Journal of computer science and software applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377-0430
Vol. 4, No. 7, 2024

Financial Risk Analysis Using Integrated Data and Transformer-Based Deep Learning

Yijing Wei[1], Ke Xu[2], Jianhua Yao[3], Mengfang Sun[4], Ying Sun[5]
[1]Northwestern University, Evanston, USA
[2]Columbia University, New York, USA
[3]Trine University, Phoenix, USA
[4]Stevens Institute of Technological, Hoboken, USA
[5]Columbia University, New York, USA
Correspondence should be addressed to Jianhua Yao; yaoboxi168@gmail.com

Abstract:

This paper explores the integration of multimodal data using Transformer models to enhance the accuracy of financial risk prediction. By combining diverse data sources such as time series, financial reports, and unstructured text (e.g., news and social media), this research offers a more holistic approach to identifying potential high-risk events in the financial market. We compare different data fusion strategies and demonstrate that the combination of textual and quantitative financial data significantly improves model performance, particularly in terms of AUC and Recall. Unlike traditional single-modality approaches, multimodal fusion enables the model to capture a wider range of risk signals, which are often latent or intertwined across data types. Our experimental results highlight the superiority of Transformer models in processing complex multimodal information, making them a powerful tool for financial risk assessment. The findings of this research are particularly relevant for regulatory bodies such as central banks, financial supervisory authorities, and risk management departments within financial institutions, as these sectors require more precise and adaptive tools to monitor market dynamics, identify systemic risks, and respond to financial crises. By leveraging the self-attention mechanisms of Transformers, this study offers an effective methodology for improving predictive accuracy in financial risk management, with practical implications for regulatory compliance and policy-making.

Keywords:

Financial risk prediction, multimodal fusion, self-attention mechanism, Transformer

1. Introduction

In the field of financial risk prediction, traditional analytical models are often limited to a single data source, such as time series data or financial reports[1]. However, the complexity of the financial market far exceeds the expression of information in a single dimension. Market fluctuations, corporate credit risks, and the formation of systemic financial crises are often affected by multiple factors. For example, financial reports and time series data can reflect the historical operating conditions and market behavior of enterprises, but sudden events, policy changes, and market sentiment are often expressed in the form of text data such as news and social media. How to effectively integrate these diverse sources of information has become a major challenge in the field of financial risk prediction[2]. To this end, the Transformer model provides a highly promising solution that can combine multimodal data to inject more dimensional information into financial risk prediction, thereby assessing risks more comprehensively and accurately[3].

Since the Transformer model made a breakthrough in the field of natural language processing (NLP), it has gradually become a mainstream tool for solving multimodal data problems with its powerful attention mechanism and parallel computing capabilities. Unlike traditional recurrent neural networks (RNNs) or convolutional neural networks (CNNs), Transformers do not need to rely on the sequential processing of sequence information and can more effectively capture long-range dependencies between data[4]. This is particularly important for financial risk prediction, because different types of information in financial data may have complex spatiotemporal correlations, and traditional models are often unable to handle these correlations[5]. By introducing the Transformer, the model can adaptively assign attention weights based on the input multimodal data, so as to more accurately understand and predict potential financial risks.

The integration of multimodal data provides a new perspective for financial risk prediction. News reports, policy changes, macroeconomic data, company financial reports and other information in the financial market often have an important impact on market trends, but these data types are different and have different forms of expression[6]. Text data can reveal market sentiment or event-driven risk factors, while financial statements can quantify the health of the company. Time series data such as stock prices and trading volumes reflect the market's feedback on this information. Therefore, how to fuse these heterogeneous data is an important task in financial risk prediction. Through the Transformer model, we can jointly model different data sources, maximize the value of each data type, and improve the model's predictive ability and robustness through multimodal learning[7].

Transformer-based multimodal financial risk prediction is not only highly feasible in theory, but also has broad significance in practical application scenarios. First, the fusion of multimodal data can significantly improve the accuracy of financial risk prediction. In traditional financial prediction, models often rely on historical data for linear regression or time series analysis, while ignoring sudden changes in market sentiment or macroeconomic factors. By introducing text data such as news and policy changes, the Transformer model can quickly capture potential risk signals in the market and make predictions in combination with historical data. Secondly, this model is particularly outstanding in dealing with emergencies in the financial market. For example, the market's response to policy changes is often rapid and drastic. Traditional models cannot process this unstructured information in real-time, while Transformers can quickly identify and process these changes through attention mechanisms, helping financial institutions make more timely decisions.

In addition, Transformer-based multimodal learning can also improve the model's generalization ability and adaptability to complex environments. The financial market is a highly dynamic and complex system that is affected by a variety of internal and external factors. When faced with such a complex environment, traditional unimodal prediction models are often prone to overfitting a certain type of data and ignoring other important sources of information. The multimodal characteristics of the Transformer model enable it to extract useful information from different data sources and comprehensively evaluate the relevance and importance of this information, thus avoiding the singleness and limitations of the model. Especially in the globalized financial market, the rapid changes in policies and economic environments and the cross-border flow of financial information make prediction models based on multimodal data particularly important.

In summary, Transformer-based multimodal financial risk prediction provides a new direction for the analysis and prediction of financial markets. By integrating multiple data such as time series, financial reports, and news texts, the Transformer model can achieve more accurate risk identification and prediction in a complex financial environment. This approach can not only improve the prediction accuracy of the model, but also enhance its ability to respond to emergencies, providing financial institutions with more timely and comprehensive decision support in a rapidly changing market.

2. Method

In Transformer-based multimodal financial risk prediction, the key to method design is how to effectively integrate different types of data sources, such as text data, time series data, financial reports, etc., to improve the model's predictive ability. As the core component of multimodal learning, the Transformer model can effectively capture the global dependencies between data through the self-attention mechanism. In this method, we first need to preprocess and represent different data sources, and then fuse them through the Transformer model to finally complete the risk prediction task.

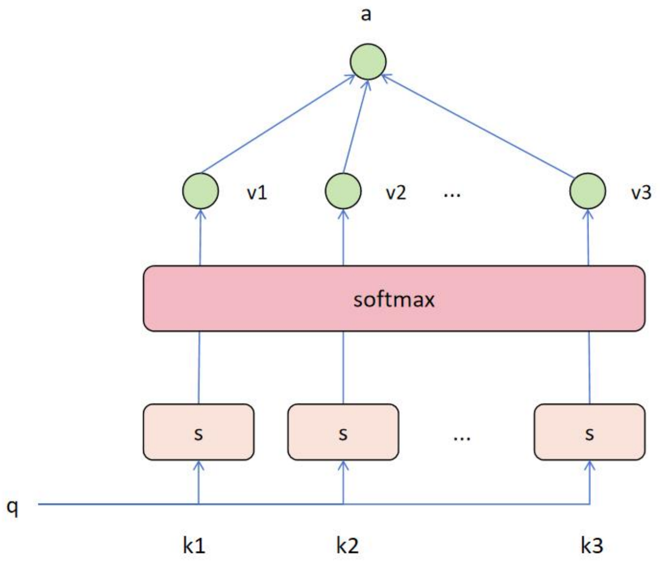

Figure 1. Attention mechanism in transformer

First, for time series data, we use embedding technology to represent it in vector form. This type of data usually has temporal order and trend. In order to capture this sequential information, we slice and normalize it appropriately and then input it into the Transformer model. Specifically, given a time series $X = \{x_1, x_2, ..., x_T\}$, we embed the input features of each time point into a high-dimensional space to obtain vector e_t ,which is then input into the encoder layer of the Transformer.

For text data, we use pre-trained language models (such as BERT or GPT) for representation learning. After each piece of text is embedded, it is converted into a vector sequence $\{w_1, w_2, ..., w_n\}$, where w_i is the vector representation of the i-th word. Transformer processes these vectors through the self-attention mechanism and calculates the correlation between them. The core of the self-attention mechanism is to calculate the attention weight of each word through the following formula:

$$Attention(Q, K, V) = soft\max(\frac{QK^T}{\sqrt{d_k}})V$$

Among them, Q, K, and V represent the query matrix, key matrix, and value matrix, respectively. In this way, the model can adaptively capture important words and information in the text and comprehensively consider their potential impact on financial risks.

When processing financial report data, it is usually structured data, which can be directly converted into vector representation through the embedding layer. Each financial indicator can be regarded as a feature f_i, which can be represented as a vector F_i and input into the Transformer model together with other modal data for fusion processing. In this way, the model can not only capture the dynamic changes of the market through time series, but also understand the fundamentals of the company through financial reports.

In order to achieve weighted fusion of different modal data, we can use linear weighted summation to combine the data of each modality. The formula is:

$$Z_i = \alpha x_i + \beta w_i + \chi F_i$$

Z_i is the data after formaldehyde fusion, α , β , and χ are all weight hyperparameters. The fused data contains valid information on all modal data. Next, a feedforward neural network is used to further process the representation and finally output a financial risk prediction result. Through this multimodal fusion method, the Transformer model can effectively integrate data from different sources and provide more comprehensive support for financial risk prediction.

3. Experiment

3.1 Datasets

The data sets used in this article are as follows: First, the time series data comes from financial indicators such as stock prices, trading volume and volatility in the market. The data comes from Yahoo Finance and covers daily trading records in the past five years. Secondly, the text data includes financial news, market reports and policy announcements, which come from financial news websites and industry report databases. The text features are extracted through the pre-trained BERT model. The data range includes major market events in the past five years. Finally, the financial data uses the financial report data of many listed companies, such as balance sheets, income statements, etc. The data comes from the Standard & Poor's and Bloomberg databases, mainly including the company's financial status in the past five years, reflecting its key indicators such as profitability, debt level and cash flow.

3.2 Experimental Results

In this experiment, we used four deep learning models as comparison models: GRU (Gated Recurrent Unit), BiLSTM (Bidirectional LSTM), TCN (Temporal Convolutional Network), and Attention-based RNN. These models were selected because they can process time series data or capture long-term dependencies in financial risks, and perform well in financial risk prediction tasks. In order to better evaluate the performance of the model in risk prediction, we selected AUC (area under the ROC curve) and Recall as evaluation indicators. AUC can effectively measure the model's ability to distinguish risk events and ensure that the model can accurately identify high-risk customers or events; Recall is used to evaluate the model's ability to capture risk events in actual predictions to avoid missing important high-risk cases. These two indicators are particularly critical for financial risk prediction tasks.

Table 1: Experimental Results

Model	AUC	Recall
GRU	0.81	0.70
BILSTM	0.84	0.72
TCN	0.86	0.75
Attention-based RNN	0.87	0.76
Transformer (ours)	0.91	0.81

From the experimental results, we can see that the Transformer model performs much better than other comparison models in the task of financial risk prediction, especially in terms of AUC and Recall. First, the GRU model has an AUC of 0.81 and a Recall of 0.70. Although it can effectively process time series data, its performance is relatively weak. As a simplified version of the RNN model, GRU can capture some dependencies in financial time series data, but its limitations are more obvious when facing the complexity and nonlinear characteristics of the financial market. Since financial risk prediction involves the integration of multimodal data, such as market news and corporate financial status, GRU lacks sufficient flexibility in processing these heterogeneous data, so it does not perform as expected in identifying and predicting high-risk events.

The performance of the BiLSTM and TCN models has improved. The AUC of BiLSTM is 0.84 and the Recall is 0.72, indicating that the bidirectional LSTM can capture the dependencies of financial data from different time directions and has certain advantages in processing asymmetric information. However, although BiLSTM can better capture information in time series, it still faces challenges in processing the integration of multimodal data, especially when dealing with text data and complex market fluctuations. TCN processes time series data through convolutional layers, with an AUC of 0.86 and a Recall of 0.75, showing that it is stronger than recurrent neural networks in capturing long-term dependencies and trends. TCN's advantages in parallel processing and long-term dependency capture make it more efficient when facing complex financial time series data, but its adaptability in processing text and financial data is not as good as more flexible models such as Transformer.

The AUC of the Attention-based RNN model reaches 0.87 and the Recall is 0.76, further demonstrating the powerful effect of the attention mechanism in financial risk prediction. By adaptively assigning attention weights to different modal data, the Attention-based RNN can better identify important financial risk signals, especially the ability to flexibly handle the interaction between text and time series data. Compared with the traditional RNN model, the introduction of the Attention mechanism improves the model's sensitivity to key market events, ensuring that the model can quickly capture high-risk events when dealing with complex financial market dynamics. However, despite its superior performance to GRU, BiLSTM, and TCN, Attention-based RNN is still slightly inferior to Transformer in multimodal data fusion.

In the end, the Transformer model is significantly ahead of other models with AUC 0.91 and Recall 0.81. Through its self-attention mechanism, the Transformer model is able to establish complex dependencies between multimodal information such as time series, text, and financial data, and effectively capture long-range dependencies. Compared with other models, Transformer not only performs well in multimodal data fusion, but also performs particularly well in complex market environments. Its higher AUC and Recall indicate that the model can effectively avoid underreporting risk events while accurately predicting high-risk financial events, providing financial institutions with more accurate risk identification tools. This also proves the strong potential of the Transformer model in the financial field, especially in complex tasks that require simultaneous processing of multi-dimensional information, where its advantages are more obvious.

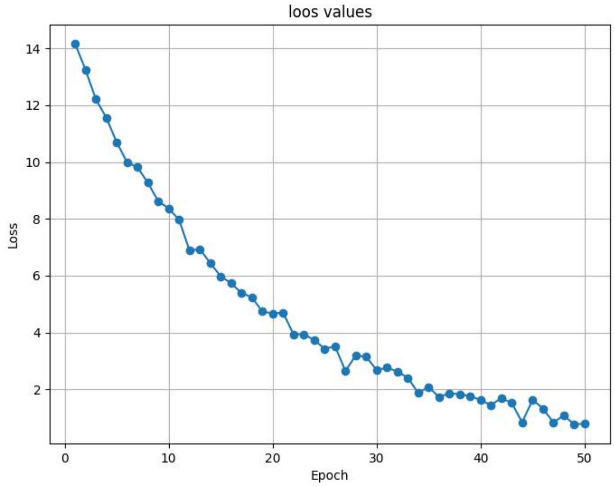

Figure 2. Loss Function Decrease Chart

We also give a graph of the loss function drop during training, as shown in Figure 2. In addition, this paper also conducted comparative experiments on different multimodal fusion strategies, and the results are shown in Table 2.

Table 2: Comparative experimental results of different multimodal fusion strategies

Strategies	**AUC**	**Recall**
Time series	0.79	0.71
Text data	0.85	0.74
Financial data	0.87	0.76
Time series+Text data	0.90	0.80
Time series+Financial data	0.89	0.79
Text data+Financial data	0.90	0.81
All	0.91	0.81

From the experimental results, different multimodal fusion strategies have a significant impact on the performance of financial risk prediction tasks. The effect of single-modal data is relatively weak, among which the AUC of time series data is 0.79 and the Recall is 0.71, indicating that when relying only on time series data such as historical prices and trading volumes in the market for prediction, the model has limitations in identifying risk events. In contrast, text data (AUC is 0.85, Recall is 0.74) and financial data

(AUC is 0.87, Recall is 0.76) perform much better, especially financial data can provide more stable company fundamentals information, which helps to identify potential high-risk events. Although text data cannot provide specific quantitative information, it can capture external factors such as market sentiment and news events, which is also extremely important for risk prediction.

When the multimodal fusion strategy is adopted, the model performance is significantly improved. In particular, after combining text data with financial data, the AUC reaches 0.90 and the Recall is increased to 0.81, indicating that the combination of multimodal data can help the model capture complex signals in the financial market more comprehensively. When all data are combined, AUC reaches 0.91 and Recall remains at 0.81. Although the improvement is limited compared to some dual-modal fusion, the model performance is most stable after combining data from all modes. This result shows that integrating information from multiple data sources can make up for the shortcomings of single data and significantly improve the model's prediction accuracy and risk identification ability, which fully demonstrates the potential of multimodal fusion in financial risk prediction.

4 . Conclusion

The experimental results of Transformer-based multimodal financial risk prediction show that the fusion of different types of data significantly improves the performance of the model. By introducing time series, text data and financial data, the model can more comprehensively capture the complex signals in the financial market, especially in dealing with market fluctuations and identifying high-risk events. The experiment shows that although single-modal data can provide effective information to a certain extent, its prediction accuracy and risk identification ability are limited. Through multimodal data fusion, especially the combination of text and financial data, the AUC and Recall indicators of the model are greatly improved, fully demonstrating the potential and importance of multimodal fusion for financial risk prediction. With its self-attention mechanism, the Transformer model can effectively process multimodal information, capture long-range dependencies and complex associations in the data, and enable it to have stronger risk identification capabilities in the financial market. In the future, with the continuous advancement of multimodal learning technology, financial risk prediction based on Transformer will hopefully provide more accurate and comprehensive decision support for financial institutions.

References

[1] Z.W. Zhang, J.N. Wang: Crane Design Manual (China Railway Press, China 1998), p.683-685. (In Chinese)

[2] Wang K, Gu T, Du X. Evaluation of multimodal data-driven financial risk prediction methods for corporate green credit[J]. Journal of Intelligent & Fuzzy Systems (Preprint): 1-13.

[3] Zhang W, Peng L. Semantic Analysis and Image Processing-Jointly Driven Multimodal Deep Learning Framework for Smart Warning of Enterprise Financial Risks[J]. Journal of Circuits, Systems and Computers, 2024.

[4] JIN X, Lin S L. An Early Prediction Model on Systemic Risk Under Global Risk: Using Finbert and Temporal Fusion Transformer to Multimodal Data Fusion Framework[J]. Available at SSRN 4706654.

[5] Ang G, Lim E P. Temporal Implicit Multimodal Networks for Investment and Risk Management[J]. ACM Transactions on Intelligent Systems and Technology, 2024, 15(2): 1-25.

[6] Jain S, Chhabra P, Neerkaje A T, et al. Saliency-Aware Interpolative Augmentation for Multimodal Financial Prediction[C]//Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). 2024: 14285-14297.

[7] Jun T, Wanting Q, Qingtao P, et al. A deep multimodal fusion and multitasking trajectory prediction model for typhoon trajectory prediction to reduce flight scheduling cancellation[J]. Journal of Systems Engineering and Electronics, 2024.

[8] Rai B K, Jain I, Tiwari B, et al. Multimodal mental state analysis[J]. Health Services and Outcomes Research Methodology, 2024: 1-28.