

Research on Financial Time Series Prediction Model Based on Multi Attention Mechanism and Emotional Feature Fusion

Jingzhi Yin

Department of Mathematics, Columbia University, New York 10017, New York, United States

Keywords: Dynamic price prediction, time series modeling, overseas warehouse location selection, customer satisfaction, SARIMA-GARCH model

Abstract: Financial time series forecasting is crucial for high-quality economic development, but data has nonlinear, non-stationary, high noise, and multi factor driving characteristics. Traditional models (such as ARIMA) and machine learning models have limitations in complex scenarios. This study constructs a dual model system: the long-term prediction model (PatchCT) captures temporal context dependencies through sparse attention and integrates cross feature trend information through channel attention branches; The short-term prediction model (MSA xLSTM) integrates FinBERT emotion index and technical indicators (filtered by Spearman), achieves feature fusion through multi-scale emotion attention module, and combines xLSTM to establish long-range dependencies to improve short-term accuracy. The data processing adopts RevIN stationary sequence and adaptive patching to enhance local feature extraction. The backtesting verification shows that the long-term prediction strategy returns are better than the baseline, and the short-term model achieves the lowest prediction error on all four stock datasets (such as RMSE/MAE indicators being better than CNN-LSTM, GRU Attention, and other models), providing better decision support for aggressive ultra short term investors. This study forms a "long-term short-term" collaborative financial time series prediction system, promoting the practical application of the model in asset allocation, risk management and other scenarios. In the future, it will expand the joint modeling of time domain frequency domain, explore cross domain generalization ability, and deepen the analysis of multiple factors related to weak fundamental stocks, promoting the evolution of financial time series prediction towards a more universal and accurate direction.

1. Introduction

As the core of the modern economic system, the stable operation of the financial market is crucial for high-quality economic development, covering multiple types of markets such as stocks, currencies, gold, bonds, etc. , and each market influences and dynamically changes with each other. The government needs to grasp market laws to formulate scientific policies to prevent systemic risks, while investors rely on accurate predictions to optimize asset allocation and decision-making. However, financial time series data has the characteristics of non stationarity, non-linearity, high noise, and multi factor driving (such as macroeconomic policies, social public opinion, emergencies, etc.). Traditional statistical models (such as ARIMA) are difficult to adapt to complex financial

scenarios because they can only capture linear relationships and rely on strict assumptions; Although machine learning models improve prediction accuracy, there are problems such as low computational efficiency and insufficient non-linear feature extraction in large-scale high-dimensional data.

With the development of deep learning, models such as CNN, RNN (such as LSTM, GRU), and Transformer have been introduced into financial prediction due to their strong fitting ability and feature extraction advantages. However, existing research still faces challenges: CNN based models are prone to losing long-term dependencies and detail information during the process of time to image conversion; Although RNN and its variants are suitable for sequence modeling, they face the risks of gradient vanishing/exploding and long-term dependency loss; Although Transformer improves feature extraction efficiency through attention mechanism, there is still room for improvement in capturing cross dimensional dependencies and addressing feature singularity issues in financial time series. In addition, existing methods for integrating external information such as market sentiment mostly remain at the feature input level, without delving into the multi-scale impact mechanism of sentiment on stock prices. The motivation for this study stems from the practical need to improve the accuracy and stability of financial time series prediction, with the goal of building a deep learning model that integrates multiple attention mechanisms and emotional features. Specific contributions include: proposing a long-term prediction model PatchCT that combines sparse attention and channel attention, and enhancing the capture of inter variable and positional dependencies through dynamic patch partitioning and improved channel attention module (RECA); Design a short-term prediction model MSA xLSTM that integrates emotional attention and extended LSTM (xLSTM), combining news/stock review emotional information and technical features (filtered by Spearman correlation coefficient) to achieve multi-scale emotional impact modeling; By verifying the effectiveness of the model through public datasets and real index backtesting, the long-term prediction strategy returns are better than the baseline, and short-term predictions provide better decision support for aggressive investors, ultimately forming a "long-term short-term" collaborative financial time series prediction system, promoting the practical application of the model in asset allocation, risk management, and other scenarios.

2. Correlation theory

2. 1. Types and Applications of Attention Mechanism Principles

The attention mechanism 错误!未找到引用源。 was inspired by the focusing characteristics of specific information in human cognition. It was first proposed in 2017 and has been widely used in language modeling, machine translation, computer vision, and time series analysis. Its core lies in dynamically allocating weights to different positions in the input sequence, allowing the model to focus on key information to enhance task processing capabilities. Specific types include dot product attention (calculating similarity through dot product of query vector and key vector, weighted sum with value vector after scaling),

$$Attention(Q, K, V) = softmax(\frac{QK^t}{\sqrt{d_k}})V \quad (1)$$

multi head attention (parallel multiple independent attention modules, capturing semantic relationships from different subspaces and integrating them for output),

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^0 \quad (2)$$

sparse attention (only calculating the similarity between some key queries and key vectors to reduce computational overhead, suitable for long sequence data), and channel attention ECA

(compressing spatial features through global average pooling, using adaptive convolution kernels to capture cross channel information, and improving parameter efficiency). These mechanisms enhance the adaptability and robustness of the model to input sequences of different lengths and structures through flexible weight allocation and feature extraction capabilities, becoming core components in deep learning tasks.

2. 2. Principles and Applications of xLSTM and RevIN Technologies

XLSTM 错误!未找到引用源。 was proposed by Beck et al. and is an extended variant of the traditional LSTM framework, which improves sequence modeling performance by introducing exponential gating and new memory structures. It includes two variants: sLSTM 错误!未找到引用源。 improves the input gate and forget gate through exponential gating (such as using sigmoid or exponential functions to control information updates for forget gates), enhances the ability to retain key information, and introduces normalized states to improve numerical stability; MLSTM adds matrix storage units (key value matrix) and covariance update mechanism on the basis of sLSTM, expanding the memory storage scale and optimizing information processing. RevIN 错误!未找到引用源。 (Reversible Instance Normalization) was proposed by Kim et al. as a method for stationary sequence data. It normalizes the input sequence using learnable affine parameters through instance normalization (linear transformation after calculating mean and variance), and restores the original distribution by inverse normalization at the output, effectively solving the problem of distribution drift in financial time series. These technologies enhance the processing capability and robustness of models for complex sequence data by improving gating mechanisms, memory structures, and data preprocessing methods, and have important application value in tasks such as financial time series prediction.

3. Research method

3. 1. PatchCT model and optimization of long-term financial time series prediction

PatchCT 错误!未找到引用源。 is a long-term financial time series prediction model with a dual channel complementary structure, aimed at improving prediction accuracy and robustness. The input of the model is a two-dimensional matrix with a historical length of L and C -dimensional features. RevIN smoothing processing (normalization and inverse normalization) is used to reduce the impact of sequence non stationarity. Subsequently, a dynamic patching strategy was adopted to divide the sequence: the patch length was dynamically calculated based on the review window time step and dataset type, with a maximum value of 16 and a step size of half the patch length, ensuring that the sequence was divided into overlapping patch blocks to enhance local semantic capture capability. The model body consists of two complementary branches: channel attention and MLP branch. The improved RECA module extracts the dependency relationships between variables. RECA uses reparameterized convolution to establish dense connections between channels, and combines the GeLU activation and gating mechanism of the MLP layer to capture local features; The Transformer branch records time point position information through two-dimensional absolute position encoding, and combines sparse self attention mechanism to reduce computational complexity and extract global features in the time dimension. The final prediction is achieved through dual channel output weighted fusion, with weights adaptively adjusted in backpropagation to enhance the model's adaptability to different financial sequences. This structure effectively solves the pain points of traditional models in variable dependency capture, long-term dependency extraction, and sequence information loss, and improves the performance of long-term financial

time series prediction.

3. 2. Experimental verification and performance analysis of PatchCT model

This article verifies the effectiveness of the PatchCT model in long-term financial time series prediction through multiple experiments. The dataset uses CSI300, ChiNext100, and exchange, where exchange covers exchange rate data from eight countries from 1990 to 2016, and CSI300 and ChiNext100 contain data from 1214 trading days from 2019 to 2023. The features include opening price, highest price, lowest price, trading volume, closing price, and closing price of three weighted stocks. The data preprocessing is standardized using Z-score [错误!未找到引用源。](#), and the experimental environment is PyTorch framework, NVIDIA RTX4090 GPU, and 32GB RAM. The optimizer is AdamW (learning rate of 0. 001), batch size is 16, training set: validation set: test set=7:1:2, and hyperparameter settings are shown in Table 1 ($\gamma=4$, $K=3$, $n=5$, $\text{head}=16$, $D=256$).

Table 1 Examples of Partial Financial Time Series Data

Date	Open	High	Low	Close	Volume	Feature1	Feature2	Feature3
1/2/2019	3017. 06	3018. 77	2958. 48	2969. 53	686630. 21	598. 98	7. 16	55. 18
1/2/2019	2963. 01	3000. 44	2953. 25	2964. 84	708671. 03	590. 00	7. 21	55. 68
1/2/2019	2940. 19	3036. 81	2935. 83	3035. 87	1033189. 72	602. 00	7. 29	56. 59
1/2/2019	3055. 15	3061. 74	3035. 91	3054. 30	1011643. 59	605. 49	7. 32	56. 30

The evaluation indicators used were MSE, RMSE, and MAE. Comparative experiments showed that in multiple prediction windows ($T1=\{10, 20, 40, 60\}$, $T2=\{24, 48, 96\}$) of CSI300, ChiNext100, and exchange, the error of PatchCT was significantly lower than that of Informer

Table2 Long-Term Financial Prediction Model Performance Across Datasets & Windows

Dataset	Prediction Window	Model	Metric	MSE	RMSE	MAE
CSI300	10	Proposed Method		0. 018	0. 132	0. 103
		Informer		0. 050	0. 225	0. 117
		Autoformer		0. 026	0. 162	0. 127
		PatchTST		0. 018	0. 134	0. 104
		ConvTimenet		0. 020	0. 141	0. 109
		iTransformer		0. 017	0. 130	0. 101
	20	Proposed Method		0. 028	0. 167	0. 133
		Informer		0. 061	0. 247	0. 195
		Autoformer		0. 040	0. 201	0. 158
		PatchTST		0. 031	0. 176	0. 139
		ConvTimenet		0. 031	0. 176	0. 142
		iTransformer		0. 033	0. 182	0. 144
	40	Proposed Method		0. 047	0. 218	0. 177
		Informer		0. 115	0. 338	0. 270
		Autoformer		0. 129	0. 357	0. 283
		PatchTST		0. 052	0. 228	0. 183
		ConvTimenet		0. 050	0. 224	0. 184
		iTransformer		0. 049	0. 221	0. 180
	60	Proposed Method		0. 068	0. 262	0. 211
		Informer		0. 093	0. 305	0. 251

Dataset	Prediction Window	Model	Metric	MSE	RMSE	MAE
Chinet100	10	Autoformer		0.148	0.386	0.316
		PatchTST		0.078	0.279	0.226
		ConvTimenet		0.072	0.268	0.219
		iTransformer		0.083	0.288	0.232
		Proposed Method		0.009	0.093	0.073
		Informer		0.029	0.170	0.133
	20	Autoformer		0.015	0.122	0.099
		PatchTST		0.009	0.095	0.073
		ConvTimenet		0.009	0.095	0.074
		iTransformer		0.009	0.096	0.076
		Proposed Method		0.017	0.130	0.103
		Informer		0.054	0.232	0.191
	40	Autoformer		0.050	0.224	0.179
		PatchTST		0.018	0.134	0.104
		ConvTimenet		0.018	0.134	0.108
		iTransformer		0.019	0.138	0.108
		Proposed Method		0.040	0.199	0.152
		Informer		0.102	0.319	0.248
	60	Autoformer		0.059	0.243	0.195
		PatchTST		0.045	0.212	0.170
		ConvTimenet		0.032	0.179	0.139
		iTransformer		0.031	0.176	0.135
		Proposed Method		0.050	0.224	0.183
		Informer		0.113	0.336	0.279
Exchange	24	Autoformer		0.092	0.303	0.249
		PatchTST		0.057	0.239	0.189
		ConvTimenet		0.052	0.228	0.190
		iTransformer		0.069	0.261	0.203
		Proposed Method		0.025	0.159	0.119
		Informer		0.431	0.657	0.545
	48	Autoformer		0.031	0.175	0.135
		PatchTST		0.024	0.154	0.120
		ConvTimenet		0.024	0.155	0.120
		iTransformer		0.026	0.163	0.123
		Proposed Method		0.048	0.219	0.166
		Informer		0.782	0.886	0.766
	96	Autoformer		0.070	0.226	0.203
		PatchTST		0.051	0.226	0.172
		ConvTimenet		0.052	0.228	0.169
		iTransformer		0.054	0.232	0.178
		Proposed Method		0.091	0.301	0.226
		Informer		0.594	0.771	0.660
		Autoformer		0.146	0.381	0.302
		PatchTST		0.104	0.322	0.238
		ConvTimenet		0.098	0.313	0.237
		iTransformer		0.129	0.359	0.265

The ablation experiment confirmed the key role of dual branch structure, RECA module, and

sparse attention mechanism in improving performance. The application case has been verified through backtesting, and on CSI300 and ChiNext100, the cumulative return rate of the prediction based fixed investment strategy is better than the baseline strategy, especially in the stage of local rapid rise and fall.

3. 3. Application verification and summary of long-term financial time series prediction models

This article verifies the practical application value of the model through backtesting experiments: using a fixed investment strategy to simulate trading scenarios on financial indices (such as CSI300, Chinet100) and public exchange rate datasets (exchange), with cumulative return as the evaluation indicator. The results show that the improved strategy based on prediction can buy at a lower price during the local rapid rise and fall stages, effectively lowering the average cost. It still shows better return performance than the baseline (daily fixed investment) in the downward trend of the test set, especially during periods of severe fluctuations in net worth (such as the 15th to 30th nodes of the ChiNext 100 and the 130-145th nodes of the Shanghai and Shenzhen 300), demonstrating the advantage of trend judgment. The limitations of the model are reflected in the limited improvement during the local fluctuation stage, and other methods need to be combined to assist in analysis. The model proposed in this article, which combines adaptive patch encoding with a dual branch structure, enhances the ability to capture the dependency relationship between time and channel dimensions by introducing financial sequence smoothing methods and an improved RECA module. On both real and publicly available datasets, its mean square error, root mean square error, and mean absolute error are generally lower than those of comparative models such as Informer and Autoformer, verifying its effectiveness in long-term prediction tasks. The actual case further proves that this model can provide valuable decision-making references for medium and long-term investors, with both theoretical optimization and practical application value.

4. Results and discussion

4. 1. Research on the Model of Emotion Index and Multi feature Fusion in Short term Stock Prediction

In the financial market, there are significant differences in investor behavior, with some investors preferring short-term trading, especially for individual stock trends that are more complex due to irrational investment behavior, requiring higher accuracy in short-term forecasting than in long-term forecasting. To improve the accuracy of short-term forecasting, this chapter combines the investor sentiment index, uses the sentiment attention module to achieve multi feature fusion, and introduces the xLSTM network to construct a stock price prediction model. By combining multi-scale attention and emotional attention modules, the model can simultaneously capture the temporal dependence of technical information (such as highly correlated technical indicators filtered by Spearman correlation coefficient) and emotional information (emotional index calculated by FinBERT model for news and stock evaluation data). The problem definition focuses on using a historical sequence of length L (including C -dimensional features) to predict the closing price of the next trading day, and mapping the feature sequence to the target value through a time series prediction model. Although existing research has focused on the application of emotional features, there is still room for improvement in terms of feature fusion timeliness, technical indicator screening, and model structure design. The MSA xLSTM model proposed in this chapter has been validated for its effectiveness on multiple stock datasets through multi-scale feature extraction and xLSTM layer design.

4. 2. Model experiment

The MSA xLSTM model proposed in this article focuses on short-term stock price prediction and improves accuracy through multi-stage data processing and feature fusion. The input integrates two types of data: textual data (sentiment index series generated by FinBERT for news/stock reviews) and sequential data (highly correlated technical indicators filtered by Spearman correlation coefficient). The data is first reduced in non stationarity through RevIN technology, and then the input sequence is split into multiple local patches through patching encoding (such as converting long sequences into several short segments), enhancing the ability to extract local features. The core of the model adopts a parallel multi-scale attention module: the input sequence is maximally pooled through kernels of different sizes, decomposed into multi-scale subsequences (such as coarse-grained capture of long-term trends, fine-grained retention of short-term details), and each subsequence is separately subjected to multi head self attention (processing technical indicators) and emotion self attention (extending emotion index to technical indicator dimension, calculating attention weights through dot product) to extract features. The results of the two types of attention are added and fused. Multi scale information is restored and accumulated to the main sequence through transposed convolution, achieving cross scale feature integration. The prediction layer adopts a three-layer xLSTM structure (m-s-m stacking method), where the second layer sLSTM introduces residual connections (input directly added to output) to enhance information transmission efficiency and model performance. Finally, the predicted values are output through the MLP layer 错误!未找到引用源。 (double fully connected layer+Gelu activation), and the data distribution is restored to the original range through reverse RevIN. This structure effectively improves short-term prediction accuracy by extracting multi-scale features, deeply integrating emotions and technical indicators, and capturing long-term dependencies using xLSTM, addressing the shortcomings of traditional methods in feature timeliness, indicator screening, and model design.

4. 3. Effect analysis

This experiment selects four different feature stock datasets, with a time range from January 1, 2020 to September 1, 2024. Technical indicators are screened for strongly correlated variables (such as moving averages, random indicator D-values, etc.) using Spearman correlation coefficients, while sentiment indices are calculated using FinBERT based on news/stock review texts (positive/negative score difference ranging from -1 to 1). Missing data are retained using the previous day's values. The data is subjected to maximum minimum normalization and the original range is restored after prediction.

Table 3 The Impact of Historical Window Length on Short term Stock Prediction Accuracy

Time Step	Ping An (601318) RMSE	MAE	Haili (600019) RMSE	MAE	Fuling Zhacai (002507) RMSE	MAE	Mindray Medical (300760) RMSE	MAE
t=5	0.594	0.451	0.222	0.149	0.254	0.180	5.327	3.922
t=7	0.575	0.442	0.201	0.136	0.245	0.179	4.492	3.477
t=10	0.598	0.452	0.209	0.143	0.252	0.181	4.849	3.570
t=15	0.662	0.508	0.239	0.160	0.269	0.188	6.463	5.121

The experimental setup adopts PyTorch framework, AdamW optimizer (learning rate 0.001),

MSE+L1 regularization loss function ($\alpha_1=\alpha_2=0.5$), batch size 16, multi head attention head count 8, and multi-scale processing $k=3$. The time step experiment shows that the historical window has the smallest error when it is 7. In the comparative experiment, the MSA xLSTM model outperformed the CNN-LSTM, GRU Attention, CNN-LSTM Attention, EEMD-TCN, and PatchCT models in terms of RMSE/MAE metrics (as shown in Table 3).

The visualization effect shows that the predicted sequence is more in line with the true values. The ablation experiment validates the effectiveness of emotional features, multi-scale decomposition, and xLSTM module. Application case: Through backtesting simulation of aggressive ultra short term trading scenarios, the cumulative return rate of the prediction based trading strategy (including securities lending) is significantly better than the "buy and hold" baseline. Among them, small cap stocks have greater volatility, and the return rate of the securities lending strategy has increased more significantly (such as stock B's return rate exceeding the baseline by 97%). The ChiNext stock needs to pay attention to risk control due to higher limit on rise and fall. This chapter's model effectively improves short-term prediction accuracy through emotion multi feature fusion and xLSTM structure, providing practical application references for ultra short term investors.

5. Conclusion

Financial time series forecasting requires data analysis, preprocessing, and feature extraction optimization due to its nonlinear, non-stationary, and high noise characteristics. This article focuses on two major tasks: long-term trend prediction and short-term prediction of emotion sensitive stocks, and constructs a dual model system: the long-term prediction model captures temporal context dependencies through sparse attention, supplemented by channel attention branches to enhance cross feature trend information integration; The short-term prediction model integrates FinBERT emotion index 错误!未找到引用源。 with technical indicators, achieves feature fusion through a multi-scale emotion attention module, and combines xLSTM to establish long-range dependencies to improve short-term accuracy. During the data processing stage, RevIN stationary sequences are uniformly used 错误!未找到引用源。 , and the ability to extract local features is enhanced through adaptive patching 错误!未找到引用源。 . The backtesting verification shows that the prediction based trading strategy demonstrates practical application value in both long and short-term scenarios. Future research will expand the joint modeling of time-domain and frequency-domain to capture multidimensional dependencies and explore the generalization ability of models in fields such as transportation and energy; At the same time, for stocks with weak fundamentals, it is necessary to deepen the analysis of the correlation between multiple factors such as finance and policies and stock prices, improve the integration framework of all factor characteristics, and promote the evolution of financial time series forecasting towards a more universal and accurate direction.

References

- [1] Xu S, Zhang W, Tao S, et al. A Dynamic Pig Face Detection Method Based on Spatial Channel Weight Optimization Attention Mechanism and Adaptive Anchor Point Selection. *Instrumentation and Measurement, IEEE Transactions on*, 2025, 74(000):1-13.
- [2] Huang Q, Chen J. xLSTM-FER: Enhancing Student Expression Recognition with Extended Vision Long Short-Term Memory Network. *Communications in Computer and Information Science*, 2025:249-259.

- [3] Dai Z, Li D. MDSE-SLSTM: A Mobility-Driven Based Deep Learning Framework for Passenger Flow Distribution Forecasting in Multimodal Transportation Hub[C]//International Conference on Traffic and Transportation Studies. Springer, Singapore, 2025.
- [4] Wang H, Zou D, Zhao B, et al. RDLinear: A Novel Time Series Forecasting Model Based on Decomposition with RevIN. 2024 International Joint Conference on Neural Networks (IJCNN), 2024:1-7.
- [5] Qi, Y. (2025). Data Consistency and Performance Scalability Design in High-Concurrency Payment Systems. *European Journal of AI, Computing & Informatics*, 1(3), 39-46.
- [6] Jiang, Y. (2025). Application and Practice of Machine Learning Infrastructure Optimization in Advertising Systems. *Journal of Computer, Signal, and System Research*, 2(6), 74-81.
- [7] Zou, Y. (2025). Automated Reasoning and Technological Innovation in Cloud Computing Security. *Economics and Management Innovation*, 2(6), 25-32.
- [8] An, C. (2025). Study on Efficiency Improvement of Data Analysis in Customer Asset Allocation. *Journal of Computer, Signal, and System Research*, 2(6), 57-65.
- [9] Huang, J. (2025). Optimization and Innovation of AI-Based E-Commerce Platform Recommendation System. *Journal of Computer, Signal, and System Research*, 2(6), 66-73.
- [10] Zhang, X. (2025). Optimization of Financial Fraud Risk Identification System Based on Machine Learning. *Journal of Computer, Signal, and System Research*, 2(6), 82-89.
- [11] Wang, Y. (2025). Exploration and Clinical Practice of the Optimization Path of Sports Rehabilitation Technology. *Journal of Medicine and Life Sciences*, 1(3), 88-94.
- [12] Li W. The Influence of Financial Due Diligence in M&A on Investment Decision Based on Financial Data Analysis. *European Journal of AI, Computing & Informatics*, 2025, 1(3): 32-38.
- [13] Sheng, C. (2025). Innovative Application and Effect Evaluation of Big Data in Cross-Border Tax Compliance Management. *Journal of Computer, Signal, and System Research*, 2(6), 40-48.
- [14] Sheng, C. (2025). Research on the Application of AI in Enterprise Financial Risk Management and Its Optimization Strategy. *Economics and Management Innovation*, 2(6), 18-24.
- [15] Tu, X. (2025). Optimization Strategy for Personalized Recommendation System Based on Data Analysis. *Journal of Computer, Signal, and System Research*, 2(6), 32-39.
- [16] Zhu, Z. (2025). Cutting-Edge Challenges and Solutions for the Integration of Vector Database and AI Technology. *European Journal of AI, Computing & Informatics*, 1(2), 51-57.
- [17] Lai L. Data-Driven Credit Risk Assessment and Optimization Strategy Exploration. *European Journal of Business, Economics & Management*, 2025, 1(3): 24-30.
- [18] Chang, Chen-Wei. "Compiling Declarative Privacy Policies into Runtime Enforcement for Cloud and Web Infrastructure. " (2025).
- [19] Sun, Jiahe. "Research on Sentiment Analysis Based on Multi-source Data Fusion and Pre-trained Model Optimization in Quantitative Finance. " (2025).
- [20] Huang, J. (2025). Promoting Cross-field E-Commerce Development by Combining Educational Background and Technology. *Economics and Management Innovation*, 2(4), 26-32.