Credit Assessment and Risk Control Modeling Method of Enterprise Digital Transformation Driven by Big Data

Yaoyi Ying

Harbin University of Commerce, Harbin, 150000, Heilongjiang, China

Keywords: Big data; Digital transformation of enterprises; Credit assessment; Wind control modeling; Digital age

Abstract: With the development of big data technology and the digital transformation of enterprises, the limitations of traditional enterprise credit assessment and risk control modeling methods are increasingly prominent. This paper focuses on the research on credit assessment and risk control modeling method of digital transformation of enterprises driven by big data. By analyzing the relevant theoretical basis, a big data-driven enterprise credit assessment system is constructed, and the assessment indicators are selected from the dimensions of finance, operation, innovation and network behavior, such as the coefficient of variation of cash flow from operating activities in recent three years, market share growth rate, etc., and the index weights are determined by comprehensive analytic hierarchy process and entropy method. At the same time, using machine learning (ML) and deep learning (DL) technology, the wind control model is constructed through data collection and pretreatment, model selection and training, and the training effects of different models are compared. Among them, the accuracy of LSTM model is 0.82, the recall rate is 0.80, and the F1 value is 0.81. The research results provide scientific and feasible methods and theoretical support for enterprises to realize accurate credit assessment and effective risk prevention and control in the process of digital transformation.

1. Introduction

Under the wave of the digital age, the deep integration of big data technology and enterprise digital transformation is reshaping the operation and development model of enterprises [1]. Big data, with its mass, diversity, high speed and low value density, provides enterprises with unprecedented insight, and the digital transformation of enterprises has become the key path to enhance competitiveness and adapt to rapid market changes [2]. In this context, enterprise credit assessment and risk control modeling are particularly important, which are important guarantees for the steady development of enterprises [3].

Traditional enterprise credit assessment and risk control modeling methods rely on limited structured data and relatively static models, which are difficult to cope with the complexity and dynamic changes of enterprise business in the process of digital transformation [4]. With the advancement of digital transformation of enterprises, data sources are more extensive, covering unstructured data such as social media data and Internet of Things device data. This brings new challenges and opportunities for credit assessment and risk control modeling [5]. How to build a more accurate, real-time and effective credit assessment and risk control modeling method driven by big data has become a key issue to be solved urgently [6].

This paper focuses on the research of credit assessment and risk control modeling methods for digital transformation of enterprises driven by big data, aiming to fill some gaps in the current research in this field and provide scientific and feasible theoretical and methodological support for risk management of enterprises in the process of digital transformation. Through in-depth analysis of the internal relationship between big data and enterprise digital transformation, effective data and technical means suitable for enterprise credit assessment and risk control modeling are mined, and an innovative method system is constructed.

  DOI: 10.25236/icmmct.2025.040

2. Credit assessment and risk control modeling related theory

Big data refers to a collection of data that cannot be captured, managed and processed by conventional software tools within a certain time range, and has the characteristics of mass, diversity, high speed and value. Its core lies in the analysis of massive and complex data, mining potential value and providing basis for decision-making [7]. The digital transformation of enterprises is a process in which enterprises make comprehensive changes in business model, organizational structure and operation process by using digital technology. It takes customers as the center, and improves the innovation ability, operational efficiency and competitiveness of enterprises by integrating digital technology and business activities [8]. Digital transformation usually goes through three stages: digitalization, digitization and intelligence, which involves changes in technology, organization and culture, and is an inevitable choice for enterprises to adapt to the development of the digital economy era.

The purpose of credit assessment is to judge the possibility of default and credit status of enterprises through the analysis of all kinds of information. Traditional credit assessment is based on limited data such as financial statements and credit records, and uses methods such as expert judgment and credit scoring model. Modern credit assessment introduces technologies such as ML and DL, which can handle more dimensional data and improve the accuracy and timeliness of assessment [9]. The development of credit assessment theory provides an important support for enterprise credit risk identification and management. Risk control modeling is a process of quantitative analysis and prediction of risks by constructing mathematical models. It follows the principles of data quality, model rationality and interpretability, and uses historical data to train and verify the model. Risk control modeling aims to identify potential risk factors, evaluate the degree of risk, provide scientific basis for enterprises to formulate risk coping strategies, and help enterprises achieve steady development.

3. The construction of enterprise credit assessment system driven by big data

3.1. The limitations of traditional enterprise credit assessment

Traditional enterprise credit assessment mainly relies on limited structured data such as enterprise financial statements and bank credit records. The assessment method is relatively simple, and most of them use expert scoring method or simple statistical model. This method has limitations in data dimension, and cannot fully reflect the complex business situation and potential risks of enterprises in the process of digital transformation. Moreover, the traditional assessment is often static, which can not track the changes of enterprise credit in real time, and it is difficult to warn the credit risk in time when the market environment is changing rapidly.

3.2. Advantages and ideas of big data-driven credit assessment

Credit assessment driven by big data has significant advantages. It can integrate multi-source data, including internal operation data, external market data, social media data, etc., and realize all-round description of enterprises. By collecting and analyzing data in real time, we can also dynamically monitor the credit status of enterprises. Based on this, to build a big data-driven credit assessment system, we need to start with expanding data sources, using advanced analysis technology, mining key indicators that can accurately reflect corporate credit, and using ML and other algorithms to establish a dynamic assessment model.

3.3. Credit assessment index selection and weight determination

Combined with the characteristics of digital transformation enterprises, this paper selects credit assessment indicators from multiple dimensions (see Table 1).

Table 1: Big Data-Driven Enterprise Credit Assessment Indicator System

Primary Indicator	Secondary Indicator	Indicator Description
Financial Dimension	Debt-to-Asset Ratio	The ratio of total liabilities to total assets, measuring a company's long-term debt repayment ability
Financial Dimension	Current Ratio	The ratio of current assets to current liabilities, assessing a company's short-term debt repayment ability
Financial Dimension	Coefficient of Variation of Operating Cash Flows in Recent Three Years	Reflects the stability of a company's operating cash inflows, with a smaller coefficient indicating greater stability
Operational Dimension	Market Share Growth Rate	The growth rate of a company's market share compared to the previous period, reflecting its market expansion capability
Operational Dimension	Customer Complaint Rate	The ratio of the number of complaining customers to the total number of customers, measuring product or service quality
Innovation Dimension	Number of Patent Applications	The number of patents applied for by a company within a certain period, reflecting its technological innovation capability
Innovation Dimension	R&D Investment Ratio	The ratio of research and development investment to operating revenue, reflecting the company's investment intensity in innovation
Online Behavior Dimension	Monthly Active Users	The number of monthly active users on a website or platform, reflecting the attractiveness of its online business
Online Behavior Dimension	Positive Mention Rate on Social Media	The ratio of the number of positive assessments on social media to the total number of mentions, reflecting the company's public image

Determining the index weight is the key link to build a credit assessment system. Analytic Hierarchy Process (AHP) and entropy method can be used comprehensively. Firstly, using AHP method, industry experts are invited to score the relative importance of each index, build a judgment matrix and calculate the subjective weight of each index. Then, using entropy method, the objective weight is determined according to the dispersion degree of index data. Finally, through the combination weighting method, the subjective weight and the objective weight are combined to get a more scientific and reasonable comprehensive weight, so as to build an accurate and effective credit assessment system.

4. Big data-driven enterprise risk control modeling methods

4.1. Deficiencies of traditional risk control modeling methods

Traditional risk control modeling methods are mainly based on historical experience and a small amount of structured data, and often adopt simple models such as linear regression and logical regression. These methods are difficult to deal with massive, complex and dynamic data. In addition, there are many assumptions in the traditional model, and in reality, the operation situation of enterprises is complex and changeable, which makes it difficult to meet the assumptions, resulting in limited adaptability and accuracy of the model, and it is impossible to identify and warn risks in time and accurately.

4.2. Big data-driven risk control modeling process and key technologies

Big data-driven risk control modeling relies on ML and DL technologies. Decision tree and random forest algorithm in ML can deal with nonlinear relations and automatically screen important features, which can be used to classify and predict risks in risk control modeling. DL neural network models, such as long-term and short-term memory networks (LSTM), can mine deep-seated features and patterns of data, and are suitable for processing time series data and complex data structures, thus improving the accuracy of risk prediction.

Data collection and pretreatment: collect the internal and external data of the enterprise extensively, including financial data, transaction records, market public opinion, etc. Data preprocessing is very important, so it is necessary to clean the data and remove duplicate and erroneous data (see Table 2). At the same time, the data are standardized and normalized to make different feature data at the same scale, which is convenient for model training.

Table 2: Comparison Before and After Data Cleaning

Data characteristics	Data before cleaning	After cleaning data
Customer age	-5 (abnormal value)	Delete the abnormal record.
Transaction amount	100,000 (wrong format, extra comma)	100000
Product name	Mobile phone (one more space)	Mobile phone

Model selection and training: select the appropriate model according to the data characteristics and wind control targets. In this paper, LSTM model is more suitable for dealing with time series risk prediction. Training data are used to train different models, and model parameters are adjusted to make the model achieve better performance on the training set.

The dimension of the input data of LSTM model depends on the number of features selected. In this paper, three characteristics of enterprise revenue, cost and market interest rate are selected as inputs, so the dimension of input layer is 3. At the same time, the time step is determined, that is, the length of continuous time series data of each input model. Based on the monthly data, this paper considers the influence of the data of the past 12 months on the current risk, and the time step is set to 12.

Hidden layer: select an appropriate number of LSTM units. Generally, the performance of the model is optimized by testing different numbers of units, such as trying 64 LSTM units first, observing the model effect, and then adjusting to 128 units. Multiple LSTM layers can be stacked to learn more complex features, but at the same time, it also increases the training difficulty and the risk of over-fitting. This paper constructs two LSTM hidden layers, the first layer has 64 units and the second layer has 32 units. The calculation formula of hidden layer h is:

$$h = Tanh(W \cdot c + b) \tag{1}$$

$$c = [v_{\omega t-1}, \omega_{\omega t-2}, \ldots, \omega_{\omega t-n+1}] \tag{2}$$

Where $h = Tanh(W \cdot c + b)$ is a weight matrix, which is used to combine the context word vectors into a hidden layer representation. $h = Tanh(W \cdot c + b)$ represents the splicing of context word vectors. b represents the bias vector of the hidden layer. The update formula of hidden status is:

$$h_t = \sigma(W_{xh} x_t + W_{hh} h_{t-1} + b_h) \tag{3}$$

Where: σ is the activation function; W_{xh} is the weight matrix input to the hidden layer; W_{hh} is the weight matrix from hidden layer to hidden layer; b_h is the offset of the hidden layer.

The dimension of the output layer is determined according to the prediction target. If it is a binary classification problem, it is necessary to judge whether there are risks in the future. The dimension of the output layer is 1, and the sigmoid activation function is used. The output value is between [0,1], and a threshold value (such as 0.5) can be set to judge the risk category. If it is a multi-classification problem, if the risk degree is divided into three categories: high, medium and low, the dimension of the output layer is 3, and the softmax activation function is used to output the probability distribution of each category.

Model assessment and optimization: using test data to evaluate the model performance, commonly used indicators include accuracy, recall, F1 value, etc. (See Table 3 for the results). If the performance of the model does not meet the requirements, it is necessary to optimize the model, such as adjusting parameters, increasing data volume or using ensemble learning method to combine multiple models to improve the stability and accuracy of the model.

Table 3: Comparison of Training Effects of Different Models

Model name	Accuracy rate	Recall rate	F1 value
Logistic regression	0.75	0.72	0.73
Support vector machine	0.78	0.76	0.77
LSTM	0.82	0.80	0.81

It can be seen that the LSTM model is outstanding in risk prediction. Its accuracy, recall and F1 value reached 0.82, 0.80 and 0.81 respectively. Based on the modeling results, risk early warning indicators and thresholds are set. When the enterprise-related indicators reach or exceed the threshold, an early warning is triggered. According to the early warning level, enterprises formulate corresponding risk coping strategies, such as adjusting capital arrangements and optimizing business structure, so as to achieve effective risk prevention and control.

5. Conclusions

This paper focuses on the credit assessment and risk control modeling method of digital transformation of enterprises driven by big data, and has achieved a series of results. In the construction of credit assessment system, we break through the traditional limitations, integrate multi-source data and select key indicators from multiple dimensions. For example, the newly added coefficient of variation of cash flow from operating activities in the past three years can more accurately reflect the financial stability of enterprises; The growth rate of market share can reflect the development trend of enterprises, and the index weight can be determined by scientific methods, which makes the assessment system more scientific and comprehensive.

In risk control modeling, the modeling process is standardized with the help of key technologies of big data. After data collection and pretreatment, the data quality is effectively improved; By comparing the training effects of different models, it is found that the LSTM model is outstanding in risk prediction, with the accuracy, recall and F1 value reaching 0.82, 0.80 and 0.81 respectively, which provides strong support for risk early warning. The risk early warning mechanism based on the modeling results can issue early warning according to the risk degree in time, and help enterprises to formulate corresponding strategies to prevent and control risks.

Generally speaking, the method system constructed in this paper provides an important theoretical basis and practical guidance for enterprises to cope with the complex and changeable market environment and achieve steady development under the wave of digital transformation. Future research can further explore how to integrate emerging technologies more efficiently and continuously optimize credit assessment and risk control modeling methods to meet the ever-changing development needs of enterprises.

References

[1] Gao Chong, Shi Xin, Liu Feng. Enterprise Digital Transformation and Auditors' Decision-making [J]. Journal of Management Science, 2023, 36(3): 116-129.

[2] Dai Wanrui, Song Liangrong. The Digital Transformation of Banks and the Mitigation of Credit Risks: Based on a New Perspective of Economic Capital [J]. Enterprise Economy, 2023, 42(12): 127-138.

[3] Dou Yongxiang, Liu Ziyuan, Yang Dong. The Impact of Digital Transformation on the Risk-taking Level of Manufacturing Enterprises [J]. East China Economic Management, 2024, 38(9): 106-115.

[4] Zhao Chenyu. A Study on the Impact of Digital Transformation on Corporate Social Responsibility [J]. Contemporary Economic Science, 2022, 44(2): 109-116.

[5] Liu Bai, Guo Shuyan, Pan Mengmeng. Digitization and Credit Default Risks in Consumer Finance [J]. Journal of Management Science, 2023, 36(6): 33-44.

[6] Zhou Can, Zhang Jiyang. Digital Transformation and the Debt Default Risks of Distressed Enterprises [J]. Journal of Shanxi University of Finance and Economics, 2023, 45(5): 113-126.

[7] Xia Yiqun, Sheng Guangyin. A Study on the Credit Risk Sharing Mechanism of the Digital Transaction System for Intellectual Property Rights [J]. Finance and Accounting Monthly, 2023, 44(9): 141-147.

[8] Du Feiye, Huang Zhe, Lü Jianglin. Has Digital Transformation Promoted the Differentiated Competition of Enterprises? [J]. Enterprise Economy, 2024, 43(11): 129-139.

[9] Xiao Zhongyi. A Study on Digital Transformation and the Governance of Corporate Financial Risks [J]. Enterprise Economy, 2024, 43(11): 5-16.