Scholar Publishing Group
International Journal of Business Management and Economics and Trade
https://doi.org/10.38007/IJBMET.2024.050118
ISSN 2790-0959 Vol. 5, Issue 1: 152-168

Investor Risk Forecast and Management Path of Listed Companies Based upon Machine Learning

Zhiwen Liu[1,a*]

[1]*College of Management, National Taiwan University, Taipei 10617, Taiwan, China*
[a]*jasonliufin@hotmail.com*
[*]*Corresponding author*

Keywords: Machine Learning, Risk Prediction, Management Path, Fund Security

Abstract: Investors must reasonably predict and optimize management of the risks they face in the investment process, so as to ensure the safety of investors' funds. Risk comes from the understanding of relevant information, cognitive analysis and prediction. For investors, this is precisely what they lack. Listed companies sometimes hide related risks with their own interests, resulting in asymmetric information differences between investors and their superior companies. This information asymmetry greatly exacerbates the risks of both parties. It is not conducive to the development of the entire investment market, nor is it conducive to risk control. This article aims to study the risk sources of listed company investors and how to avoid investors' risks. This article proposes to predict the investment risks of investors and manage the risks faced by investors. With the help of machine learning model, this article minimizes the risk of investors and effectively guarantees the safety of funds. The experimental results of this paper show that risk prediction of investor funds and portfolio management investment can minimize risks and improve the security of funds for more than 20% of investors.

1. Introduction

As an indispensable part of the modern economy, the financial market plays an important role in the entire market economy system. But the characteristic of finance is that it is very risky. Once the crisis is triggered, it will have a ripple effect on financial stability and development, such as the butterfly effect. However, risk means the possibility of development. Listed companies participating in venture capital activities can create channels or channels for their abundant capital to appreciate. In the process of participating in venture capital activities by listed companies, the company's business risks can be reasonably avoided. Listed companies can play a good benchmarking role in venture capital. For investors, the risk management of listed companies can better protect the interests of investors. In the long run, risks still exist. Therefore, the key to risk prevention is to anticipate and avoid risks reasonably. Machine learning is a course that uses the powerful computing functions of computers to simulate human learning behaviors. It acquires and learns

skills from scratch, and continuously improves the accuracy of skills. At the same time, human beings are engaged in complex and redundant local subject work. Machine learning can promote the development of the financial sector. It can analyze, model and predict changes in the financial market, assess risks, manage customer relationships, make decisions that support the operation of financial companies, and better prevent financial risks.

For risk prediction management, domestic and foreign experts and scholars have carried out a lot of research. Luo believes that predictive modeling is the key to solving many risk problems. Among all predictive modeling methods, machine learning methods can usually achieve the highest predictive accuracy. But the long-standing openness problem hinders their wide application in special fields. He proposed the first complete method that can automatically interpret the results of any machine learning predictive model without reducing accuracy. For the champion machine learning model of the competition, this method explained 87.4% of the correct prediction results [1]. Agrawal S uses risk prediction in the medical field to study and evaluate the performance of a comprehensive risk assessment model for preeclampsia in predicting adverse maternal outcomes. Performing statistical analysis of ratios and ratios by evaluating χ^2-tests and odds ratios. It turns out that breathing difficulties, visual impairment, and upper abdominal pain seem to be very important risk factors. Among the biochemical variables studied, serum creatinine and serum uric acid have a significant correlation [2]. Aven T reviews the progress of the principles and methods of risk assessment and management, paying particular attention to the basic ideas based on the principles and methods of risk assessment and management, as well as theoretical platforms and practical models and procedures. He is looking for trends in viewpoints and methods, and he is also thinking about the areas of risk that need further development and should be encouraged [3]. Reim W research puts forward a product service system (PSS) risk management decision-making framework for PSS operations, which can enable global manufacturing companies to successfully provide PSS. He conducted 25 semi-structured interviews with different interviewees from different functional departments. The findings include identifying and proposing the interrelationships between operational risks related to the provision of PSS, possible risk management responses, and decision-making criteria, all of which enable decision makers to select appropriate risk management responses [4]. Calomiris C W studied bank governance and risk selection in the 1890s. During this period, there were no losses due to deposit insurance or other government assistance to banks. He links the differences in management ownership with different corporate governance policies, risks, and risk management methods. The study found that formal corporate governance is negatively related to the ownership of senior managers. When formal governance is adopted, the higher the management ownership, the greater the salary and self-loan of managers. Banks with high management ownership (low formal governance) aim for a lower risk of default. Senior management ownership and informal governance are related to relying more on cash rather than equity to limit risk [5]. Kerr J investigates the under-researched area of insurance. It uses empirical research data to focus on a case study in the art world in London to analyze how the global art insurance industry "ensures" safety and how it makes risk and safety acceptable. The article examines how the industry plays a key role in art safety and the art world itself, and believes that the role of the global art insurance industry is largely beneficial to the art world. Because insurance allows risks to be accepted, it predicts crimes and post-criminal reactions, as well as the impact of "surpassing" insurance. It has inspired and promoted the vibrant and flourishing global art world [6]. Qazi A proposes a new process that connects project complexity and risk management. It captures the interdependence between complexity drivers, risks and goals. He conducted empirical research to determine complexity/risk management practices, and the modeling method used was based on the framework of EUT and BBN. Process helps to prioritize complexity drivers, risks and strategies [7]. Kingwell R briefly discussed the nature of the price risk faced by large-scale farmers in

Australia, and outlined some of the impact of price risk on farm management. The article describes the changes in price risk over time and commodities. The potential form of price distribution faced by farmers has proven to have an important impact on farm management. He also discussed the possibility of increased price risk in the next two decades [8]. These studies have their own research value in their respective fields, but there is very little research on the risk prediction management of investors in listed companies.

Based on the risk management model of listed companies and investors, this paper proposes to build a machine learning model to better realize risk prediction. Data analysis of the reasons for the formation of risks and how to better avoid risks. Experiments have proved that machine learning can better realize the accuracy of prediction results and realize the risk control of listed companies and investors.

2. Risk Prediction Management Methods and Machine Learning

2.1 Risk Prediction Management

Investment risks are mainly manifested in the failure of venture capital companies to invest in equity plans due to factors such as poor management of the invested company and unsuccessful listing plans, resulting in the failure of venture capital companies to achieve the expected return as agreed in the investment agreement, and even withdraw the principal [9]. The risks can be divided into the following categories:

(1) Operational risk: This is an internal risk, which causes losses due to external events caused by imperfect or even failed internal processes, labor, and systems.

(2) Financial risks: These are the risks of financial failure. Possible reasons for errors in the financial management system and improper use of other financial management methods. In the process of financial management, business is inevitable, and this problem cannot be eliminated. Unless managers adopt a proactive management approach, they must reduce the loss of financial risks. But this kind of risk lurks in the company's financial management all the time.

(3) Secondary market equity risk: Secondary market capital risk refers to the price fluctuation of financial products.

(4) The risk of investing in stocks: The so-called equity investment risk refers to the economic loss of the enterprise due to the uncertainty of economic development and the uncertainty of the external environment, or the possible deviation from the expected economic return. Tangible assets and intangible assets can essentially be summarized as the instability and extreme volatility of investment returns.

(5) Policy and regulatory risks: The success of equity investment is measured by the transfer of equity. In the transaction process, it is very likely to be affected by many factors such as the current national economic environment and policies. It relies heavily on national policy projects, and when the initial policy changes, the risk will greatly increase. Nowadays, many investors focus on high-yield industries and are attracted by the high rate of return, but they often ignore the risks of policy regulation. In addition to factors such as restricted industry approvals, increased risks of policy adjustments or urban planning changes, and temporal and spatial differences, the implementation of policies will also have a greater impact on equity investment. It will further threaten the mismanagement of normal enterprises and cause excessive economic losses.

(6) Other risks: The risks encountered suddenly in the normal production and operation of the enterprise may even lead to the survival of the enterprise. Whether it is a natural disaster or an illegal or criminal act, it has been dealt with in a timely and efficient manner. If it is not handled in time, it will cause irreparable losses to the enterprise.

Risk management is the use of measurable and practical methods to identify risks in business

processes or operating units. A quantifiable formula is used to evaluate the size of the risk, and then effective economic or technical means are adopted to control the risk, and to obtain the highest possible and safe-guaranteed benefits at the smallest possible cost [10]. Risk management mainly includes four major elements: (1) Risk identification, which is mainly to monitor risks, collect information, identify and make quantitative calculations. (2) Risk assessment, the magnitude of risk is relatively speaking, and is affected by factors such as the strength of the enterprise and the risk preference of decision makers. Risk assessment is mainly to judge whether the risk identified in the previous step is within the acceptable range of the company and the degree of impact on the company. (3) The risk report is mainly to consider the company's risk management, development strategy, and business performance as a whole. It provides guidelines for the company's risk management. (4) Risk control means to implement risk reports and supervise daily business behaviors. It takes specific measures to reduce the level of risk faced by the company. The relationship between the four elements can be expressed as shown in Figure 1.

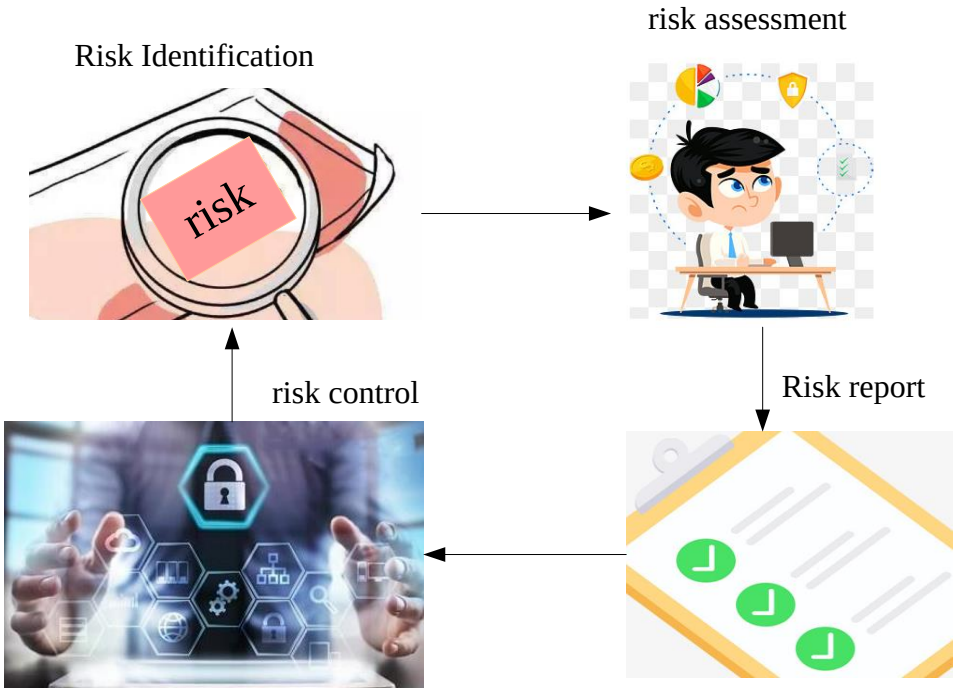

Figure 1. Risk management control elements

2.2 Machine Learning

Machine learning is a technical discipline formed by the integration and development of multiple disciplines [11]. This algorithm is an application algorithm for predicting data with the help of data laws. The specific subdivision algorithm is shown in Figure 2. According to the characteristics of manual labeling, supervised learning, semi-supervised learning, unsupervised learning, etc., they are one of the methods of machine learning. Decision trees, support vector machines, clustering, and deep learning techniques are all branches of classic machine learning. There are applications of machine learning in various fields. More mature applications include: detecting spam, detecting anomalies, segmenting users, etc. In addition, the application of machine learning technology is becoming more and more widely used in fields such as data mining and speech recognition.

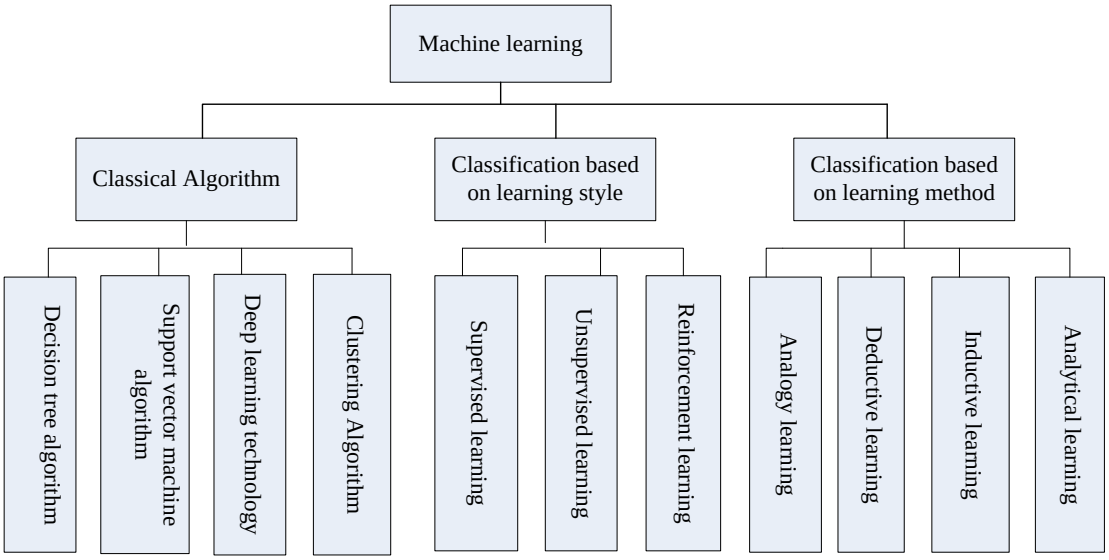

Figure 2. Introduction to machine learning classification

(1) Decision tree algorithm

A common learning method in machine learning is a decision tree. Its function is manifested in the fields of classification data, prediction data, extraction rules, etc. The structure is shown in Figure 3. Decision tree algorithm is an algorithm that subdivides data to build a decision tree in a recursive manner [12]. This article takes C4.5 algorithm as an example to analyze the principle of decision tree.

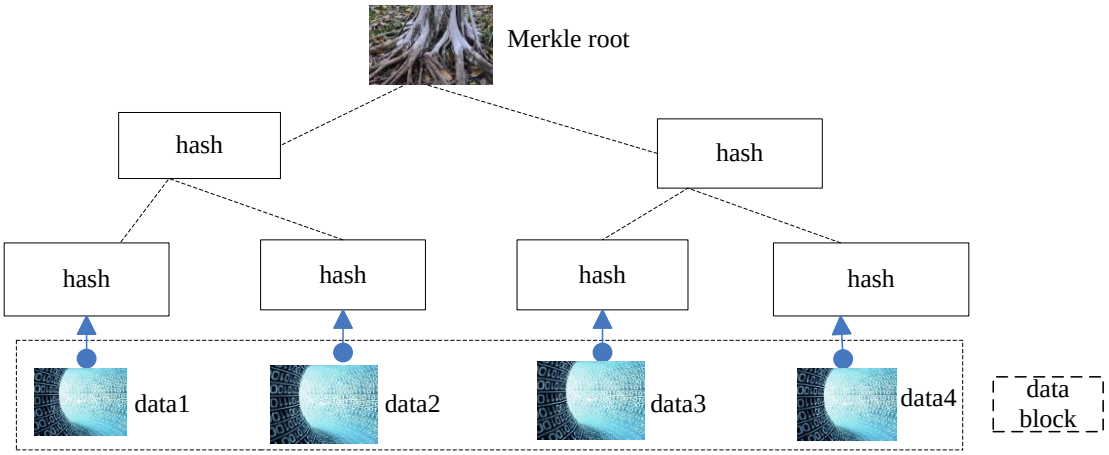

Figure 3. Merkle tree structure diagram

If the training data set is represented by A, then its information entropy can be expressed as:

$$Entropy(A) = -\sum_{x=1}^{y} c_x \log_2 c_x \quad (1)$$

Among them, the probability of category attributes with n category labels occupying the overall population can be represented by $c_x(x = 1,2,...,n)$.

Now suppose that the tuples in A are divided according to attribute B, and attribute B divides A into g different classes $\{A_1, A_2, ..., A_g\}$. Then $\{B \to A\}$ information entropy can be expressed as:

$$Entropy(A) = -\sum_{x=1}^{y} c_x \log_2 c_x \quad (2)$$

Among them, |Ai| and |A| are the number of samples contained in Ai and A, respectively.

Gain(A,B) can be expressed as:

$$Gain(A,B) = Entropy(A) - Entropy_B(A) \quad (3)$$

With the help of attribute splitting information, the C4.5 algorithm can adjust the information gain:

$$SplitE(B) = -\sum_{x=1}^{g} \frac{|A_x|}{|A|} \log_2 \frac{|A_x|}{|A|} \quad (4)$$

The information gain rate is:

$$GainRatio(B) = \frac{Gain(B)}{SplitE(B)} \quad (5)$$

(2) Support vector machine algorithm

With the deepening of the development of statistical theory, Support Vector Machine (SVM) has continued to develop in depth [13]. Its advantage is that it takes into account both minimizing empirical errors and maximizing geometric edges, which can avoid the disaster of dimensionality, and is more valuable for high-dimensional data applications. Especially nonlinear and high-dimensional binary classification and regression problems have outstanding advantages for the processing of small samples.

Support vector machine is proposed for the binary classification problem, and successfully applied the sub-solution function regression and the first-class classification problem. In statistical learning theory, we often use support vector machine classification. Because its effect is very good, its central idea is to apply the principle of structural risk minimization to the field of classification [14].

How to find a support vector from training samples to construct the best classification hyperplane is the core content of support vector machines. It is developed from the optimal classification hyperplane in the case of linear separability. To describe in mathematical language is to solve a quadratic programming problem. The constraints of this problem can be expressed in the form of inequalities. We divide the training sample set into two categories. According to whether the sample data can be divided into two cases, one is linear and the other is nonlinear. The specific structure is shown in Figure 4.

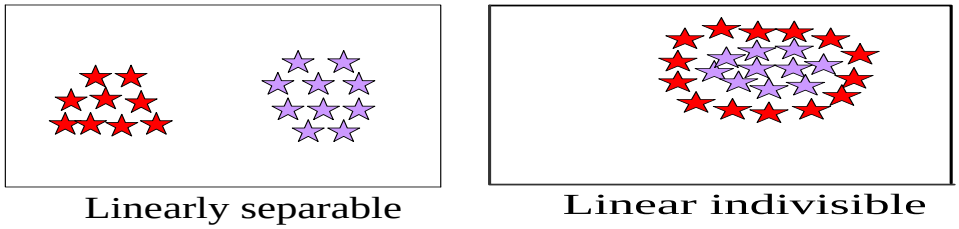

Figure 4. Case diagram of linearly separable and linearly indivisible SVM

Suppose the training sample set is $\{(z_1, z_{x1}), (z_2, z_{x2}), ..., (z_s, z_{sn})\}, z_{sn} \in \{-1,1\}$, where the amount of data is denoted by s, and the attribute is denoted by n. The training sample set F can be covered by matrix $s*(n+1)$:

$$F=\begin{Bmatrix} z_{11} & \Lambda & x_{1n} & w_1 \\ z_{21} & \Lambda & x_{2n} & w_2 \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} \\ z_{s1} & \Lambda & z_{sn} & w_{\mathrm{s}} \end{Bmatrix} \quad (6)$$

If there is a hyperplane, all vectors in the training sample set can be correctly divided, and maximum marginalization can be achieved. Then this hyperplane is called the optimal hyperplane. Assume that there is a hyperplane that can divide this sample set linearly. Suppose this hyperplane is $\mathrm{d}+T^K M=0$, and the parameter c is the intercept, $T=\{\mathrm{t}_1,t_2,...,t_n\}^K$. The support vector is the vector in the training sample set closest to the hyperplane.

First consider the linearly separable case, assuming that the support vector satisfies the condition:

$$\mathrm{d}+T^K M_x>0+c\rightarrow Z_{\mathrm{x}}=+1 \quad (7)$$

$$\mathrm{d}+T^K M_x>0\text{-}c\rightarrow Z_{\mathrm{x}}=\text{-}1 \quad (8)$$

We can get:

$$\mathrm{Z_x}(\mathrm{d}+T^K M_x)>c \quad (9)$$

The relationship between c and T satisfies the following conditions:

$$Z_{\mathrm{x}}\left(d+T^K M_x\right)\geq 1-\chi_x \quad (10)$$

$$\chi_{\mathrm{x}}\geq 0, x=1,2,...,s \quad (11)$$

The measure of the total degree of error is represented by $\sum_{x=1}^{s}\chi_x$, and the classification effect of χ_{x}=0 is 100% indicating linear separability. At this time, the distance between the two parallel boundaries is $2\mathrm{c}=\frac{2}{\|T\|}$. If c is to be the largest, then ||T|| is the smallest. The function to maximize the edge is:

$$\min f(T)=\min\frac{\|T\|^2}{2}=\min\frac{1}{2}T^K T \quad (12)$$

When $\chi_{\mathrm{x}}\neq 0$, a penalty parameter should be added to adjust the error division rate of the objective function. The objective function is as follows:

$$\min f(T)=\min(\frac{\|T\|^2}{2}+J\sum\nolimits_{\mathrm{x}=1}^{s}\chi_x) \quad (13)$$

Among them, the penalty parameter can be represented by J. Using Lagrangian multiplier method, the simultaneous solution of (10)(11)(13) is:

$$L(T,\mathrm{p},q)=\frac{\|T\|^2}{2}+J\sum\nolimits_{\mathrm{x}=1}^{s}\chi_x-\sum\nolimits_{x=1}^{s}p_x\left(Z_x\left(d+T^K\right)M_x-1\right) \quad (14)$$

The Lagrange multiplier can be represented by p_x, and $p_x\geqslant 0$. Finding the minimum value of R with respect to T and p, we can get:

$$T=\sum\nolimits_{\mathrm{x}=1}^{s}p_x Z_x M_x \quad (15)$$

$$\sum_{x=1}^{s} p_x Z_x = 0 \quad (16)$$

When the nonlinearity is separable, according to the Lagrangian function, the inner product after the conversion process determines the result when mapped to a high-dimensional space. The calculation of inner product introduces the concept of kernel function [15]. The decision function is:

$$Q(M) = \text{Sign}\left(d + T^K M\right) = \text{Sign}\left[d + \sum_{x=1}^{s} p_x Z_x \left(M^K M_x\right)\right] \quad (17)$$

(3) Logistic regression algorithm

In machine learning, a common application model is logistic regression. It is a generalized linear regression. Binary classification is the most commonly used dependent variable in logistic regression [16]. The formula for a linear regression is as follows:

$$U = \varepsilon_0 a_0 + \varepsilon_1 a_1 + ... + \varepsilon_n a_n \quad (18)$$

The input of the sample can be represented by a, and ε is the relevant parameter. Then the linear formula can be expressed as:

$$U = \psi^K Y \quad (19)$$

Use the sigmoid function to convert linear to nonlinear:

$$j(u) = \frac{1}{1+e^{-u}} = \frac{1}{1+e^{-\psi^K M}} \quad (20)$$

The graphical representation of sigmoid is shown in Figure 5:

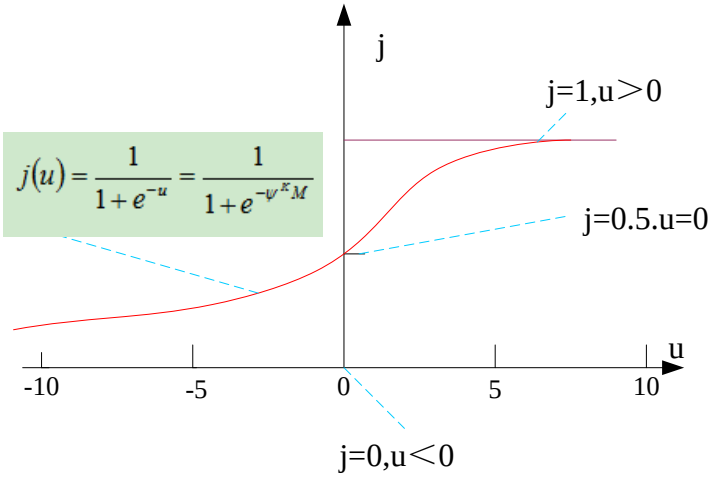

Figure 5. Graphical diagram of sigmoid

If the linear regression model is converted to logistic regression, it will cause different values on both sides of the equation. The value on the right side of the equation will be close to infinity or infinitesimal, so logistic regression calculation appears. Then get a positive sample:

$$j_\psi(t) = Q(m=1|t;\psi) \quad (21)$$

Negative sample:

$$1\text{-}\, j_\psi(t) = Q(m=0|t;\psi) \quad (22)$$

Then use the maximum likelihood method to solve the loss function, first get the probability function:

$$Q(m|t,\psi) = \left(j_\psi(t)\right)^m * \left(1 - j_\psi(t)\right)^{1-m} \quad (23)$$

The likelihood of taking the function is:

$$K(\psi)=\prod_{x=1}^{s}\left(j_{\psi}\left(t^{x}\right)\right)^{m^{x}} * \left(1-j_{\psi}\left(t^{x}\right)\right)^{1-m^{x}} \quad (24)$$

The log likelihood is:

$$k(\psi)=\log K(\psi)=\sum_{x=1}^{s}\left(m^{x}\log j_{\psi}\left(t^{x}\right)+\left(1-m^{x}\right)\log\left(1-j_{\psi}\left(t^{x}\right)\right)\right) \quad (25)$$

To maximize ψ, $h(\psi)=-\frac{1}{d}k(\psi)$, the gradient descent method can be used to solve the loss function:

$$\cos t\left(j_{\psi}(t),m\right)=\begin{cases}-\log\left(j_{\psi}(t)\right), if m=1\\-\log\left(1-j_{\psi}(t)\right), if m=0\end{cases} \quad (26)$$

$$h(\psi)=-\frac{1}{d}k(\psi)=-\frac{1}{d}\sum_{x=1}^{s}\left(m^{x}\log j_{\psi}\left(t^{x}\right)+\left(1-m^{x}\right)\log\left(1-j_{\psi}\left(t^{x}\right)\right)\right) \quad (27)$$

The goal is to find the best value of the parameter through the training sample.

2.3 Investor Behavior

The corporate governance theory in classical economics is based on the hypothesis of rational "economic man" and efficient market. But the behavioral finance theory that questioned the classic financial theory shows that the investment decisions made by investors are affected by their subjective perceptual cognition. This sensibility is affected by factors such as personality, gender, age, educational background, etc., as shown in Figure 6. The result completely violates the logical thinking and expected judgment of rational decision theory, and shows some irrational behavior [17].

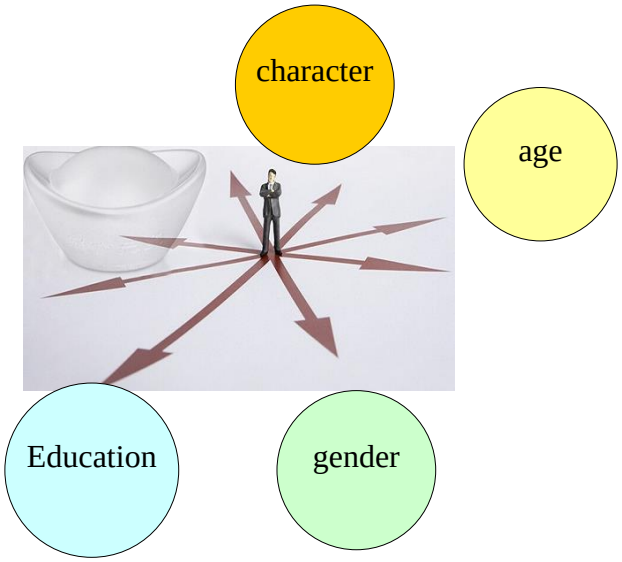

Figure 6. Influencing factors of investor irrational behavior

Investors' investment behavior rarely depends on the level of corporate governance. Retail investors still dominate the Chinese stock market. Because of the heterogeneous beliefs of investors, investors will form irrational behaviors, such as a lot of noise trading, herding effect, etc. [18]. The heterogeneous beliefs of investors are serious, so the positive signals released by listed companies in the stock market will affect the investment behavior of investors. Through the top-down

transmission mechanism, China's securities market will develop more sustainably and healthily. The investment decisions made by investors in the market are counter-promoting the governance of listed companies, and the two have a subtle mutual influence. The market must not only supervise and manage the governance level of listed companies, but also cultivate investors' good investment ideas, and actively guide investors to make correct investment decisions [19-20].

3. The Management Path of Listed Companies and the Construction of Investor Risk Experiments

3.1 Selection of Relevant Indicators

This article refers to the existing literature and selects corporate governance indicators for principal component analysis from four aspects: the distribution of equity capital, the composition of management, the incentives of management, and the relevant attributes of the company [21].

Based on the existing data, the heterogeneous investor beliefs have established relevant indicators from the four aspects of analysts' different forecasts of returns, the volatility of excess returns, the holding ratio of each member fund, and the turnover rate. This paper conducts a quantitative analysis of the heterogeneity of investor willingness [22].

The risk is related to the volatility of the expected result sample. This volatility is caused by volatility. Therefore, risk is also defined as the possibility that the actual outcome of an investment is different from the expected outcome. At present, the measurement of risk in academia mainly includes variance and standard deviation, Hurst index and lower partial moment method of return. The partial moment method under the rate of return can be divided into the variance method and the target semi-variance method [23]. Since the expected return standard deviation is the most widely used in academia, the expected return standard deviation method is chosen to measure the risk of investing in stocks. In addition, factors such as the size of the company, the degree of assets and liabilities, the life of the listed company, and the nature of the company are also important factors affecting risk [24-25].

In summary, the definition and description of each variable in this article are shown in Table 1:

Table 1. Definition and description of each variable

Variable type	Variable name	Variable code	
Dependent variable and independent variable	Stock investment risk	sir	Annual standard deviation of stock returns
	Investor heterogeneous beliefs	ihb	Annual average turnover rate
	Corporate Governance Index	gg	Make use of related variables
	Company Size	cs	Log value of annual total assets
Control variable	Company Type	ct	Value 1 for state-owned enterprises, 0 for others
	Company years	cy	Logarithmic value of listed company years
	Assets and liabilities	aal	Total assets/total responsibility
	Stock systemic risk	ssr	Company stock beta coefficient

3.2 Sample Selection

In order to eliminate the interference of other factors other than accounting information, the selection of the original sample was carried out strictly in accordance with the following conditions: First, due to the annual losses of ST and PT companies, the financial status and dividend policy will be abnormal. Therefore, in order to ensure the reliability of the research conclusions, these companies are excluded. Second, due to the valuation of this issuance and the particularity of the B and H shares, it is difficult to obtain data on such companies. Therefore, these companies are excluded. Third, since the life of listed companies is selected as the control variable when selecting variables, empirical analysis is taken into consideration to reduce the heteroscedasticity of the life data of listed companies. It can better meet the needs of the model, logarithmic transformation of the life data of listed companies, and exclude companies listed in 2015 and beyond. The fourth is to eliminate listed companies with special functions such as insurance. Fifth, to ensure the continuity of the selected data, companies with abnormal data are eliminated. After the screening, it is finally determined that the sample data is 862 samples per year, with a total of 4198 observation samples for 5 years. Principal component analysis is used to construct the corporate governance index, calculate the turnover rate and Sir coefficient data from the Guotaian database.

3.3 Data Analysis

Descriptive statistics are performed on the sample data, and the specific results are shown in Table 2.

Table 2. Variable descriptive statistics

	max	minimum	mean	Standard deviation	Median	quantity
gg	4.2544	0.1120	0.2345	1.2421	0.3245	4198
sir	1.1139	0.0088	0.4351	0.2748	0.4872	4198
ihb	3.3624	-1.0027	0.9812	1.0394	0.4113	4198
cs	36.5128	15.6751	27.4488	2.3355	25.6146	4198
ct	2.1120	-0.0048	1.2712	0.4899	1.5671	4198
sy	4.0084	1.3358	2.5103	0.4716	2.2267	4198
ssr	3.4532	0.0782	1.5657	0.7324	2.6327	4198
aal	1.2	-0.3	0.3312	0.5118	0.6	4198

Table 2 shows the basic situation of the main variables in the model. It measures the mean and standard deviation of the dependent variables, independent variables and control variables selected to reflect the overall characteristics of the sample.

From the data in the table, the data of the corporate governance index (gg) shows that the governance level of listed companies in China is generally unequal. The equity investment risk coefficient (sir) data shows that the risk of equity investment in China is not low. This may be caused by the influence of foreign economies, the fall of China's stock market, and the lack of investor confidence. The stock market plummeted suddenly, and lack of optimism about the future of the stock market led to increased risks of investing in the stock market. Investor heterogeneous beliefs (ihb) data show that investors have different views on different listed companies and make different decisions. The company size (cs) data shows that there is a big gap between different companies. Chinese enterprises are still dominated by oligopoly, and there are many small and micro enterprises, but the development of medium-sized enterprises is slightly insufficient. The asset-liability ratio (ssr) data shows that there are large differences in the asset-liability of different

companies. The median value indicates that the overall economic development of the industry is slow, the industry as a whole is in a downturn, and the corporate debt is serious. The data on lifespan (sy) of listed companies shows that the lifespans of listed companies in the sample vary greatly. The average value of ultimate controller (ct) ownership indicates that there are nearly half of the companies in the entire sample, indicating that the proportion of companies in the sample is medium.

3.4 Correlation Analysis

In order to initially understand the relationship between different variables, we conducted Pearson correlation test on the model variables [26]. The Pearson correlation coefficient test is a statistical test that describes the degree of linear correlation between variables. This is an important way to check whether there is multicollinearity between variables. The test range is [- 1.1]. Empirical analysis shows that, in general, when the value is within the interval [0,5,0,8], there is a significant correlation between the two variables. When the absolute value of the correlation coefficient between two variables is within the range of [0.8, 1.0], the correlation between the two variables is strong. This paper tests the correlation coefficients of related variables, as shown in Table 3.

Table 3. Correlation analysis of variables

	gg	sir	ihb	cs	ct	cy	aal	ssr
gg	1							
sir	-0.0657***	1						
ihb	-0.2015**	0.3126***	1					
cs	-0.0418	-0.092***	-0.273***	1				
ct	-0.057***	-0.007	-0.019**	0.518***	1			
cy	-0.4220***	0.237***	-0.109	0.219***	0.186***	1		
aal	-0.378***	-0.105***	-0.323***	0.317***	0.099***	0.178***	1	
ssr	-0.014*	0.296***	0.048***	-0.051	0.006***	0.010	0.049***	1

***(p<0.01),**(p<0.05),*(p<0.1)

3.5 Collinearity Test

The main purpose of the collinearity test is to prevent the repeatability between variables from causing errors in the regression results [27]. Therefore, in order to avoid the influence of Multi-co-linear between variables, variance inflation factor (VIF) and tolerance (tolerance) are used to diagnose Multi-co-linear problems between model variables [28]. If VIF≥10 or tolerance≤0.1, there is strong collinearity between the independent variables. The collinearity test is shown in Figure 7 and Figure 8.

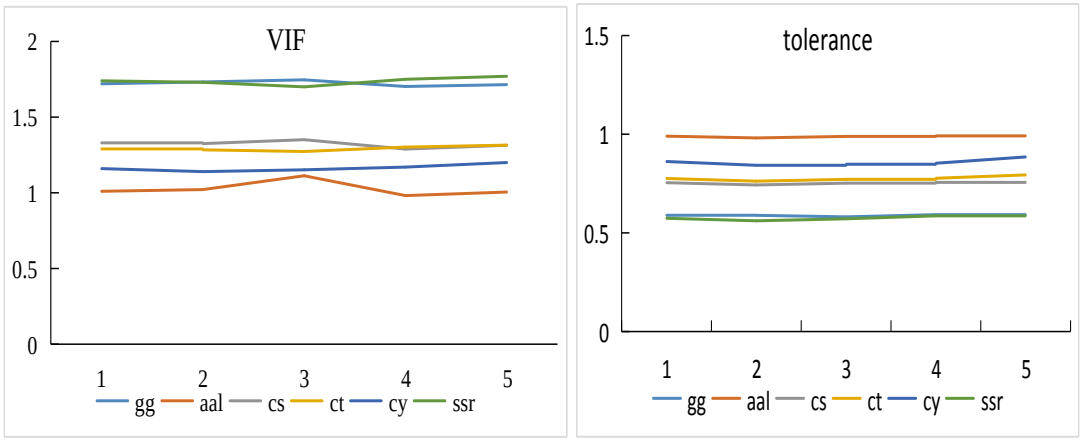

Figure7. Comparison diagram of collinearity test of various indicators in VIF and tolerance models

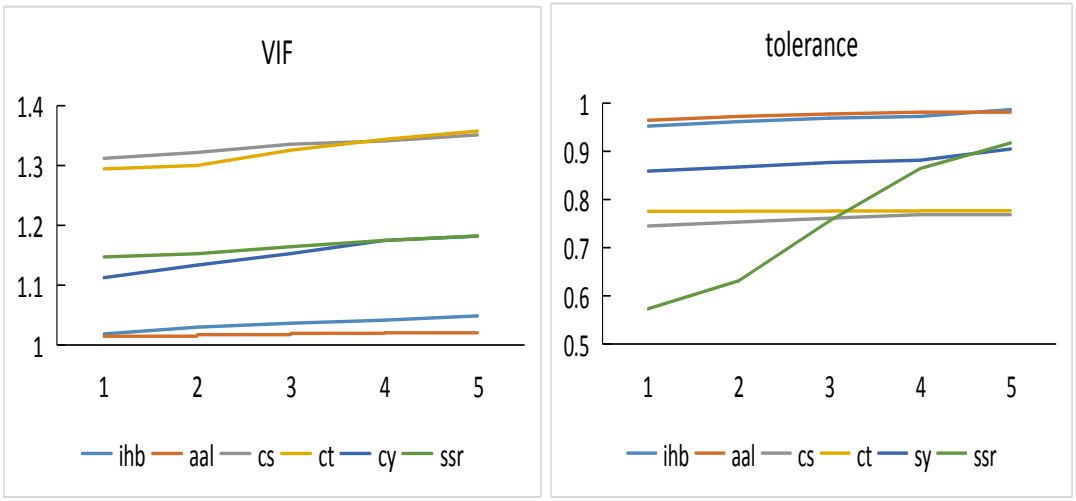

Figure 8. Comparison of collinearity test between VIF and tolerance model parameters

It can be seen from Figure 7 and Figure 8 that the tolerance of each variable in each model is greater than 0.1 and close to 1. It shows that the degree of collinearity between variables is very low. VIF is about greater than 1, far less than 10, because the higher the value, the higher the degree of collinearity. It can be seen that there is no collinearity problem between the variables that affects the regression effect of the equation.

3.6 Analysis of Regression Results

4. The regression results in Figure 9 show that the corporate governance index (gg) and equity investment risk (sir) are negatively correlated at the 0.01 level. This means that the higher the governance level of a listed company, the lower the risk of investors investing in its stocks. The corporate governance index (gg) and investor belief heterogeneity (ihd) are negatively correlated at the 0.01 level, indicating that the higher the level of corporate governance, the lower the investor belief heterogeneity.

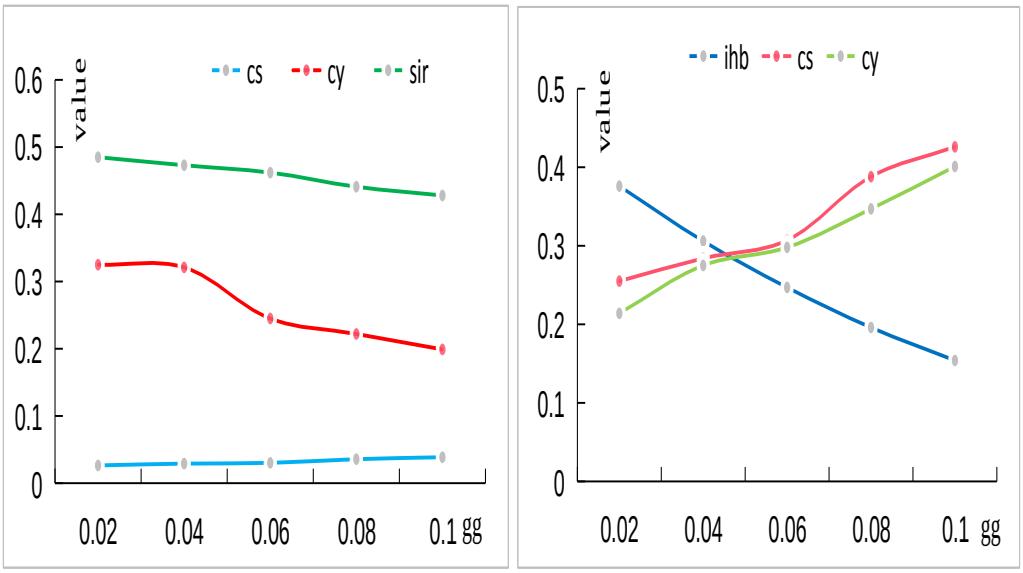

Figure 9. Correlation comparison between corporate governance index and stock investment risk and investor heterogeneous beliefs

3.7 Robustness Test

In order to verify the reliability of the research conclusions, the adjusted daily average turnover rate (Turnover) was used as a substitute for the turnover rate, and the return rate of non-dividend stocks was replaced by the return of the return rate of non-dividend stocks [29]. Secondly, the measurement results of related indicators may have a lagging effect on the market. That is, the influence of the corporate governance index on the heterogeneous beliefs of investors and the risk of investing in stocks may show up in the next fiscal year. This article will perform regression analysis on each model with lagged data [30]. If the regression results are consistent or similar, the model design and data selection are relatively stable. The other variable data used in the robustness test is consistent with the data used in the main test in this article. The specific regression results are shown in Figure 10.

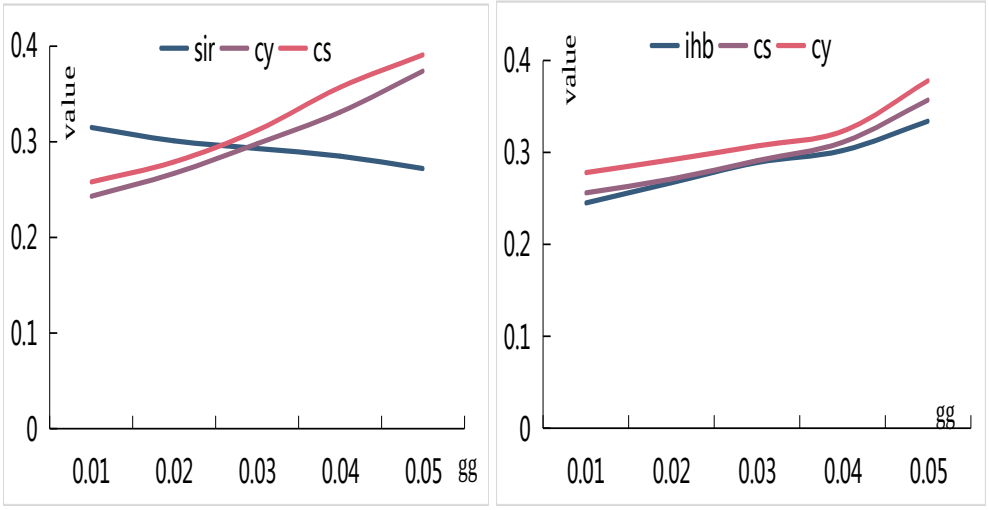

Figure 10. Comparison of robustness of corporate governance index with stock investment risk and investor heterogeneous belief data

Figure 10 shows that the corporate governance index (gg) is negatively correlated with equity investment risk (sir), and heterogeneous investor belief (ihb) is positively correlated with equity investment risk (sir). The mediating role of investor heterogeneous beliefs (Turnover) is significant. The regression results of the robustness test are consistent with the previous empirical results, which further strengthens the reliability of the results of this study.

5. Discussion

The article verifies the connection between corporate governance and investors. There is a positive correlation between corporate governance and investors' beliefs within a certain range. This requires listed companies to strengthen the management of investors. However, because investors and listed companies are not equal in receiving relevant information. Therefore, investors cannot effectively obtain the relevant information of listed companies and their specific operating conditions, so it is easy to be misled by false information to make wrong decisions. This behavior will not only damage the property of investors, but also affect the development of listed companies. Based on this, it is necessary to promote two-way communication between listed companies and investors, and theoretically attach great importance to investor management. It is necessary to do a good job in information disclosure and interpretation, respond to investor inquiries, establish a diversified and multi-form investor management model, maintain investor relations, and improve investor management supporting measures as much as possible.

5. Conclusion

Risk management is a major measure to reduce investment risks. Finding risk indicators and determining the source of risk are the prerequisites for good risk management. This article starts with listed companies and investors, investigating the relationship between corporate governance and investors' investment risks. This article finds that there is a certain relationship between them, which is conducive to better risk management. But it also has certain problems. This paper selects 6 indicators to reflect the governance level of listed companies from different aspects. However, the indicators to measure corporate governance are not rich enough, and explanations are limited. In the selected data samples of Shanghai and Shenzhen A-share listed companies from 2015 to 2020, there is no distinction between the main board, the small and medium board, and the ChiNext board. In the future, we can conduct in-depth research from specific sectors.

References

[1] Luo, Gang. Automatically explaining machine learning prediction results: a demonstration on type 2 diabetes risk prediction[J]. Health Information Science & Systems, 2016, 4(1):1-9.

[2] Agrawal S , Maitra N . Prediction of Adverse Maternal Outcomes in Preeclampsia Using a Risk Prediction Model[J]. Journal of Obstetrics & Gynecology of India, 2016, 66(S1):1-8.

[3] Aven T . Risk assessment and risk management: Review of recent advances on their foundation[J]. European Journal of Operational Research, 2016, 253(1):1-13.

[4] Reim W , Parida V , Sjodin D R . Risk management for product-service system operation[J]. International Journal of Operations & Production Management, 2016, 36(6):665-686.

[5] Calomiris C W , Carlson M . Corporate Governance and Risk Management at Unprotected Banks: National Banks in the 1890s[J]. Journal of Financial Economics, 2016, 119(3):512-532.

[6] Kerr J . The art of risk management: the crucial role of the global art insurance industry in enabling risk and security[J]. Journal of Risk Research, 2016, 19(3-4):1-16.

[7] Qazi A , Quigley J , Dickson A , et al. Project Complexity and Risk Management (ProCRiM): Towards modelling project complexity driven risk paths in construction projects[J]. International Journal of Project Management, 2016, 34(7):1183-1198.

[8]Kingwell R . Price Risk Management for Australian Broad acre Farmers: some observations[J]. Australasian Agribusiness Review, 2016, 08(4):416-20.

[9]Camiciottoli B C . Using English as a lingua franca to engage with investors: An analysis of Italian and Japanese companies' investor relations communication policies[J]. English for Specific Purposes, 2020, 58(6):90-101.

[10] Huang W . The use of management forecasts to dampen analysts' expectations by Chinese listed firms[J]. International Review of Financial Analysis, 2016, 45(may):263-272.

[11]Tugba, Karagoez. The Influence of Investor-Centered Values in the Operation of Political Risk Insurance[J]. Journal of world investment and trade, 2018, 19(1):118-153.

[12] Yan J , Tseng F M , Lu L Y Y . Developmental trajectories of new energy vehicle research in economic management: Main path analysis[J]. Technological Forecasting and Social Change, 2018, 137(DEC.):168-181.

[13] Alemu D S , D Shea. A path analysis of diagnosis of organizational levels of functionality Implications to educational organizations[J]. The International Journal of Educational Management, 2019, 33(7):1515-1525.

[14]Lowry, Paul, Benjamin, et al. Gender deception in asynchronous online communication: A path analysis[J]. Information Processing & Management: Libraries and Information Retrieval Systems and Communication Networks: An International Journal, 2017, 53(1):21-41.

[15]Mackay M M , Allen J A , Landis R S . Investigating the incremental validity of employee engagement in the prediction of employee effectiveness: A meta-analytic path analysis[J]. Human Resource Management Review, 2017, 27(1):108-120.

[16] Balbi M , Petit E J , Croci S , et al. Title: Ecological relevance of least cost path analysis: An easy implementation method for landscape urban planning[J]. Journal of Environmental Management, 2019, 244(AUG.15):61-68.

[17] Xu L , Lin T , Xu Y , et al. Path analysis of factors influencing household solid waste generation: a case study of Xiamen Island, China[J]. Journal of Material Cycles and Waste Management, 2016, 18(2):377-384.

[18] Rosa F L , Bernini F , Verona R . Ownership structure and the cost of equity in the European context: The mediating effect of earnings management[J]. Meditari Accountancy Research, 2020, 28(3):485-514.

[19]Alikaj A , Nguyen C N , Medina E . Differentiating the impact of CSR strengths and concerns on firm performance: An investigation of MNEs and US domestic firms[J]. Journal of Management Development, 2017, 36(3):401-409.

[20] Chen S , Shahi C , Chen H , et al. Trade-offs and Synergies Between Economic Gains and Plant Diversity Across a Range of Management Alternatives in Boreal Forests[J]. Ecological Economics, 2018, 151(SEP.):162-172.

[21]Helma C , Cramer T , Kramer S , et al. Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds[J]. J Chem Inf Comput, 2018, 35(4):1402-1411.

[22]Buczak A , Guven E . A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection[J]. IEEE Communications Surveys & Tutorials, 2017, 18(2):1153-1176.

[23]Singh A , Ganapathysubramanian B , Singh A K , et al. Machine Learning for High-Throughput Stress Phenotyping in Plants[J]. Trends in Plant Science, 2016, 21(2):110-124.

[24]Holzinger A . Interactive machine learning for health informatics: when do we need the human-in-the-loop?[J]. Brain Informatics, 2016, 3(2):119-131.

[25] Byrd R H , Chin G M , Neveitt W , et al. On the Use of Stochastic Hessian Information in Optimization Methods for Machine Learning[J]. Siam Journal on Optimization, 2016, 21(3):977-995.

[26]Raccuglia P , Elbert K C , Adler P , et al. Machine-learning-assisted materials discovery using failed experiments[J]. Nature, 2016, 533(7601):73-76.

[27]Obermeyer Z , Emanuel E J . Predicting the Future - Big Data, Machine Learning, and Clinical Medicine.[J]. N Engl J Med, 2016, 375(13):1216-1219.

[28] Ying S , Babu P , Palomar D P . Majorization-Minimization Algorithms in Signal Processing, Communications, and Machine Learning[J]. IEEE Transactions on Signal Processing, 2016, 65(3):794-816.

[29] Lary D J , Alavi A H , Gandomi A H , et al. Machine learning in geosciences and remote sensing[J]. Geoence Frontiers, 2016, 7(1):3-10.

[30] Wang J X , Wu J L , Xiao H . Physics-Informed Machine Learning for Predictive Turbulence Modeling: Using Data to Improve RANS Modeled Reynolds Stresses[J]. Physical Review Fluids, 2016, 2(3):1-22.