establishment, and performance evaluation. In terms of insurance risk management, many insurance companies now use machine learning methods to mine insurance risk management rules and identify new cases. Machine learning methods are increasingly regarded as the key means of insurance risk management. This paper first proposes to apply the random forest algorithm to the management of underwriting risk. Next, we will consider continuing the specific empirical research. Other related applications are gradually beginning to attract people's attention, but still need further study.

References

[1] Choi T M, Chan H K, Yue X. Recent Development in Big Data Analytics for Business Operations and Risk Management [J]. IEEE Transactions on Cybernetics, 2016:1-12.
[2] Kobayashi K, Takara K, Sano H, et al. A high-resolution large-scale flood hazard and economic risk model for the property loss insurance in Japan [J]. Journal of Flood Risk Management,
2016, 9(2):136-153.

[3] Woodard J. Big data and Ag-Analytics: An open source, open data platform for agricultural & environmental finance, insurance, and risk [J]. Agricultural Finance Review, 2016, 76(1):15-26.

[4] Samhan, Bahae. Can cyber risk management insurance mitigate healthcare providers’ intentions to resist electronic medical records? [J]. International Journal of Healthcare Management, 2017:1-10.

[5] Cerchiello P, Giudici P. Big data analysis for financial risk management [J]. Journal of Big Data, 2016, 3(1):18.

[6] Savitt A. Insurance as a tool for hazard risk management? An evaluation of the literature [J]. Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards, 2017, 86(2):1-17.

[7] Adams M, Hoejmose S, Kastrinaki Z. Corporate Philanthropy and Risk Management: An Investigation of Reinsurance and Charitable Giving in Insurance Firms [J]. Business Ethics Quarterly, 2017, 27(01):1-37.

[8] Bernard C, Le Courtois O. Asset Risk Management of Participating Contracts [J]. Asia-Pacific Journal of Risk and Insurance, 2012, 6(2):5.

[9] Thistlethwaite J, Wood M O. Insurance and Climate Change Risk Management: Rescaling to Look Beyond the Horizon [J]. British Journal of Management, 2018, 29.

[10] Lin Y, Macminn R D, Tian R, et al. Pension Risk Management in the Enterprise Risk Management Framework[J]. Journal of Risk and Insurance, 2017, 84(S1):345-365.