

Research Article

Corporate Risk Information Disclosure Based on Semantic Analysis Methods

Ning Yang 

College of Finance and Accounting, Yellow River Conservancy Technical Institute, Kaifeng, 475004 Henan, China

Correspondence should be addressed to Ning Yang; 2007820373@yrcti.edu.cn

Received 18 February 2022; Revised 16 March 2022; Accepted 23 March 2022; Published 12 April 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Ning Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In China's growth enterprise market, information disclosure has become the main basis for investors to make investment decisions. A sound information disclosure system is the cornerstone of maintaining the normal order of the market and a strong guarantee for the sustainable and healthy development of the New Third Board market. This study mainly discusses the enterprise risk information disclosure based on a semantic analysis method. The method of emotion dictionary construction based on dictionary has a large scale, but it is suitable for general domain emotion dictionary construction; the method of emotion dictionary construction based on corpus has strong domain. Obtain a plurality of modeling corpora to be identified; from each modeling corpus to be identified, obtain at least one emotional feature word; for any emotional feature word, obtain at least two labeling results for the emotional feature word, at least two the labeling results are labelled by the publisher and multiple receivers, and each labeling result corresponds to a classified sentiment label; according to at least two labeling results of sentiment feature words. In this paper, sentiment analysis is carried out for web financial reviews, so the domain applicability of sentiment dictionary is required to be high. The basic characteristics of the company are as follows: Indep and Concen. As an independent third party with no interest relationship with listed companies, independent directors play an important role in reducing agency costs and insider control, protecting the interests of investors, and improving the company's decision-making ability. ROA returns on total assets, which is intended to measure the ability of a company to use all its assets (including liabilities and shareholders' equity) to generate a percentage of pretax profit before interest. When calculating ROA for a certain period, the denominator (total assets) in the formula can be the average of the total assets at the beginning of the period and the total assets at the end of the period as the denominator. To be more precise, take the weighted average. To a certain extent, the concentration of equity can eliminate the defects of equity dispersion, so as to achieve the strategic purpose of enterprises. The financial characteristics of the company are measured by ROA, Leve, and OPG. Once the company's operating revenue growth rate is not ideal, it will affect the company's listing, so the management has the motivation to affect the company's relevant information disclosure. The shareholding ratio of venture capital is measured by the total proportion of all venture capital. The average value is 8.5%, the maximum value is 74%, and the minimum value is 0.04%. In the future, not only the information disclosure requirements of regulatory agencies for listing and applying for listing will be more stringent, but also the information disclosure requirements of external investors will be more stringent. This study helps to make up for the lack of gem data. Combined with the relevant experience accumulated in the actual work and the relevant regulations of the mature OTC capital market, some suggestions are put forward, in order to be beneficial to the improvement of the market information disclosure system.

1. Introduction

The performance of some companies has changed greatly, the operational and financial risks are relatively high, and the information disclosure is not standardized, which makes it difficult for investors to obtain key information in a timely

manner to make correct investment judgments, resulting in many unnecessary losses, which is not conducive to the continued health of the entrepreneurial market developing. Most of the researches on the risk factors of stock price crash are based on agency problem and information asymmetry theory, focusing on information quality, institutional

investors, company management, external factors and macromarket factors. At present, the research on the impact of corporate nonfinancial information disclosure on stock price crash has been started from the disclosure of internal control information and corporate social responsibility information, but the research on the impact of risk information disclosure in corporate annual report on stock price crash is still rare. It is found that the construction speed of the information disclosure system is far behind the market development speed, and there is currently a lack of specific regulations for detailed rules.

Comprehensive, true, and complete information disclosure can reduce the impact of adverse selection and moral hazard problems caused by information asymmetry on the market and help companies listed on the New Third Board to improve their corporate governance structures, so as to optimize resource allocation and improve financing efficiency, it is an important guarantee for the sustainable and healthy development of the New Third Board market. At present, there are relatively few studies on risk information disclosure at home and abroad, and relevant scholars have not reached a consensus on the definition of risk information disclosure. Risk information disclosure literally refers to the specific description of the main risks faced by listed companies in the nonfinancial information disclosure of annual reports, which will bring a certain degree of uncertainty to the future development of enterprises.

This paper studies the listing information disclosure mechanism of the national SME share transfer system and analyzes the possible adverse effects of listing through illegal information disclosure on business performance and investors. Deconstruction is looking for any meaning, semantics, and concepts, and then explaining how they seem to cause confusion, and is always on the boundary of meaning duality. Valibeigi and Ashuri aim to study urban identity from the perspective of deconstruction. In the first step, they proposed deconstructive philosophy based on Derrida's point of view. Then, he identified some common semantic features of the first city. Finally, he made some conclusions based on the semantic interpretation of the identity of the first city. Although it seems that all cities in his research tend to inspire a special imaginative meaning in the text, the research process is not logical enough [1]. Phu et al. believe that natural language processing has been studied for many years and has been applied to many research and commercial applications. They proposed a new model and used it for English document-level sentiment classification. In the investigation, they used the ID3 algorithm of decision tree to classify the semantics (positive, negative, and neutral) of English documents and proposed a new model. The semantic classification of his model is based on many rules. Although the accuracy of the model in his research is very high, it is not authoritative due to too few data [2]. Nguyen et al. provide a logical understanding of the development of human resource management (HRM) in Vietnam over the past 30 years. Although previous studies have reviewed the adoption of human resource management practices, it is still necessary to understand the current development of human resource management in Vietnam. Using the method of

semantic analysis, they systematically analyzed the themes and concepts of 100 journal articles about Vietnam's human resource management selected from academic databases. Although the development of human resource management in their research is related to the development of Vietnam at the critical stage of the economy, the sample age of the research is too long [3]. Augustyn et al. examined the use of nontechnical terms to refer to the SARS-Cov-2 virus, which emphasized the origin of the virus in an intentional or unintentional way. They aim to study the dynamics of meaning interpretation of selected expressions in Internet discourse. Although these potentials in his research will be activated differently according to the context in the process of meaning interpretation, the results of the research are not always so intuitive [4]. Information is the basic element of the capital market, which reflects the financial status, operation status, and external environment of listed or listed companies and is an important basis for the public to make investment decisions. Improving the information disclosure system of the capital market, enhancing the transparency of the market, and improving the operating efficiency of the capital market are the urgent needs to protect the vital interests of market participants and promote corporate governance.

All investors should have the right to receive timely and equal access to material information about the company's business and finances. A strong and effective information disclosure system is a prerequisite for the company's shareholders to effectively exercise their shareholder rights. The emotional dictionary based on the dictionary construction method is large in scale but suitable for the construction of the emotional dictionary in the general field; the emotional dictionary based on the corpus method is strong in domain. This article conducts sentiment analysis for Web financial reviews, so the domain applicability of the sentiment dictionary is highly required, and it is obtained by means of expansion based on expert annotation and manual annotation. The measurement indicators of the basic characteristics of the company are as follows: the proportion of independent directors (Indep) and the degree of concentration of equity (Concen). Independent directors, as an independent third party with no interest relationship with the listed company, are crucial in reducing agency costs and insider control, protecting the interests of investors, and improving the company's decision-making ability. A certain degree of concentration of equity can eliminate equity dispersion to a certain extent. The shortcomings are used for the realization of corporate strategic goals. The measurement indicators of the company's financial characteristics include return on assets (ROA), debt to assets ratio (Leve), and year-on-year growth rate of operating income (OPG).

2. Corporate Risk Information Disclosure

2.1. Information Disclosure. Based on the particularity of its disclosure of "risk," whether the disclosure of risk information in the annual report of a company has increased the risk perception of investors is the focus of current debate. There are currently three opposing views:

useless view, heterogeneous view, and same view. In the quality view [5, 6], the useless view believes that the company's annual report does not materially disclose risk-related information, but in order to cope with the template disclosure of relevant regulatory authorities, risk information disclosure without information content does not play its real role and cannot increase investors' risk perception: The heterogeneous view believes that the disclosure of risk information is more of unknown information in the market. These unknown risk factors and risk items increase the uncertainty of the company's future development, increase people's risk perception, and lead to stock price fluctuations [7].

In the measurement method of risk information disclosure, foreign scholars generally use content analysis and index construction methods [8, 9]. Among them, the method of content analysis is to judge whether risk information is included in the semantics by screening whether risk words such as uncertainty and risk appear in the article, using the number of risk words and the number of entries containing risk word sentences as a measure of the degree of risk information disclosure [10, 11]. With the continuous advancement of research content and methods, the index construction method can more comprehensively and three-dimensionally measure the quality of risk information content disclosure from multiple dimensions and perspectives, but due to the lack of a unified measurement standard, it is usually the academic self-built index [12, 13].

Risk information disclosure provides information increment. We all know that stock price fluctuations are caused by unexpected news in the market. Through the disclosure of external risk information, companies can convey to investors' more unknown factors and risk events that the company may face in the future development, increase the amount of information supplied, and enable investors to have reasonable psychological expectations for the future development of the company. When these risk events and unfavorable situations actually occur, they belong to the expected news, and within the acceptable range of investors, they will have less impact on the market and will not cause substantial fluctuations in stock prices [14]. Risk information is the company's forecast information for future development, which can provide a more reliable reference for investors' decision-making, prompt investors to reasonably estimate the value of the company, correctly guide investment decision-making behavior, and reduce blindly following trends in the capital market, chasing rises and falls (irrational behavior [15, 16]).

2.2. Stock Price Decline Risk. Risk information disclosure enhances investor confidence [17]. Information plays an important role in the capital market. The side with information advantages can make more accurate analysis and judgments on the business development of the enterprise, while the side with information disadvantages faces a complex investment environment due to the lack of reliable reference basis more likely to produce cognitive bias [18, 19]. Faced with the complexity and variability of the investment environment, investors will choose to use procedural rationality to control risks when they cannot accurately predict the final

investment results. For those companies with high-quality information disclosure, investors believe that they have strong procedure evaluation standards and risk response capabilities and open and transparent risk information disclosure effectively participated in market supervision, so the disclosure of risk information did not increase investors' risk perception, but increased investors' trust in the company. The confidence of the company's future development will bring positive market returns, thereby maintaining the stability of individual stocks and reducing the risk of stock price collapse [20, 21]. First of all, it is necessary to calculate the specific yield index of individual stock i in t week:

$$R_{i,t} = \beta_0 + \beta_1 R_{m,t-2} + \beta_2 R_{m,t-1} + \beta_3 R_{m,t-1} + \beta_4 R_{m,t+2} + \beta_5 R_{m,t+1} + \varepsilon_{i,t},$$

$$W_{i,t} = \ln(1 + \varepsilon_{i,t}). \quad (1)$$

Second, calculate the stock price risk based on the weekly rate of return:

$$NCSKEW_{i,t} = - \frac{[n(n-1)^{3/2} \sum W_{i,t}^3]}{[(n-1)(n-2)(\sum W_{i,t}^2)^{3/2}]}. \quad (2)$$

Among them, n is the number of trading weeks per year for a stock i , and the negative return skewness coefficient $NCSKEW_{i,t}$. The larger the value, the more serious the skewness of company i 's stock skewness coefficient, and the greater the possibility of a crash [3, 22].

$$DUVOL_{i,t} = \frac{\log \left[\frac{(n_u - 1) \sum_{\text{Down}} W_{i,t}^2}{(n_d - 1) (\sum_{\text{Up}} W_{i,t}^2)^{3/2}} \right]}{1}. \quad (3)$$

n represents the number of weeks $W_{i,t} = \ln(1 + \varepsilon_{i,t})$ that a stock i is listed and traded each year. Ideally, the fluctuation of the W value is unbiased, but in reality, the company's management will conceal related negative news out of self-interested behavior. Before the release, the probability of W rising and volatility is relatively large, and after the outbreak of negative news, the magnitude of the downward fluctuation of W will be greater than the magnitude of its upward fluctuation [23].

2.3. Semantic Analysis. Although information disclosure can also affect supply and demand while guiding investment behavior, leading to short-term stock price fluctuations, in the long run, the continuous penetration of information in the capital market will narrow the information gap between market participants. The price will gradually converge to its true value, so stock prices tend to be more stable [24, 25].

The prior information is updated through historical data, and the posterior probability density function $f(x|\text{data})$ is obtained. Its expression is as follows:

$$f(x|\text{data}) = \frac{g(x|\text{data})f(x)}{\int g(x|\text{data})f(x)dx} \propto g(x|\text{data})f(x). \quad (4)$$

The probability density function of beta(a, b) prior distribution is as follows:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \propto x^{a-1}(1-x)^{b-1}, \quad a > 0, b > 0. \quad (5)$$

The expectation and variance of beta(a, b) are as follows:

$$E(x) = \frac{a}{a+b}, \quad (6)$$

$$V(x) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Use the sample reflected by the probability density function to estimate the expected value of the distribution:

$$E(x) = \lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N x_i f(x_i)}{\sum_{i=1}^N f(x_i)}. \quad (7)$$

Among them, N is the number of iterations [26]. Define the texture gradient as the distance between these two histograms:

$$\chi^2(g, h) = \frac{1}{2} \sum \frac{(g_i - h_i)^2}{g_i + h_i}. \quad (8)$$

In order to ensure stability, the improved characteristics are defined as

$$\tilde{f}(x) = f(x) \cdot \left(\frac{-f''(x)}{|f'(x) + \lambda|} \right). \quad (9)$$

Among them, λ is the parameter used to optimize features [27]. The semantic framework combines the prior probability and the posterior probability; that is, the observation likelihood probability obtains the final pixel-level significance detection result:

$$f(\text{sal}|I) = \frac{f(\text{sal})f(I|\text{sal})}{f(\text{sal})f(I|\text{sal}) + f(b)f(I|b)}, \quad (10)$$

$$f(b) = 1 - f(\text{sal}).$$

2.4. Semantic Text Classification Model. In the input layer, the text is converted into a two-dimensional matrix through word2vec, which is similar to the process of the convolutional neural network model applied to the image, but the image can use the three RGB channels as different channels for input. In the input matrix of TextCNN, one row represents the word2vec word vector of a word in the input, and one column represents the word sequence

of the input. The size of the input matrix is the dimension vec_dim of the word vector multiplied by the sequence length seq_len. Here, the specified word vector dimension is 64 dimensions, and the length of the sequence is 650. If the length is insufficient, the sequence is filled with padding.

Convolutional layers convolve the input to extract features. kernel_size represents the number of input words contained in the vertical sliding window of the convolution kernel. If kernel_size is specified as 5, it is analogous to the n -gram model, which is equivalent to extracting the word features of the context 5gram. After each convolution kernel is calculated, we get a column vector, which represents the features extracted from the sentence by the current convolution kernel. Different from the convolutional neural network model used in image processing, the convolution kernel of the TextCNN model only slides from top to bottom in a single dimension.

Pooling layer, Max pooling layer, or mean pooling layer is often used to select the most important or representative features. The Max pooling operation is to extract the maximum value after convolution, and then combine these maximum values to obtain the final feature. In this way, after the Max pooling operation, we will obtain a vector with the same dimension as the number of convolution kernels num_filters, which is to connect the maximum value in the vector obtained after each convolution operation. If there is no process of complementing padding before the convolution operation, then, the length of the vector obtained after convolution will be different, and after the feature is merged through the Max pooling operation, the final feature vector will be the same as the number of convolution kernels, so the Max pooling operation can also solve the difference in the length of the input text The problem [28].

The external corpus required to train the classification model comes from THUCNews, which is a corpus for classifying and labeling news text. The specific classification categories are planned for the following ten categories: finance, real estate, home furnishing, education, technology, fashion, life, games, and entertainment. For each category, 7000 pieces of marked data are taken. The sizes of the training set, validation set, and test set are 5000, 1000, and 1000, respectively. The text classification model is shown in Figure 1.

Solve the model by constructing a likelihood function, and generate the following likelihood function based on the above example. The solution process of the likelihood function generally uses the gradient descent method, and the semantic feature model is obtained after the training is completed.

$$L(\beta) = \sum_{j=0}^{k-1} \sum_{i=1}^n y_{ij} \log(\pi_j(X_i)). \quad (11)$$

Train through backpropagation and labeled training data to obtain entity text classification model. The entity predicts the link result according to the result of the

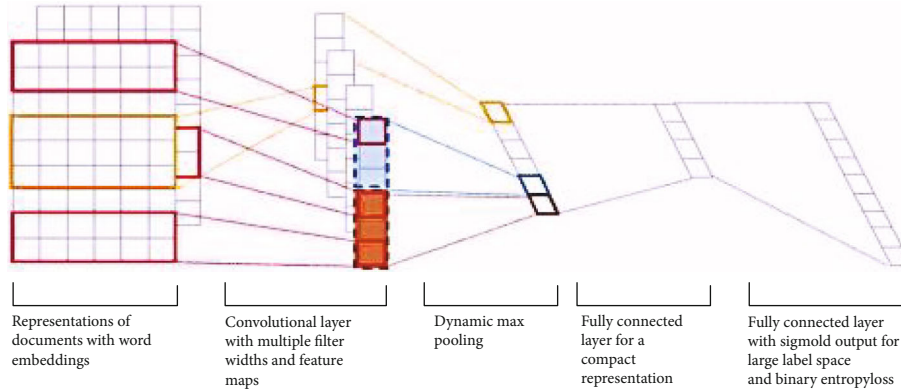


FIGURE 1: Text classification model (from <http://alturl.com/z6h5j>).

classification and the Euclidean distance result of the category attribute of the entity in the knowledge base; the formula is as follows:

$$d = \sqrt{\sum_{k=1}^n (x_{1k} - x_{2k})^2}. \quad (12)$$

3. Enterprise Risk Information Disclosure Control Experiment

3.1. Generation of Emotional Dictionary in Financial Domain. The emotional dictionary based on the dictionary construction method is large in scale but suitable for the construction of the emotional dictionary in the general field; the emotional dictionary based on the corpus method is strong in domain. This article conducts sentiment analysis for Web financial reviews, so the domain applicability of sentiment dictionary is high. And because the research focus of this article is not on the construction of sentiment dictionary, this article adopts expert annotation and manual annotation based on the expansion method. The basic emotional data information of the enterprise is shown in Table 1.

3.1.1. Experts Mark Seed Emotion Words. After comparison, the “Xinhua 08 Chinese-English Financial Dictionary” is selected as the expert annotated dictionary. The “Xinhua 08 Chinese-English Financial Dictionary” contains a total of more than 2,400 entries, involving macroeconomics, fixed income, stock funds, foreign exchange currencies, commodity futures, and other financial fields. Several experts from CITIC Securities marked the dictionary with word segmentation, part of speech, and emotional polarity and obtained 197 emotional words with praise and derogation polarity.

3.1.2. Manual Annotation on the Corpus. According to the corpus documents of the company research on Sina Finance, 383 emotional words are collected.

3.2. Selection of Financial Indicators. GEM-listed companies have disclosed a series of financial indicators that indicate the company’s profitability, solvency, and operating capabilities in their prospectus and annual reports. The selection principles are as follows: (1) Select comparable indicators,

and the increase or decrease of such indicators can be used. Compare and analyze the company vertically or horizontally. (2) Select the indicators disclosed in the prospectus and annual reports. (3) Select financial indicators from the perspective of the risk information needs of investors and regulators.

Here, it is easy to understand that the risk information disclosed by listed companies affects the changes of the company’s financial indicators by influencing investors’ decisions. We can analyze whether there is a difference between the disclosed risk information and the changes in the company’s financial indicators. Relevant relationships are used to test whether the risk factors disclosed by the issuer are true, accurate, and complete. Therefore, we mainly select the following indicators from the indicators that reflect the company’s profitability, solvency, operating efficiency, and asset management capabilities: main business profit margin, return on net assets, earnings per share, debt ratio, total asset turnover, inventory turnover ratio, and operating cash flow per share.

3.3. Theoretical Model Design of the Study. The research model of this article is divided into two aspects. The first aspect mainly analyzes whether there is venture capital investment and how the four different characteristics of venture capital affect the quality of information disclosure of GEM and listed companies. First, add the five variables of venture capital: whether there are venture capital holding VC, venture capital state-owned background VC-Country, venture capital private background VC-Private, venture capital holding ratio VC-Share, and venture capital Jointly held VC-Count, to independently return the information disclosure quality of the explained variable GEM-listed companies. The second aspect is to analyze the moderating effect of company size on the relationship between venture capital and its characteristics and the information disclosure quality rank of GEM-listed companies. By observing and judging whether the regression coefficient of the interaction term is significant, analyze the adjustment effect of the company size.

3.4. Control of Variables. The basic characteristics of the company are measured by the following:

TABLE 1: Basic emotional data information of the enterprise.

The name of the stock	Stock code	Number of emotional sentences	Number of emotional words
Pathfinder	3 00005	830	1 795
Beilu Pharmaceutical	300016	711	1430
SuperMap Software	300036	597	1092
Interactive entertainment	300043	696	1357
Bishuiyuan	300070	667	1279
Digital video	300079	689	1254
Changxin Technology	300088	745	1444
Yibai Pharmaceutical	600594	706	1459
Xin'an shares	600396	673	1447
Tsing Tao beer	600600	963	1952
Total	—	—	14509

- (1) Proportion of independent directors (Indep): the ratio of independent directors to the total number of board members in the year that the GEM company was listed. The independent director system is a very important part of the modern corporate governance design. As an independent third party with no interest relationship with listed companies, independent directors have the most important role in reducing agency costs and insider control, protecting the interests of investors, and improving the company's decision-making ability.
- (2) Concentration of equity (Concen): the degree of equity concentration is measured by the proportion of shares held by the largest shareholder at the end of the year when a GEM company is listed. The concentration of equity to a certain extent can eliminate the defects of equity dispersion to a certain extent, which can be used for the realization of corporate strategic goals. Therefore, this article predicts that the degree of equity concentration of listed companies on GEM is positively related to the quality of information disclosure.

The measurement indicators of the company's financial characteristics are as follows:

- (1) Return on assets (ROA): it is obtained by dividing net profit by the total assets at the end of the year. It mainly measures the profitability of listed companies; on the other hand, if the profitability of the company is poor, the management has a strong incentive to cover up the operation the fact of loss and decline in performance. Therefore, this article predicts that the rate of return on assets is positively

correlated with the quality of information disclosure of listed companies.

- (2) Asset-liability ratio (Leve): this is the impact of asset loading rate on the quality of information disclosure of listed companies. Solvency indicates the financial risk faced by the company. The higher the asset loading rate, the higher the company's financial risk. When the company's asset-liability ratio is higher than a certain level, the company's creditors will take certain measures to protect its interests, such as adding some restrictive clauses when signing a loan contract with the company to prevent the company from using its loans for those risks' larger projects and ultimately lost money. In addition, the company's management is often willing to increase the company's information disclosure costs to strengthen the creditors' trust in the company, thereby reducing the company's borrowing costs. Therefore, this article predicts that the asset-liability ratio is negatively correlated with the quality of information disclosure of listed companies.
- (3) Year-on-year growth rate of operating income (OPG): this mainly measures the growth capacity of listed companies. Once the company's operating income growth rate is not satisfactory, it will affect the company's listing. Therefore, the management has an incentive to affect the company's relevant information disclosure. Therefore, this article predicts that the quality of information disclosure of GEM-listed companies is positively correlated with the year-on-year growth rate of their operating income.

3.5. *Adjusted Variables.* In the company size (size), take the logarithm of the total assets of the listed company at the end of the year when it was listed, and this indicator is used as an adjustment variable. The larger the company, the greater the need and reliance on external funds, and the greater the economic and political costs caused by improper information disclosure. Therefore, larger companies have the motivation and ability to improve the quality of information disclosure. In addition, as a public company, the higher the attention and media exposure of large-scale companies, the relatively low cost of publishing information for large-scale companies and the relatively high quality of information disclosure; therefore, this article predicts that the size of the company has a positive regulatory effect on venture capital and its characteristics and the quality of information disclosure of listed companies on the GEM.

4. Corporate Risk Information Disclosure Control

4.1. *Data Analysis of Information Disclosure Quality of Listed Companies.* The data on the quality of information disclosure of listed companies in this article comes from the official website of the stock exchange, the relevant information on venture capital comes from the Wind database, and the financial data of listed companies and basic company information comes from the Guotai'an CSMAR database. The

TABLE 2: The total number of virtual finger evaluation objects and the number of correct replacements.

Company listing year	Number of IPO companies	Number of companies participating in venture capital	Venture capital as a percentage of IPO	Number of risk-free investment companies	Risk-free investment as a percentage of IPO
2018	117	68	58.1%	49	41.9%
2019	128	89	69.5%	39	30.5%
2020	74	60	81.1%	14	18.9%

TABLE 3: Overall description of the sample.

Enterprise	Frequency		Correct rate (%)
	Total number of discoveries	Correct replacement times	
Manufacturing	3424	2585	75.5%
Financial industry	4600	3824	83.1%

total number of false finger evaluation objects found in the two industries and the number of correct replacements are counted, as shown in Table 2. The overall description of the sample is shown in Table 3. Figure 2 shows the comparison between the number of companies participating in venture capital and the number of nonrisk investment companies.

4.2. Descriptive Statistics of Variables. After finishing the sample data sorting, we performed descriptive statistics on the variables involved in the model, including four types of variables: explained variables, explanatory variables, controlled variables, and adjusted variables. In addition, the descriptive statistics of the variables are shown in Table 4, so that the overall situation of each variable can be found more intuitively. The distribution of evaluation results is shown in Figure 3. From Table 2, we can see intuitively that the average information disclosure rating of 453 GEM-listed companies is 2.359, and the median is 3. There are 374 GEM-listed companies with a rating of 3, representing “good (B),” accounting for 82.56% of the total number of samples; 30 GEM-listed companies with a rating of 4, representing “excellent (A);” and 48 GEM-listed company with a rating of 2, which means “qualified (C);” only one GEM-listed company (Zhendong Pharmaceutical 300158) is rated 1, which means “unqualified (D).” This shows that the overall quality of information disclosure of listed companies on the Growth Enterprise Market is good, with a passing rate of 99.78%.

The number of venture capital is shown in Figure 4. The total proportion of venture capital holdings is shown in Figure 5. As can be seen from Figures 4 and 5, among the top ten shareholders of 307 GEM-listed companies with venture capital participation, each has an average of 2.391 venture capital, the maximum value is 7, and the minimum and mode are both 1. Among them, 95 companies have only one venture capital among the top ten shareholders, and 212 companies have more than one venture capital among the top ten shareholders, accounting for 69.05% of all GEM-listed companies with venture capital participation. This shows that joint venture capital investment is common in

GEM-listed companies. In addition, the proportion of venture capital holdings is measured by the total proportion of all venture capital. The average value is 8.5%, the maximum value is 74%, the minimum value is 0.04%, and the standard deviation is 0.093. It can be seen from Figure 4 that in the first year of listing for most startup listed companies, the total shareholding ratio of venture capital is concentrated in the range of 0-10%.

4.3. Empirical Test Results. In order to study the impact of annual report risk information disclosure on the risk of stock price collapse of listed companies, this article will finally obtain the industry distribution of sample companies and the industry distribution of all A-share-listed companies. After screening, the industry distribution of the sample data finally obtained and the industry distribution of all A-share-listed companies disclosed by the China Securities Regulatory Commission in 2018 can be seen from the above chart; in the final sample selected, the manufacturing industry accounted for the main Status, more than 60%, the three industries of accommodation and catering, scientific research and technical services, health, and social work accounted for less than 0.8%. Similarly, we have also observed that all A-share-listed companies also present a similar distribution pattern: the manufacturing industry accounts for close to 60%, which makes it occupy the main position, while the accommodation and catering industry, scientific research and technical service industry, the three industries, health, and social work also accounted for less than 0.8%. In summary, the final sample industry distribution is similar to the overall industry distribution of capital market, which shows that after removing the sample, it will not bring industry bias to the final regression results. The distribution of sample industries is shown in Figure 6.

4.4. Descriptive Statistical Analysis. In order to observe the characteristics and differences of the various indicators in the sample data, this paper carried out descriptive statistics on the main variables, including the average, standard deviation, median, minimum, and maximum. The results of descriptive statistics are shown in Table 5. From the perspective of the full sample, the average values of the two indicators of average stock price collapse risk, NCSKE and DUVOL, are -0.35273 and -0.15695, respectively, indicating that stock prices exhibit greater volatility during the sample period. The standard deviations of NCSKE and DUVOL are 0.859608 and 0.737948, respectively. The values are relatively large, which indicates that these two indicators are quite different in the sample listed companies and further

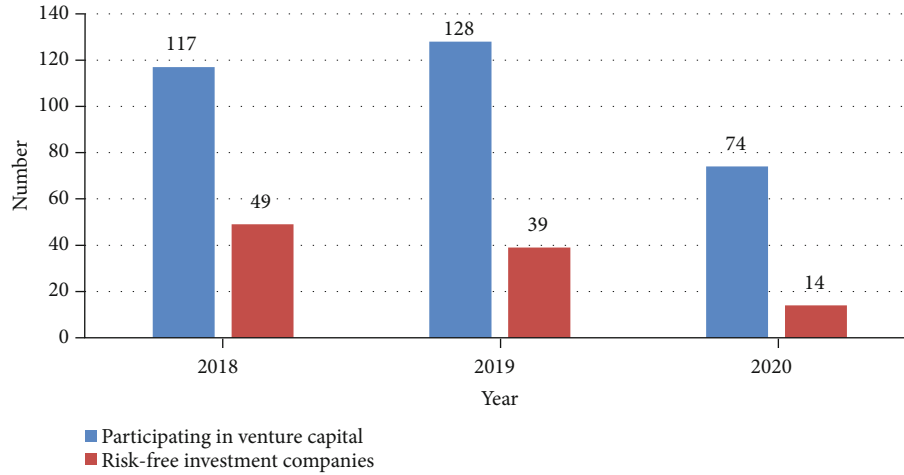


FIGURE 2: Comparison of the number of companies participating in venture capital and the number of risk-free companies.

TABLE 4: Descriptive statistics of variables.

Variable code	Average value	Median	Max	Min	Standard deviation	Number of samples
Rank	2.956	3	4	1	0.424	453
VC	0.678	1	1	0	0.468	453
VC-Country	0.147	0	0.74	0	0.354	307
VC-Private	0.853	0.06	7	0	0.354	307
VC Share	0.085	2	24.212	0.0004	0.093	307
VC-Count	2.391	20.601	26	19.491	1.335	307
Size	20.846	0.333	0.6	1	0.554	453
Age	11.082	0.333	0.882	0.25	4.217	453
Indep	0.375	9.446	27.505	0.042	0.055	453

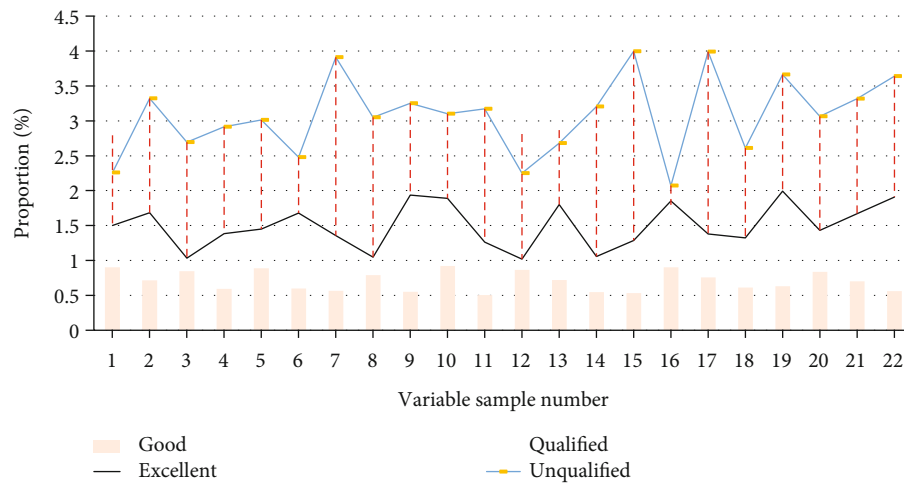


FIGURE 3: Distribution of assessment results.

indicate that the degree of stock price collapse risk between companies is quite different. This is also roughly the same as the previous research. The average and standard deviation analysis is shown in Figure 7.

Capital market is very different from the more mature Western capitalist markets. State-owned listed companies

account for a relatively large proportion, and the external environment and corporate governance characteristics of state-owned listed companies and non-state-owned listed companies are very different. In state-owned listed companies, the state, as the largest shareholder, has dominant control rights, and its owner agents are often directly appointed

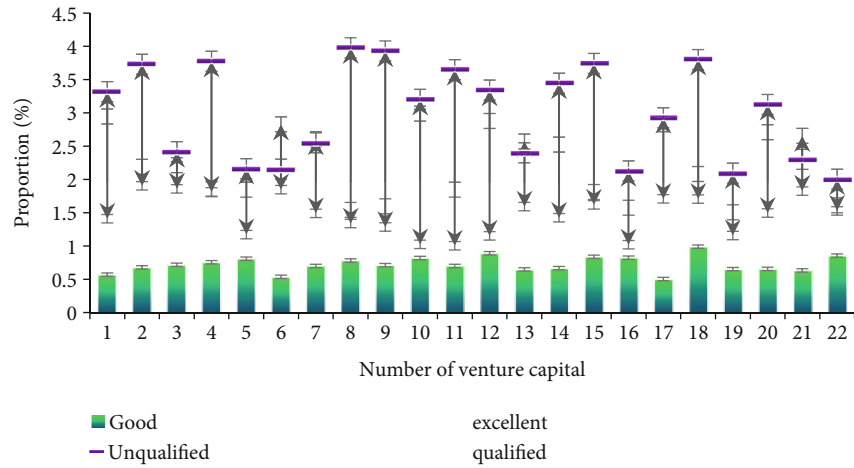


FIGURE 4: Number of venture capital.

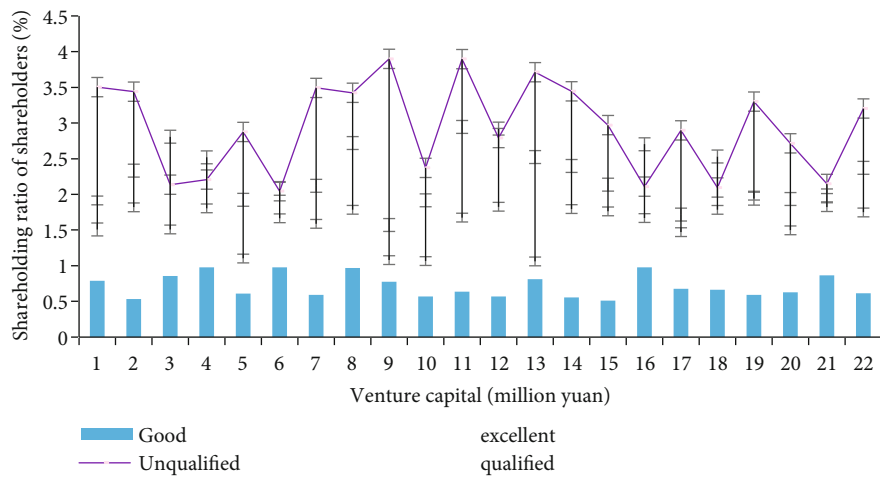


FIGURE 5: Total venture capital holdings.

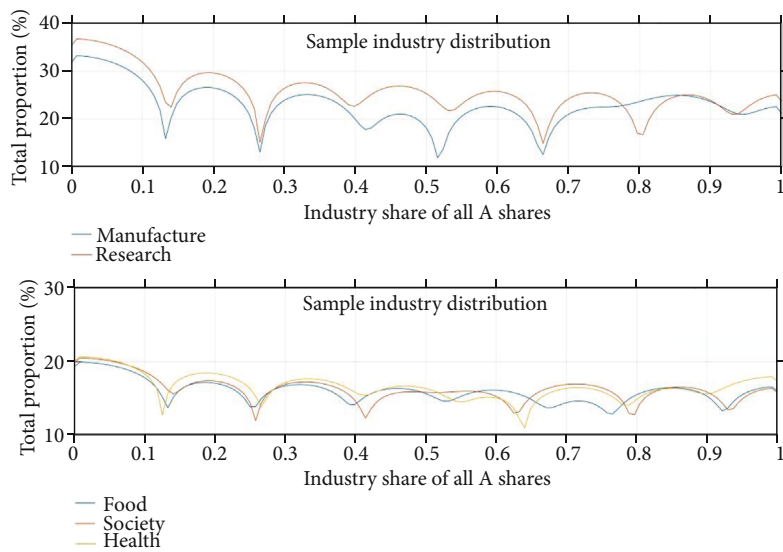


FIGURE 6: Distribution of sample industries.

TABLE 5: Descriptive statistics results.

Variable name	Model variable name	Average value	Standard deviation	Median
Negative weekly return skewness	NCSKE	-0.15695	0.859608	-0.36935
Stock volatility ratio	DUVOL	3.681745	0.737948	-0.22566
Risk information disclosure score	Risk	1.779564	1.428184	4
Disclosure of internal risk information	Risk1	1.883812	1.046066	2
External risk information disclosure	Risk2	0.080101	1.053377	2
Average excess turnover	Dturn	-0.00036	0.429211	0.093949
Weekly return	RET	0.060879	0.00845	-0.00095
Weekly return standard deviation	Sigma	0.706007	0.021681	0.057316
Market to account ratio	MB	22.14004	0.753703	0.465022
Company size	Size	0.422304	1.237735	22.0009
Financial leverage	LEV	0.041794	0.2092	0.408265

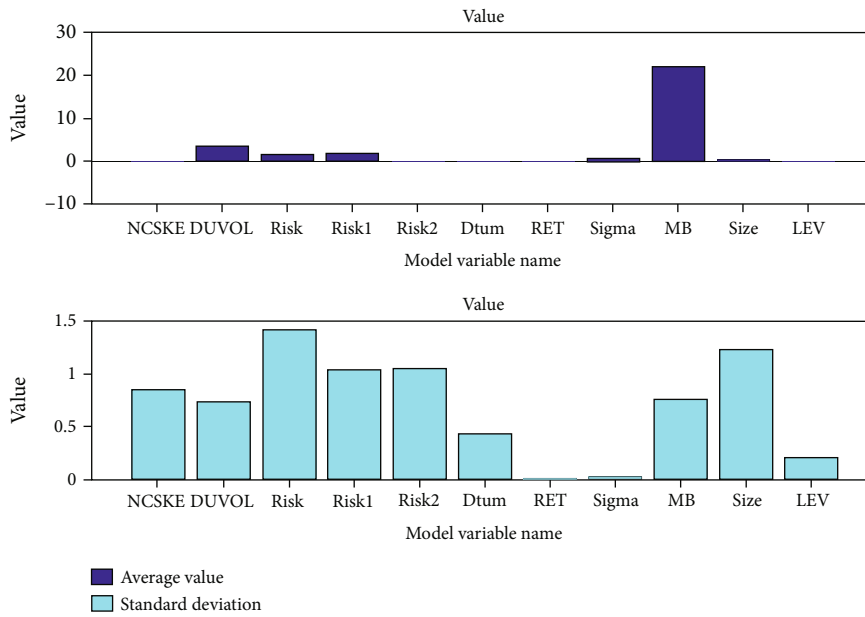


FIGURE 7: Analysis of mean and standard deviation.

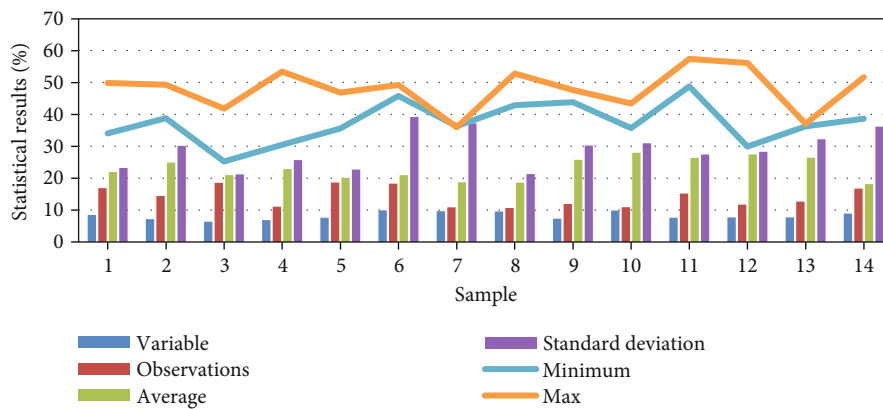


FIGURE 8: Descriptive statistical analysis results.

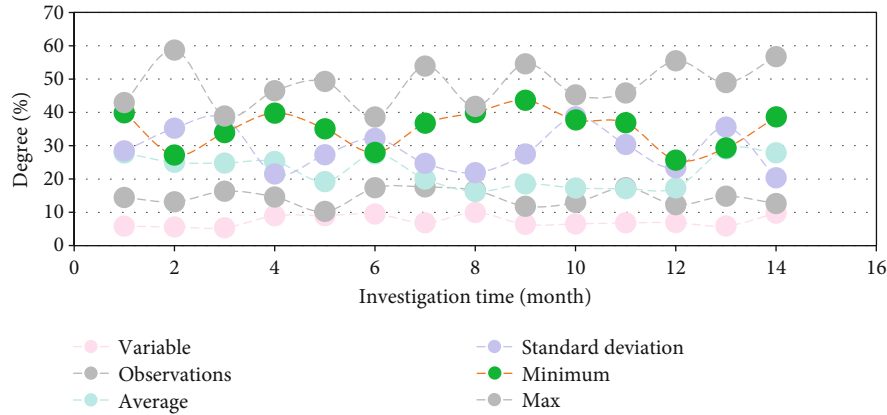


FIGURE 9: Degree of information disclosure of non-state-owned enterprises.

TABLE 6: Descriptive statistical results of risk information disclosure distinguishing the nature of property rights.

Property right	Variable	Observations	Average	Standard deviation	Minimum	Max
State-owned	Risk	1497	3.546426	1.493977	0	10
Non-state-owned	Risk	2857	3.752888	1.387663	0	10

by the government, that is, officials at all levels of government functional agencies. This has led to the inability of small and medium shareholders to govern and control the company. External information disclosure exercises effective supervision functions. However, non-state-owned listed companies, that is, private enterprises, have substantial controlling shareholders due to their independent management characteristics. Compared with state-owned listed companies, major shareholders can more actively and effectively exercise supervisory functions on the management. In order to further examine the relationship between risk information disclosure and stock price crashes, we divide the sample into two groups of state-owned enterprises and non-state-owned enterprises according to the nature of property rights and substitute the two sets of data into the regression equation to examine the impact of different types of corporate risk information disclosure on stock price crashes. It can be seen that the average value of risk information disclosure risk in state-owned enterprises is 3.546426, which is less than the average value of risk information disclosure risk in non-state-owned enterprises (3.752888). Therefore, on the whole, the risk information disclosure of non-state-owned enterprises in corporate annual reports the situation is better than that of state-owned enterprises. This shows that in state-owned listed companies, the management has more motivation to hide “bad news” for their own political promotion and reputation image and less disclosure of risk information the company faces, which exacerbates the risk of stock price collapse. The results of descriptive statistical analysis are shown in Figure 8. Figure 9 shows the degree of information disclosure of non-state-owned enterprises. The descriptive statistical results of the risk information disclosure that distinguish the nature of property rights are shown in Table 6.

5. Conclusion

GEM-listed companies refer to the general term for companies that, upon application and approval, issue and circulate stocks in the GEM market of various countries and regions are subject to the evaluation and constraints of the GEM market system in various countries and regions. GEM-listed companies have disclosed a series of financial indicators that indicate the company’s profitability, solvency, and operating capabilities in their prospectus and annual reports. The following are the selection principles: select comparable indicators, and the increase or decrease of such indicators can be used for the company’s vertical or horizontal comparative analysis; select the indicators disclosed in the prospectus and annual reports to select financial indicators from the perspective of the risk information needs of investors and regulators.

Here, it is easy to understand that the risk information disclosed by listed companies affects the changes of the company’s financial indicators by influencing investors’ decisions. We can analyze whether there is a difference between the disclosed risk information and the changes in the company’s financial indicators. There are relevant relationships to test whether the risk factors disclosed by the issuer are true, accurate, and complete.

The research model of this article is divided into two aspects. The first aspect mainly analyzes whether there is venture capital investment and how the four different characteristics of venture capital affect the quality of information disclosure of GEM and listed companies. First, add the five variables of venture capital: whether there are venture capital holding VC, venture capital state-owned background VC-Country, venture capital private background VC-Private,

venture capital holding ratio VC-Share, and venture capital Jointly held VC-Count, to independently return the information disclosure quality of the explained variable GEM-listed companies. The second aspect is to analyze the moderating effect of company size on the relationship between venture capital and its characteristics and the information disclosure quality rank of GEM-listed companies. By observing and judging whether the regression coefficient of the interaction term is significant, analyze the adjustment effect of the company size. This research helps to make up for the lack of data on GEM companies. To enhance the market's price discovery ability, it is necessary to truly focus on information disclosure, form a market supervision concept centered on information disclosure, and create a multilevel capital supervision system, thereby making this market the cornerstone of multilevel capital market.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- [1] M. Valibeigi and F. Ashuri, "Semantic analysis of the first cities from a deconstruction perspective," *Architecture and Engineering*, vol. 5, no. 3, pp. 43–48, 2020.
- [2] V. Phu, V. T. N. Tran, V. T. N. Chau, N. D. Dat, and K. L. D. Duy, "A decision tree using ID3 algorithm for English semantic analysis," *International Journal of Speech Technology*, vol. 20, no. 3, pp. 593–613, 2017.
- [3] D. T. N. Nguyen, S. T. T. Teo, and M. Ho, "Development of human resource management in Vietnam: a semantic analysis," *Asia Pacific Journal of Management*, vol. 35, no. 1, pp. 241–284, 2018.
- [4] R. Augustyn and E. M. Pramo, "The spread of Chinese virus in the internet discourse: a cognitive semantic analysis," *Gema Online Journal of Language Studies*, vol. 20, no. 4, pp. 209–227, 2020.
- [5] I. Tzankova and E. Cicognani, "Youth participation in psychological literature: a semantic analysis of scholarly publications in the PsycInfo database," *Europe's Journal of Psychology*, vol. 15, no. 2, pp. 276–291, 2019.
- [6] E. C. Simatupang, "Syntactic and semantic analysis on slogans of aviation companies in ASEAN countries," *English Journal Literacy UTama*, vol. 3, no. 2, pp. 123–131, 2019.
- [7] D. A. K. Khotimah and R. Sarno, "Sentiment analysis of hotel aspect using probabilistic latent semantic analysis, word embedding and LSTM," *International Journal of Intelligent Engineering and Systems*, vol. 12, no. 4, pp. 275–290, 2019.
- [8] F. Horasan, H. Erbay, F. Varçın, and E. Deniz, "Alternate low-rank matrix approximation in latent semantic analysis," *Scientific Programming*, vol. 2019, no. 2, 2019.
- [9] E. Y. Jang, "A cognitive semantic analysis on the polysemy 'Down'," *The New Korean Journal of English Language & Literature*, vol. 61, no. 2, pp. 231–250, 2019.
- [10] T. Sun and B. Zhong, "A tale of four cities: A semantic analysis comparing the newspaper coverage of air pollution in Hong Kong, London, Pittsburgh, and Tianjin from 2014 to 2017," *Newspaper Research Journal*, vol. 41, no. 1, pp. 37–52, 2020.
- [11] W. Wu, Y. Liu, C. H. Wu, and S. B. Tsai, "An empirical study on government direct environmental regulation and heterogeneous innovation investment," *Journal of Cleaner Production*, vol. 254, p. 120079, 2020.
- [12] O. Vasylyk, "The names of Uman suburbs dating from the 17th to the 18th centuries: structural-semantic analysis," *Alfred Nobel University Journal of Philology*, vol. 15, no. 15, pp. 233–238, 2018.
- [13] L. Yu and Z. Junhuai, "Semantic analysis of the Wa tautology in subordinate clauses with minimal independence degree," *International Journal of Advanced Network Monitoring and Controls*, vol. 3, no. 2, pp. 40–43, 2018.
- [14] E. Khuzina and R. Mukhtarova, "The lexico-semantic analysis of the common and various in world of gender picture in the English and Tatar linguistic cultures," *Journal of History Culture and Art Research*, vol. 7, no. 4, p. 214, 2018.
- [15] T. Rohani and S. Arsyad, "Semantic analysis of figurative language expressions in "feature" of the Jakarta post," *Journal of Applied Linguistics and Literature*, vol. 1, no. 1, pp. 97–114, 2018.
- [16] A. Czech-Rogoyska and M. Krawiec, "It-related Anglicisms in der Spiegel: a semantic analysis," *Social Communication*, vol. 4, no. 1, pp. 12–24, 2018.
- [17] K. Salehuddin, "The value of knowledge in the Malay mind: a cognitive semantic analysis," *International Journal of Asia-Pacific Studies*, vol. 14, no. 2, pp. 39–59, 2018.
- [18] R. N. Peclat and G. N. Ramos, "Semantic analysis for identifying security concerns in software procurement edicts," *New Generation Computing*, vol. 36, no. 1, pp. 21–40, 2018.
- [19] R. Fernandez-Beltran, J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and F. Pla, "Multimodal probabilistic latent semantic analysis for Sentinel-1 and Sentinel-2 image fusion," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 9, pp. 1347–1351, 2018.
- [20] N. Andrienko and G. Andrienko, "State transition graphs for semantic analysis of movement behaviours," *Information Visualization*, vol. 17, no. 1, pp. 41–65, 2018.
- [21] S. B. Tsai, "Using the DEMATEL model to explore the job satisfaction of research and development professionals in China's photovoltaic cell industry," *Renewable and Sustainable Energy Reviews*, vol. 2018, no. 81, pp. 62–68, 2018.
- [22] W. Cui, J. Du, D. Wang et al., "Short text analysis based on dual semantic extension and deep hashing in microblog," *ACM transactions on intelligent systems*, vol. 10, no. 4, pp. 1–38.24, 2019.
- [23] J. Zhang, Y. Mu, S. Feng, K. Li, Y. Yuan, and C. H. Lee, "Image region annotation based on segmentation and semantic correlation analysis," *IET Image Processing*, vol. 12, no. 8, pp. 1331–1337, 2018.
- [24] Y. Shin, "What can tripartite semantic network analysis do for media framing research?," *Information Communication and Society*, vol. 33, no. 1, pp. 121–137, 2020.
- [25] D. Tian, "Research on PLSA model based semantic image analysis: a systematic review," *Journal of Information Hiding*

- and Multimedia Signal Processing*, vol. 9, no. 5, pp. 1099–1113, 2018.
- [26] A. Lenci, G. E. Lebani, and L. C. Passaro, “The emotions of abstract words: a distributional semantic analysis,” *Topics in Cognitive Science*, vol. 10, no. 3, pp. 550–572, 2018.
- [27] R. U. Ahrorova, “Semantic analysis of phraseological units representing age periods in French and Uzbek languages,” *Theoretical & Applied Science*, vol. 83, no. 3, pp. 186–189, 2020.
- [28] C. Andujar, C. R. Vijulie, A. Vinacua, B. S. Santos, and G. Alford, “Syntactic and semantic analysis for extended feedback on computer graphics assignments,” *IEEE Computer Graphics and Applications*, vol. 40, no. 3, pp. 105–111, 2020.