Evaluation and Prognosis Analysis of Corporate Financial Risks Based on Big Data Context

Qu Rong

Email: u3600839@connect.hku.hk

The University of Hong Kong, Faculty of Business and Economics, Hong Kong 999077, China

Abstract: With the rapid development of the Chinese economy, Chinese manufacturing companies have grown rapidly in this context, but along with this development comes many difficulties for these manufacturing companies. In addition, the outbreak of the new crown epidemic has increased the risk to these manufacturing companies, and among these risks, financial risk is particularly important, so it is even more important to be able to evaluate financial risk effectively. However, most companies only consider the impact factors of financial indicators when evaluating financial risk, while ignoring the impact factors of non-financial indicators, which inevitably results in an overly one-sided and inaccurate financial risk evaluation. Therefore, companies has to balance financial data information and non-financial data information when evaluating financial risks. The issue tracking system is an important part of Github and is used by an increasing number of users and developers to submit issue reports. Github provides a tagging mechanism to speed up the processing of issues. Tags allow developers to keep track of issue types and prioritise their processing, but the use of tags is still rare on Github, largely due to the manual nature of tag allocation.

Keywords: Big Data; Corporate Finance; Risk assessment; Prognostic analysis

1 INTRODUCTION

In software engineering development, both developers and users play an important role and a successful project cannot be achieved without the efforts of all [1]. In practice, software projects often go through many iterations, and each update requires timely feedback from users [2]. Only the continuous development of various industries can continuously drive the development of the economy. Among the many industries, the automobile manufacturing industry has a very big role in promoting the development of China's economy, but along with the economic development, the market transformation has been intensified, and the traditional automobile manufacturing industry faces many risks in the process of its development, for example, the automobile manufacturing industry is facing risks such as the transformation and upgrading situation [3]. These risks are bound to cause significant financial crises for automotive manufacturing companies and are bound to cause difficulties in the development of the automotive manufacturing industry. Therefore, it is necessary for automotive manufacturing

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

DOI 10.4108/eai.19-5-2023.2334402

companies to conduct an effective financial risk assessment in order to be able to effectively identify potential financial crises [4].

2 FINANCIAL RISK-RELATED OVERVIEW

Financial risk is narrowly defined as the risk of financing a company's production and operating activities alone, as a result of the company's liabilities taking up a disproportionate share of its capital structure, thus making it unable to pay off its debts as they fall due in the short term out of earnings [4]. From a broad perspective, the company's financial risk is not only the financing risk, but can be broadly divided into four categories, one is the risk brought by financing, the second is the risk brought by investment, the third is the risk brought by the company's operations, and the fourth is the risk brought by the distribution of earnings after the company has made a profit [5].

2.4 Approach to financial risk evaluation

(1) Entropy method

The specific equation is shown in 1. where H represents the entropy, and when the signal source is uncertain n, the likelihood of its occurrence is i.

$$H(U) = -\sum_{i=1}^{n} P_i log_p i \tag{1}$$

(2) Definition of the entropy value method

With the continuous development and improvement of information entropy, the entropy method also emerged, which is still based on the data provided by the user, which makes it unstable, if the user provides too much information, the calculation result will be correspondingly smaller, and vice versa [6].

(3) Hierarchical analysis

Hierarchical analysis, also known as AHP, will be calculated in three layers, firstly the target layer, then a second layer below this layer, the criterion layer, then a third layer below it, the indicator layer, based on this layer, and finally the corresponding weights are calculated through these three layers [7].

(4) Efficacy factor method

The efficacy coefficient method, also known as the efficacy function method, focuses on setting the satisfaction value to find out how well the satisfaction level of each indicator is achieved, and then setting the disallowance value to finalise the final score of each indicator, before finally making a comprehensive evaluation [8].

2.5 Calculation method of the efficiency factor method

This paper focuses on an overview of the traditional efficacy factor method, while the improved efficacy factor method that will be adopted in this paper stands on top of the traditional efficacy factor method. In contrast, the traditional efficacy coefficient method is used to assign values to individual indicators. The traditional efficacy factor method is formulated as:

$$Z = \sum [60 + (A_i - B_i)/(M_i - B_i) \times 40] \tag{2}$$

Where Z refers to the overall efficacy score, A_i is the actual value for a given value up to a certain point, B_i is the impermissible value for a given value up to a certain point and M_i is the satisfactory value for a given value up to a certain point [9].

3 EXPERIMENTAL PROCEDURE

3.1 Experimental steps and evaluation

In this paper, the trained model is called for each issue text to obtain the probability distribution over different topics, and the topic vector and word vector together form the new features. Once the fused feature vectors are available, the classification of the issues is carried out according to the convolutional neural network model constructed and trained, and the classification results are saved in a text file for subsequent statistics and preservation. The data set was divided into 10 parts, 9 of which were selected as the training set and 1 as the test set.

Once the classification of the issues had been completed through the above steps, two questions were posed for study based on existing experience in order to judge the reasonableness of the classification results:

(1) RQ1: How accurate is the issue classification?

(2) RQ2: How valid is the issue classification?

3.2 Analysis of the accuracy of issue classification

In this paper, we hope that for an issue text, adding a global topic vector will improve the effectiveness of the classification model. The results of the experiments are shown in Figure 1: the word vector and the text topic vector have a slightly worse effect than the other dimensions when the dimension is set to 50, and the classification results are similar in 100, 150 and 200 dimensions, with 150 being the best and decreasing when the dimension is 200, indicating that a large word embedding dimension is not conducive to the expression of word features, which increases the time complexity and makes the model more complex. Since vector dimensionality does affect classification accuracy, given the experimental environment, a vector dimension of 150 dimensions was eventually used as the input to the convolutional neural network [10].

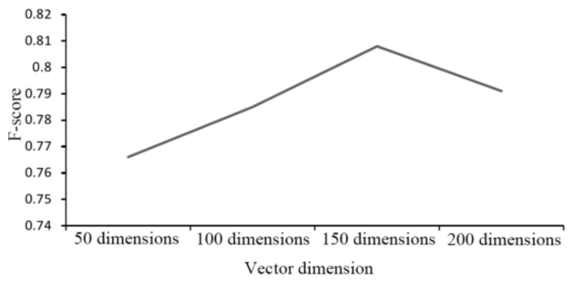

Figure 1: Chart of F-score by dimension

The issue classification method in this article is based on the Github platform, where issues generally have a title, description, and comment information. Given the need for timeliness in classification, each issue in the dataset does not contain comment information, as it can be a long time after an issue is submitted before a user comments on it, and it is clearly not appropriate to wait until the user and project staff have discussed it before starting classification. Therefore on the basis of such a dataset one tests one's own method on the one hand and compares it with other classification methods on the other, the experiments compared in this paper consider two main points [11]:

(1) FastText-based issue classification.

(2) Issue classification based on convolutional neural networks using only word2vec for vector representation.

The classification performance of the method proposed in this paper is shown in Figure 2. Good results were achieved in different categories, with an accuracy rate of over 80% in classification and 80.8%, 80.7% and 80.7% in the three metrics of accuracy, recall and F-score respectively in overall terms.

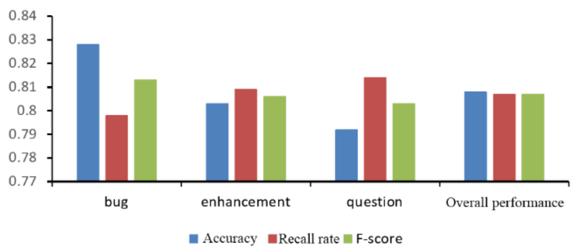

Figure 2: Effect of issue classification

Figure 2 visualises the classification effect on the different issue categories in the form of a bar chart, and in general the experimental approach meets expectations.

The results of the FastText-based issue classification method are shown in the figure below, which shows that the method in this paper is effective, and all three metrics have improved to some extent. Specifically the accuracy metric improved by 1.7%, the recall metric improved by

2.5% and the F-score metric improved by 2.1%, confirming that the method in this paper learns key information about the issue very well.

As show in figure 3&4.

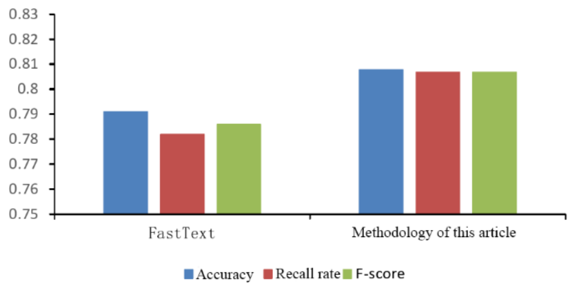

Figure 3: Comparison between FastText classification and this method

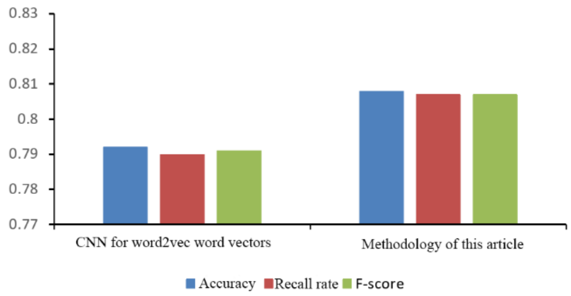

Figure 4: Comparison of CNN of word2vec word vectors and this paper's approach to classification

3.3 Analysis of the validity of issue classification

To answer RQ2, this paper requires a manual experiment to verify that issue classification tags can help developers to review issues more effectively. Specifically, eight postgraduate students with around two years of development experience were selected as researchers to participate in this experiment. Three open sources projects of ongoing interest and familiarity were selected for the experiment, with 10 issues selected for each project, and participants were divided equally into two groups, with the participants not knowing the category information in advance. The first group of participants first reviewed the first 5 issues without tags, based only on the title and description information, and then reviewed the content of the next 5 issues in conjunction with the generated categorical tags, while the second group reviewed the first 5 issues with tag information and the next 5 issues without tags, and each participant in the experiment was asked to give the time spent reviewing each issue.

Box plots 5, 6 and 7 show the distribution of user research results for each of the three project issues, and from the box plots it can be seen the time taken to review issues using the tags generated by the issue categories is generally less than the time taken to review the native issues. As show in figure 5&6&7.

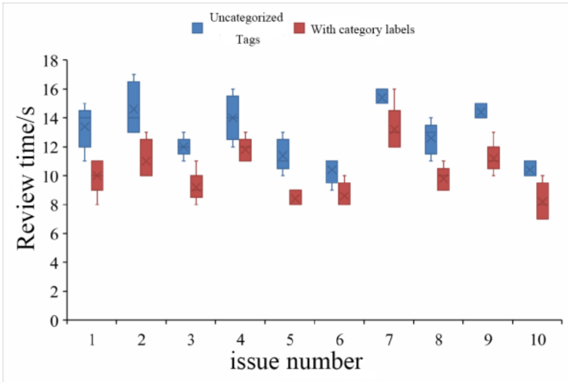

Figure 5: User review results for grpc-java

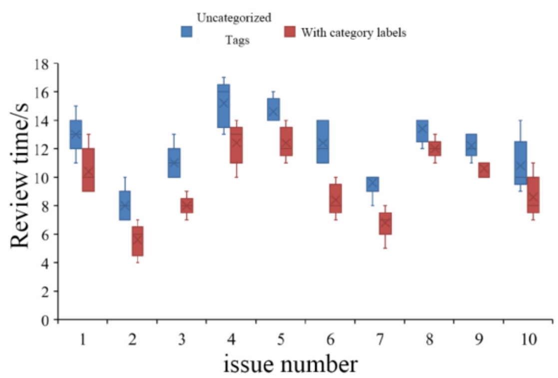

Figure 6: User review results for vue

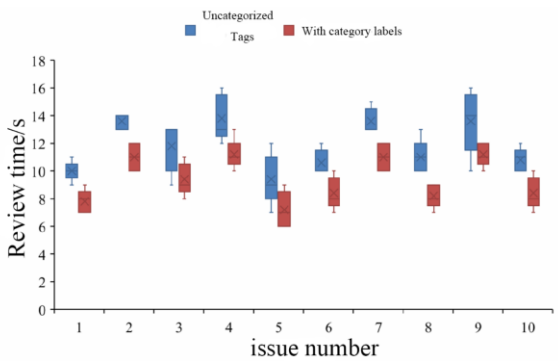

Figure 7: Results of the user review of react

4 CONCLUSIONS

This paper focuses on the issue classification model designed for the experiments and the analysis of the results. The experimental setting and evaluation metrics are first described, then validated in terms of accuracy and validity. In terms of accuracy, the paper focuses on three metrics and provides a comparative analysis, and in terms of validity, the manual experiments demonstrate the value of categorised labels in reducing the time overhead of reviewing issues. Overall, the work met expectations and any shortcomings will continue to be improved in the future. It is an analysis of the evaluation method currently used by HM and how it is actually evaluated. After analysis, it is found that it mainly compares its own statement data with previous years and then compares it with the corresponding industry averages, and although the company's financial risk problems can be obtained through these two angles of analysis, it is not possible to obtain the level of risk of the financial indicators, and it is not possible for managers to accurately identify which specific indicators of the company have risk problems, nor can they know how big the level of risk of the indicators is. In addition, the company's selection of indicators prior to the evaluation process included only financial indicators. Although this dimension alone could reflect the company's risk level to a certain extent, the degree of reflection was limited, so an important element of non-financial indicators was added to the indicator selection step before the improved efficacy factor method was used for the evaluation process.

REFERENCES

[1] Fan Yu He. Research on financial risk evaluation of Shenzhou information based on deep neural network[D]. Shihezi University, 2022. DOI:10.27332/d.cnki.gshzu.2022.000723.

[2] Gu Huan. Research on enterprise financial risk management in the context of big data[J]. Investment and Entrepreneurship,2023,34(04):43-45.

[3] Gou Linyuan. Thinking about enterprise financial risk management in the context of big data[J]. DOI:10.13665/j.cnki.hzjjykj.2022.23.054.

[4] Li Xiaoyan. Prevention and control measures of enterprise financial risks in the context of big data[J]. PR World,2022(19):91-93.

[5] Liu Rui. Exploration of enterprise financial risks and internal control measures in the context of big data[J]. China Collective Economy,2022(23):160-162.

[6] Sun Ran, Man Haihong. Financial risk management of enterprises in the context of big data[J]. Cooperative Economics and Technology,2022(21):144-145.DOI:10.13665/j.cnki.hzjjykj.2022.21.065.

[7] Song Wenxiu. Analysis of enterprise financial risks and preventive and control measures in the context of big data[J]. Mall Modernization, 2022(09):139-141. doi:10.14013/j.cnki.scxdh.2022.09.024.

[8] Xu Shan. A study on enterprise financial risk management in the context of big data - taking Alibaba Group as an example [C]//Chongqing City Dingying Culture Communication Co. 2022 Social Development Theory Symposium Proceedings (III). 2022 Social Development Theory Symposium Proceedings (III), 2022:17-20. doi. 10.26914/c.cnkihy.2022.049321.

[9] Zhang S. Lemon. Research on early warning mechanism of enterprise financial risk in the context of big data[J]. Small and medium-sized enterprise management and technology,2022(06):79-81.

[10] Zuo Qianqi. Research on the prevention and control of enterprise financial risks in the context of big data[J]. Bohai Economic Outlook, 2022(11):168-170. doi:10.16457/j.cnki.hbhjjlw.2022.11.035.

[11] Zhang Yining, Gu Quangen. Research on financial risk control of enterprises in the context of big data[J]. Shanghai Business,2022(09):117-119.