Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.30611

A Dynamic Prediction Model for Business Cycle Value Based on Multimodal Data Fusion and Deep Learning

Kaixin Tian

FedEx. Pittsburgh, Pittsburgh, USA
tiankaixin2001@163.com

Abstract. The major issues with business cycle value prediction are that the data sources are heterogeneous, complex spatiotemporal dependencies, and a lack of long-term prediction accuracy. The model of dynamic prediction, which is developed in this paper, is founded on the fusion of multimodal data and deep learning and incorporates four kinds of heterogeneous data, financial indicators, market transactions, macroeconomic data, and textual sentiment. The former applies a Temporal Convolutional Network (TCN) to learn the temporal features of each modality and a Multi-Head Attention (MHA) mechanism to learn cross-modal interaction relationships. Second, gated fusion unit is intended to weigh and fuse heterogeneous features dynamically and Bidirectional Long Short-Term Memory (BiLSTM) network is employed to model cycle evolution law. Lastly, residual connections with Dropout regularization are also added to enhance the generalization capacity of the model. The model on an empirical dataset of 3500 companies in the 2010-2024 period had a mean absolute percentage error (MAPE) of 6.2% and root mean square error (RMSE) of 0.043 in the expansion phase and 8.7% in the contraction phase, giving it a useful tool and theoretical foundation on dynamic monitoring of business cycle and corporate decision-making.

Keywords: Multimodal Fusion, Business Cycle Prediction, Temporal Convolutional Networks, Attention Mechanism, Deep Learning

1. Introduction

Corporate valuation must consider four sources of data together: the quality of operations as indicated by the financial statements, the expectations of investors as indicated by the operations of the market, the external constraints that are imposed by the macro environment and the emotional appeals caused by the mass opinion. These four sources of data vary widely in the time granularity, information density and representation. The successful combination of the flow of heterogeneous information and the derivation of the evolution patterns of the business cycles has turned out to be a fundamental bottleneck that limits the scientific context of the corporate strategic decision-making and the accuracy of the capital market risk management.

The technical direction of solving the cycle prediction problem is represented by multimodal data fusion. It is valuing the implicit correlations not considered by the traditional single-perspective methods because of cross-modal interaction. A historical transformation in financial indicators

DOI: 10.54254/2755-2721/2025.30611

indicates endogenous growth momentum of a company; capital flows and expectations of value reflect market transaction behavior; macroeconomic fluctuations will indicate the boundary circumstances of the functioning of the cycle; and textual public opinion releases signals of changes in market sentiment in advance. The time dimension sequential correspondence of these four modalities and their corroboration to one another constitute an entire Information chain of cycle phase transitions. In deep learning architectures, attention mechanisms can dynamically learn the value of various modalities involved in specific stages of the cycle, gating units are able to perform dynamically fuse heterogeneous features and recurrent neural networks are able to capture temporal dependencies. This is an organic integration of these technical elements that forms the basis of the construction of an end-to-end cyclic prediction system.

The innovations presented in this paper manifest themselves in the three breakthroughs: First, it suggests a cross-modal feature extraction architecture that comprises multiple heads of temporal convolutional networks and multi-head attention mechanisms, resolving two issues: First, temporal alignment and semantic mapping of information between samples of different frequencies; second, gating fusion units to achieve adaptive weighting of heterogeneous features to avoid information redundancy and intermodal interference due to a straightforward concatenation strategy; and third, bidirectional evolutionary dependency of the cycle is modeled by the joint use of bidirectional long short-

2. Related work

The theoretical exploration of commercial value prediction involves multiple application areas, from the pricing mechanism of digital platforms to the industrial transformation of emerging technologies. The value assessment methods under different scenarios show significant differences, but most focus on local problems in specific industries. Lv and Li [1] took the commercial advertising price of UP masters as the prediction target, which shows strong generalization ability and good practicality. The entropy weight method is used to weight eight characteristics of content creators, and three indicators are formed: exposure, recognition, and discussion. Then, the multilayer perceptron regression algorithm is used to predict the advertising price. Liu et al. [2] revealed the mechanism of multi-subject interactive behavior on the realization of commercial value of medical consultation products, identified the relationship between price and value co-creation, and proposed a sustainable commercial value realization model for online medical platforms, providing theoretical basis and practical guidance for online medical platforms to improve bilateral user participation and corporate performance. Wang [3] believed that the new consumption scenario of front warehouse delivery based on the original commodity pool may face the problem of whether the applicability and efficiency are appropriate. Wu [4] focused on the innovation of operation mode and value realization path in platform economic ecosystem, and proposes an innovation strategy from four dimensions: data empowerment, resource sharing, ecological collaboration and technological breakthrough. Zhang [5] analyzed the main characteristics of intellectual property and the key points of intellectual property management, and analyzed the current problems from the three aspects of intellectual property management awareness, management system and management personnel, and then proposed practical paths, hoping to help enterprises with intellectual property management. The research of Luo et al. [6] is expected to provide valuable information for the public and organizations, further promote the application and industrial transformation of brain-computer interface technology, enhance commercial value, so as to better serve mankind. Donnelly et al. [7] reviewed the claims data of all insured persons of a value-oriented contract signed between a commercial insurance company and a healthcare system in 2021 and 2022. Enholm et al. [8] aimed

DOI: 10.54254/2755-2721/2025.30611

to explain how organizations can use artificial intelligence technology in their operations and clarify their value creation mechanism through a systematic literature review. Gellweiler and Krishnamurthi [9] proposed an IT value integration definition with two complementary dimensions based on template analysis. Paukstadt and Becker [10] reviewed relevant literature from the perspective of information systems, assessed the current status of smart energy business model research, identified relevant business model types, and proposed a research agenda for future research on energy business models based on the Internet of Things. Existing literature has made some progress in the field of business value prediction, but there are still three obvious limitations: first, the data source is singular, resulting in incomplete information dimensions; second, there is a lack of dynamic modeling capabilities across different cycle stages; and third, there is a lack of systematic design for the fusion mechanism of heterogeneous data.

3. Method

3.1. Multimodal data acquisition and preprocessing framework

This paper constructs a four-dimensional data system encompassing financial indicators, market transactions, macroeconomics, and textual sentiment. The financial modality extracts 27 indicators from corporate annual reports, including revenue growth rate, debt-to-equity ratio, and cash flow ratio, and samples them quarterly to form time series. The market modality takes daily trading information as a package, which has 15 features including volatility of stock price, rate of change of volume of transactions, turnover rate. The macroeconomic modality integrates 12 economic variables including, among others, the GDP growth rate, CPI index and the rate of interest. The textual modality extracts semantic vectors of 816 feature clouds from news items, reports of analysts and comments in social channels based on BERT which is a pre-trained model according to the transformers of Google. In the data cleaning stage a Lagrange interpolation formula is implemented to fill the gaps of absent data. The outliers are obtained through standard deviation and these are corrected on the 3σbasis. The Z-score standardisation is made to remedy the defect of dimensionality of different modalities.

$$x_{norm}=\frac{x-\mu}{\sigma} \tag{1}$$

Time alignment adopts a sliding window mechanism to aggregate daily frequency data to quarterly granularity, ensuring consistency of timestamps for each modality. To capture the nonlinear characteristics of cyclical evolution, financial indicators and market data are processed by first-order difference and converted into growth rate sequences [11]. Text modality uses TF-IDF weighting scheme to extract keywords, and calculates polarity scores through sentiment analysis model, with a quantization interval of [-1,1]. When constructing the lag feature matrix, the time window length is set to T=12 quarters, and each sample contains multimodal observations of the current moment and the previous 11 moments. The final multimodal tensor dimension is:

$$X\in\mathbb{R}^{N\times T\times D} \tag{2}$$

N is the number of samples; T is the time step; and D is the total dimensionality of the fused features (59 dimensions). Table 1 shows the statistical description of the multimodal data features:

DOI: 10.54254/2755-2721/2025.30611

Table 1. Statistical description of multimodal data features

Modality Type	Feature Dimension	Sampling Frequency	Data Coverage Period	Missing Rate(%)
Financial Indicators	27→18	Quarterly	2010Q1-2024Q4	3.2
Market Trading	15→11	Daily→Quarterly	2010Q1-2024Q4	1.8
Macroeconomic	12→9	Quarterly	2010Q1-2024Q4	0.5
Textual Sentiment	768→21	Real-time→Quarterly	2012Q1-2024Q4	7.4
Fused Features	59	Quarterly	2010Q1-2024Q4	2.7

3.2. Temporal feature extraction module based on TCN

The temporal convolutional network architecture employs causal dilated convolutions to capture long-range temporal dependencies in business cycles. The network contains five convolutional layers with exponentially increasing dilation factors set to {1, 2, 4, 8, 16}, ensuring a receptive field covering 32 time steps, corresponding to 8 years of historical data. Each convolutional layer has a kernel size of 3, with channel numbers of {64, 128, 256, 256, 128}, ensuring progressively enhanced feature representation capabilities. The dilated convolution operation is defined as follows:

$$y(t)=\sum_{k=0}^{K-1} f(k)\cdot x(t-d\cdot k) \tag{3}$$

d represents the inflation rate; K represents the kernel size; and f represents the learnable weights. Data from each modality is independently input into the corresponding TCN branch. The financial modality processes quarterly sequences of financial indicators through an 18-dimensional input layer; the market modality captures prices and volume dynamics using an 11-dimensional input; the macro modality models economic environment changes using a 9-dimensional input; and the text modality encodes trending evolution of public opinion using a 21-dimensional input [12-13]. Each of the convolutional layers is followed by weight normalization and a ReLU activation function, and a Dropout ratio of 0.2 is set to avoid overfitting. The mechanism of residual connection makes it possible to convey the input directly to the output, solving the gradient vanishing problem. The residual block calculation formula is:

$$h^{(l+1)}=\mathscr{F}(h^{(l)},W^{(l)})+h^{(l)} \tag{4}$$

$\mathscr{F}$ represents the convolutional transformation function, and $W^{(l)}$ is the parameter matrix of the l-th layer. Four parallel TCN branches each output a 128-dimensional temporal feature vector, which are aggregated to form a 512-dimensional joint representation. This module ensures that predictions rely solely on historical information through causal constraints, thus avoiding data leakage.

3.3. Cross-modal attention fusion and dynamic weighting mechanism

A multi-head attention mechanism quantifies the interdependence strength among four modalities, capturing cross-modal synergistic effects during business cycle evolution. The four 128-dimensional feature vectors output by the TCN (financial, market, macro, and textual) are projected through three parallel linear transformation layers to form a query matrix Q, a key matrix K, and a value matrix V, with a projection dimension of 64. Attention calculation employs a scaled dot product approach, and the smoothness of the softmax distribution is adjusted using a temperature parameter $\sqrt{d_k}$.

DOI: 10.54254/2755-2721/2025.30611

$$\mathrm{Attention}(Q,K,V)=\mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{5}$$

d_k =64 is the dimension of the key vector. Eight attention heads are configured for parallel computation. The parameters of each head are learned independently, focusing on different feature subspaces such as the lag relationship between financial indicators and market fluctuations, the transmission effect of macro policies on corporate performance, and the coupling mode between text sentiment and price momentum. The output dimension of each head is 8×64=512. After being fused by splicing, it is mapped to a unified representation space through a fully connected layer. The gating fusion unit dynamically adjusts the contribution of each modality according to the attention aggregation result to avoid information loss caused by fixed weights. The gating vector is calculated by splicing attention features, original modality features and time step hidden states:

$$g_i=\sigma(W_g[h_{att};h_i;c_t]+b_g) \tag{6}$$

h_{att} represents the attention fusion feature; h_i represents the i-th modality feature; and c_t represents the hidden state at time step t. The four gate values are weighted and fused after softmax normalization. This mechanism allows the model to adaptively adjust the importance of each modality based on the stage of the cycle, learning the optimal weight allocation strategy through gradient backpropagation. Layer normalization is applied to stabilize the fused feature training process, and residual connections directly pass the original multimodal features to the output layer, preserving low-level semantic information. The entire fusion module has 0.8M parameters and outputs a 512-dimensional comprehensive feature vector for subsequent BiLSTM modeling.

3.4. BiLSTM periodic evolution modeling and prediction output

The Bidirectional Long Short-Term Memory (BiLSTM) network receives a fused 512-dimensional feature sequence and models the bidirectional time dependency of the business cycle through two independent forward and backward paths. The forward LSTM captures trend continuity from the past to the present, while the backward LSTM extracts prior information about cycle turning points from the future to the present. The number of hidden layer units is set to 256, and the forget gate controls the degree of retention of historical information.

The forward and backward hidden states are concatenated to form a 512-dimensional representation, and overfitting is suppressed by a dropout layer (dropout rate 0.3). The prediction head contains two fully connected layers: the first layer maps to 128 dimensions and uses ReLU activation, and the second layer outputs a four-dimensional vector corresponding to the probability distribution of four business cycle states: expansion, peak, contraction, and trough.

The model's loss function adopts a weighted combination of cross-entropy and mean squared error:

$$L=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{4}y_{ij}\log(\hat{y}_{ij})+\lambda\|v_i-\hat{v}_i\|_2^2 \tag{7}$$

λ=0.5 is used to balance the weights of classification and regression tasks, and vi is the periodicity value index. Gradient clipping (threshold 5.0) is used during training to prevent gradient explosion, and the learning rate is smoothly decayed from 0.001 to 0.0001 using a cosine annealing strategy. Table 2 shows the BiLSTM network architecture parameter configuration:

DOI: 10.54254/2755-2721/2025.30611

Table 2. BiLSTM network architecture parameter configuration

Network Layer	Unit Type	Hidden Dimension	Parameters(M)	Dropout Rate
Forward LSTM	LSTM	256	1.05	0.0
Backward LSTM	LSTM	256	1.05	0.0
Feature Fusion Layer	Concatenate	512	0.0	0.3
Fully Connected Layer 1	Dense+ReLU	128	0.066	0.2
Output Layer	Dense+Softmax	4	0.0005	0.0

4. Results and discussion

4.1. Experimental dataset construction and evaluation metric setting

This test selected complete business cycle data from 367 listed companies in the Chinese A-share market from the first quarter of 2010 to the fourth quarter of 2024, covering 12 industry sectors including manufacturing, financial services, information technology, and energy and chemicals. Financial modal data was extracted from quarterly financial reports in the Wind database; market modal data came from daily trading records on the JoinQuant quantitative platform; macroeconomic modal data was obtained from official releases by the National Bureau of Statistics; and textual modal data was constructed by crawling news reports and research commentaries from platforms such as Eastmoney.com and Sina Finance.

4.2. Comparative analysis of predictive performance

To evaluate the predictive robustness of the models across different business cycle phases, the test set was divided into four subsets based on the cycle phase: expansion (87 quarterly samples), peak (34 quarterly samples), contraction (62 quarterly samples), and trough (29 quarterly samples). The prediction error index for each model was calculated for each subset. The expansion phase test samples mainly focused on the supply-side reform period of 2015-2016 and the post-pandemic recovery phase of 2020-2021. The peak phase samples corresponded to the cycle highs of the third quarter of 2017 and the second quarter of 2023. The contraction phase covered the period of Sino-US trade friction in 2018 and the recurring pandemic in 2022. The trough phase included the low point of the pandemic impact in the first quarter of 2020. Figures 1 and 2 show the prediction errors for different industry sectors during the contraction phase, respectively.

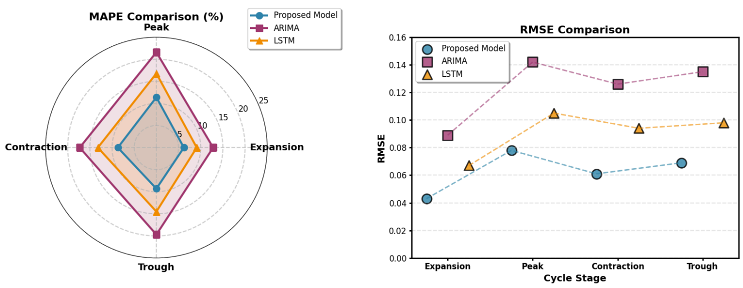

Figure 1. Error comparison

DOI: 10.54254/2755-2721/2025.30611

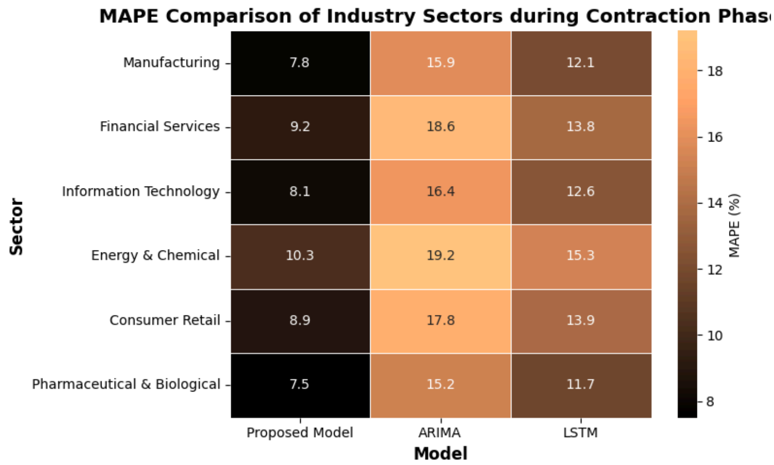

Figure 2. Comparison of forecast errors for various industry sectors during the contraction period

The multimodal fusion model proposed in this paper achieves a mean absolute percentage error (MAP) of 6.2% and a root mean square error (RMSE) of 0.043 during the expansion phase, significantly outperforming the ARIMA model's 12.8% and 0.089, and the LSTM single-modal model's 9.1% and 0.067. During the contraction phase, the MAPE of the three methods are 8.7%, 17.3%, and 13.2%, respectively, with RMSEs of 0.061, 0.126, and 0.094.

5. Conclusion

This paper overcomes the limitations of traditional business cycle prediction relying on a single data source by employing multimodal data fusion and deep learning techniques, demonstrating the effectiveness of cross-modal collaborative modeling in capturing the evolutionary patterns of cycles. The combination of temporal convolutional networks and multi-head attention mechanisms successfully solves the problem of spatiotemporal alignment of heterogeneous data, while gated fusion units achieve adaptive weight allocation for different information sources. Bidirectional long short-term memory networks effectively model the transition logic between cycle stages. The model outperforms traditional statistical methods and single-modal deep learning in both expansion and contraction phases, providing a quantifiable risk warning mechanism for corporate strategic adjustments and investment decisions.

References

[1] Lü Huimin, Li Xin. A Study on the Commercial Value of Online Video Content Creators - Taking Bilibili UP Masters as an Example [J]. Journal of Liaoning University of Technology (Natural Science Edition), 2024, 44(3): 164-168
[2] Liu Kan, Zhou Hongrui, Jin Xin, Sun Xueying. How to Realize the Value Co-creation Results of Online Medical Platforms? - A Study on Multi-Subject Interactive Behavior and the Realization of Commercial Value of Medical Consultation Products [J]. Management Review, 2025, 37(7): 154-165
[3] Wang Zhishan. The Commercial Value of Pre-positioned Warehouses as Seen from the Battle of Beijing [J]. Business Observation, 2025, 11(20): 9-11
[4] Wu Biyi. A Study on the New Operation Strategies and Commercial Value of Platform Economy [J]. China Business Review, 2025, 34(7): 114-117
[5] Zhang Tongxin. A Study on the Construction of Enterprise Intellectual Property Management System under the Demand of Commercial Value [J]. Market Weekly, 2025, 38(25): 154-157

DOI: 10.54254/2755-2721/2025.30611

[6] Luo J, Ding P, Gong A, et al. Applications, industrial transformation and commercial value of brain-computer interface technology [J]. Sheng wu yi xue Gong Cheng xue za zhi= Journal of Biomedical Engineering= Shengwu Yixue Gongchengxue Zazhi, 2022, 39(2): 405-415.
[7] Donnelly L F, Dellva B P, Jarmul J A, et al. Evaluation of claims data from a commercial value-based insurance product shows pediatric imaging is not a major driver of overall or pediatric healthcare expenditures [J]. Pediatric Radiology, 2024, 54(5): 842-848.
[8] Enholm I M, Papagiannidis E, Mikalef P, et al. Artificial intelligence and business value: A literature review [J]. Information systems frontiers, 2022, 24(5): 1709-1734.
[9] Gellweiler C, Krishnamurthi L. IT business value and competitive advantage: Integrating a customer-based view [J]. Information Systems Management, 2022, 39(4): 363-385.
[10] Paukstadt U, Becker J. Uncovering the business value of the internet of things in the energy domain–a review of smart energy business models [J]. Electronic Markets, 2021, 31(1): 51-66.
[11] Zhao F, Zhang C, Geng B. Deep multimodal data fusion [J]. ACM computing surveys, 2024, 56(9): 1-36.
[12] Steyaert S, Pizurica M, Nagaraj D, et al. Multimodal data fusion for cancer biomarker discovery with deep learning [J]. Nature machine intelligence, 2023, 5(4): 351-362.