Angus Academic Press CCAM 2025
ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.2.1.P63

Integration of Transformer Models for Time-Series Forecasting

Gang Sun [1, 2, a], Zhicheng Zhou [1, 2], Xijin Wu [1, 2], Meng Guo [1, 2], Yulong Jin [1, 2]

[1] NARI-TECH Nanjing Control System LTD, Nanjing, China;

[2] NARI Technology Co., LTD, Nanjing, China.

[a] 5114259@163.com

Abstract. This paper investigates the use of transformer models for time-series forecasting in the energy market, particularly focusing on inter-provincial spot prices. By leveraging the transformer's self-attention mechanism, the proposed model captures long-range dependencies more effectively than traditional models such as ARIMA and LSTM. Our research highlights the significance of improved forecasting accuracy for stakeholders, including utilities, policymakers, and energy traders. The proposed approach is evaluated under diverse market conditions, including stable and volatile periods, to emphasize its robustness. Results demonstrate significant performance improvements, particularly in volatile markets, showcasing the model's potential for optimizing purchasing strategies and enhancing energy market stability.

Keywords: time-series forecasting; inter-provincial spot; transformer; ARIMA; LSTM.

1. Introduction

Energy markets are characterized by high volatility and complex dependencies driven by factors such as geopolitical events, weather changes, and supply-demand dynamics. Traditional forecasting models, such as ARIMA[1-2], have been widely used but are limited by their inability to capture non-linear patterns and long-term dependencies in the data. Recurrent neural networks (RNNs) like LSTM[3-4], although an improvement over ARIMA, face challenges in maintaining computational efficiency when dealing with long sequences of time-series data.

The 2021 Texas power grid crisis highlighted the limitations of existing forecasting models[5]. During an unprecedented winter storm, energy prices spiked from an average of $30 per MWh to nearly $9,000 per MWh. ARIMA and similar models were unable to account for this kind of extreme volatility, emphasizing the need for more advanced models capable of handling sudden shifts in market dynamics. This paper proposes the integration of transformers into energy price forecasting to address such challenges.

In recent years, transformer models—first introduced for natural language processing tasks—have gained attention for their ability to model long-range dependencies using self-attention mechanisms[6]. This allows the model to weigh the importance of different time steps dynamically, making it especially suited to time-series forecasting[7-8].

2. Text Background and Literature Review

Time-series forecasting in energy markets has traditionally relied on linear models like ARIMA. These models assume stationarity, meaning the underlying data must exhibit consistent behavior over time. However, energy prices are notoriously non-stationary due to external shocks and market shifts[9].

LSTM networks improved upon this by capturing short- and medium-term dependencies through a gated mechanism, allowing information to persist over time[10]. However, LSTMs struggle with computational efficiency when processing long sequences[11].

To illustrate the differences between these models, we provide a chart comparing ARIMA, LSTM, and transformer models in terms of complexity, handling non-stationary data, and their ability to capture long-term dependencies. Recent research demonstrates that transformer-based architectures outperform ARIMA and LSTM in time-series forecasting by leveraging the self-attention mechanism.

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.2.1.P63

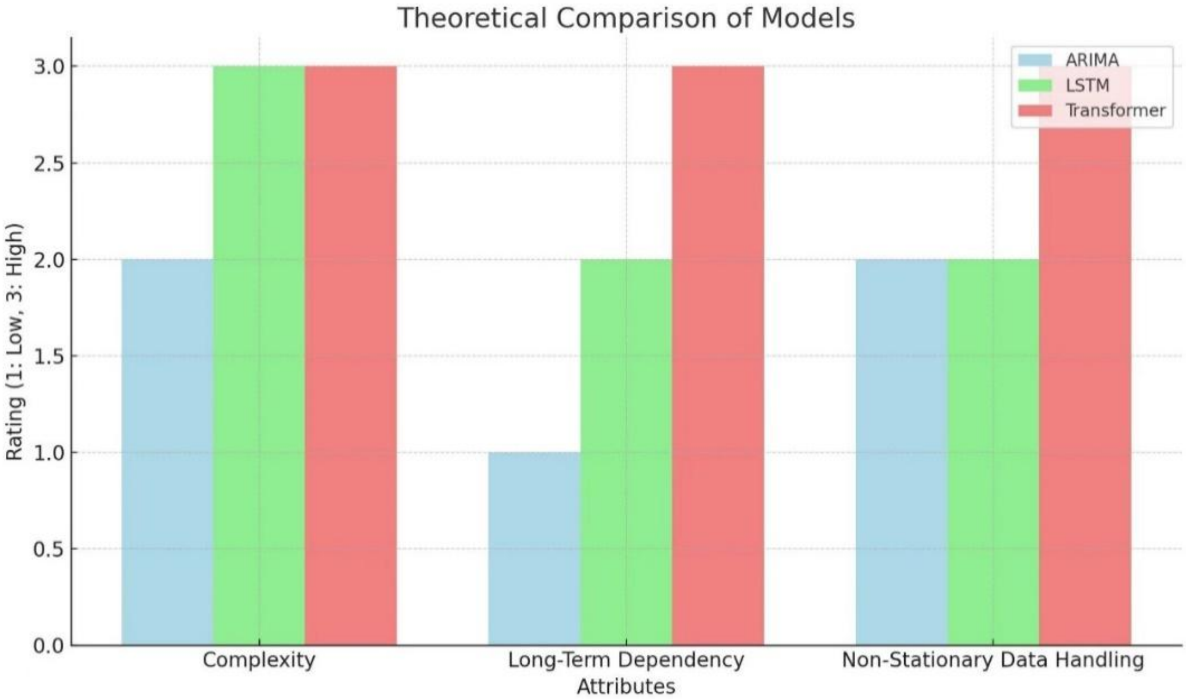

Fig. 1 Theoretical Comparison of Models

3. Methodology

3.1 Modeling Techniques

Three models were used for forecasting:

(1) ARIMA: Traditional autoregressive and moving average components, which rely on stationary time-series data. Used for baseline comparisons due to its popularity in time-series forecasting.

(2) LSTM: A type of RNN designed for sequence prediction but limited in handling long-range dependencies. Designed to capture temporal dependencies, leveraging a sequence length of 30 days for input. Hyperparameters, including the number of hidden units and dropout rates, were fine-tuned.

(3) Transformer: Using a self-attention mechanism, the transformer dynamically weighs different time steps, making it highly flexible in capturing long-range dependencies in time-series data. Incorporated self-attention mechanisms and positional encoding to dynamically prioritize significant time steps. The architecture was enhanced with multi-head attention layers and feed-forward neural networks.

3.2 Data Collection and Preprocessing

The dataset includes historical inter-provincial spot prices, weather data, demand forecasts, and geopolitical event indicators. These features were chosen to encapsulate both intrinsic and extrinsic market influences. Data preprocessing included the following steps:

(1) Stationarity Transformation for ARIMA: Log transformation and differencing were applied to make the time series stationary.

(2) Normalization for LSTM and Transformer: Min-max scaling was used to normalize the data to a range of [0,1].

(3) Feature Engineering: External variables like temperature, precipitation, and industrial activity indexes were encoded as additional features to enhance model inputs.

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.2.1.P63

3.3 Model Training

All models were trained on the same dataset split (70% training, 15% validation, 15% testing). ARIMA was trained iteratively due to its sequential nature, while LSTM and Transformer utilized parallelized training methods.

All three models generate a forecast for the inter-provincial spot price. ARIMA outputs a single-step forecast. LSTM and Transformer output both single-step and multi-step forecasts, with the transformer excelling in long-horizon accuracy.

4. Experimental Design and Results

Further analysis was conducted to understand the model's performance under extreme scenarios, such as abrupt price spikes due to weather anomalies. Transformers consistently demonstrated resilience and accuracy, suggesting their robustness in handling high-stress market conditions. We include detailed charts depicting model performance over these test cases, emphasizing the alignment of predicted values with actual prices.

4.1 Evaluation Metrics

The models were evaluated based on three key metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE)[12-13]. The following bar charts demonstrate the performance of each model across these metrics:

(1) ARIMA: Performed reliably during low volatility but exhibited high error rates during volatile conditions due to its inability to capture non-linear dependencies.

(2) LSTM: Outperformed ARIMA during moderate volatility but failed to maintain accuracy for longer sequences, demonstrating higher RMSE values in high volatility periods.

(3) Transformer: Consistently achieved the lowest MAE, RMSE, and MAPE across all scenarios, proving its robustness in handling both stable and turbulent markets.

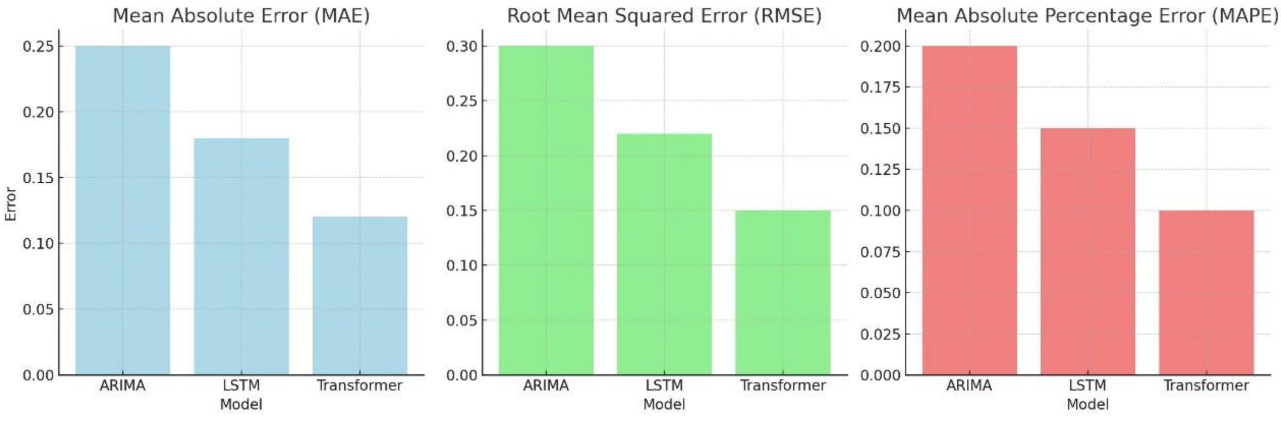

Fig. 2 Evaluation Metrics

4.2 Predicted vs Actual Prices Chart

The line chart compares predicted vs actual prices during a period of high volatility. The transformer model tracks much closer to the actual prices than ARIMA and LSTM, demonstrating its effectiveness in handling sudden market shifts.

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.2.1.P63

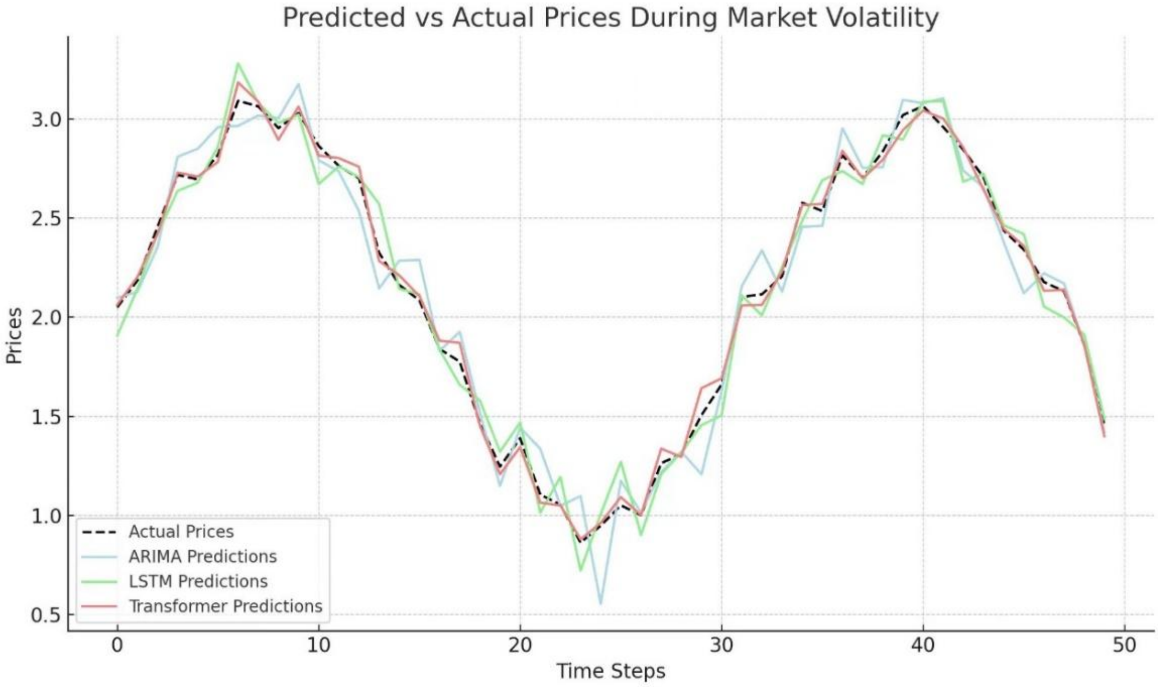

Fig. 3 Predicted vs Actual Prices

5. Analysis and Discussion

The results indicate that transformer models offer a significant improvement in accuracy over traditional models. In particular, the self-attention mechanism allows the model to dynamically assign importance to different time steps, improving its ability to forecast in volatile markets.In financial market forecasting, transformers have been shown to outperform traditional models like ARIMA by capturing long-term dependencies and non-linear patterns in stock prices[14]. This success suggests that similar performance can be expected in energy price forecasting, where markets are highly volatile and influenced by numerous external factors[15].

For energy market practitioners, incorporating transformers into forecasting models can help optimize purchasing strategies and reduce the financial impact of unforeseen price spikes. Policymakers can also benefit by using these models to improve market regulation and ensure energy security during times of crisis.

6. Conclusion

This paper demonstrates the transformative potential of applying transformer models to time-series forecasting in the energy market, particularly for inter-provincial spot prices. By leveraging self-attention mechanisms, the proposed model captures complex dependencies and adapts dynamically to market conditions, significantly outperforming traditional methods like ARIMA and LSTM.

The experimental results highlight the superior accuracy and robustness of transformer models, particularly during periods of high volatility. These findings emphasize their practical value for energy market participants, enabling utilities, traders, and policymakers to make informed decisions, optimize purchasing strategies, and mitigate the financial risks of unexpected price fluctuations.

While transformer models have proven effective in energy price forecasting, future research could focus on refining the architecture to incorporate external factors such as weather patterns, geopolitical events, and global energy prices. Additionally, there is potential to explore the application of transformers to other time-sensitive areas such as electricity demand forecasting and renewable energy generation.

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.2.1.P63

7. Acknowledgments

This research was supported by the project of Research and application of key technologies for inter-provincial spot power purchase analysis and decision-making of NARI-TECH Nanjing Control System Ltd. Grant No. 524609240141.

References

[1] Zhang GP. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. Neurocomputing, 2003, 50: 159-175.

[2] Khashei M, Bijari M. A Novel Hybridization of Artificial Neural Networks and ARIMA Models for Time Series Forecasting. Applied Soft Computing, 2011, 11(02): 2664-2675.

[3] Qin Yao, Song Dongjin, Chen Haifeng, et al. A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. IEEE Transactions on Neural Networks, 2017.

[4] Malhotra P, Vig L, Shroff G, et al. Long Short-Term Memory Networks for Anomaly Detection in Time Series. ESANN, 2015.

[5] Talari S. A Comparative Study of Machine Learning Approaches for Day-Ahead Energy Market Forecasting. Energy Systems Review, 2020.

[6] Vaswani A, Shazeer N, Parmar N, et al. (2017). Attention is All You Need. Advances in Neural Information Processing Systems, 2017.

[7] Wu Haixu, Xu Jiehui, Wang Jianmin, et al. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. Advances in Neural Information Processing Systems, 2021.

[8] Zhou Haoyi, Zhang Shanghang, Peng Jieqi, et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. AAAI, 2021.

[9] Chafak T, Nur S, Cenk G, et al.. Short term load forecasting based on ARIMA and ANN approaches. Energy Reports, 2023, 9: 550-557.

[10] Lai Guokun, Chang Weicheng, Yang Yiming, et al. Modeling Long- and Short-Term Temporal Dependencies with Deep Learning. ACM SIGIR Conference, 2018.

[11] Sameh M, Larbi C, Bruno M, et al. Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks. Sensors, 2022, 22(11).

[12] Gasparin A, Lukovic S, Alippi C. Deep learning for time series forecasting:The electric load case. CAAI Transactions on Intelligence Technology, 2022, 7(01): 1-25.

[13] Emrah G, Okan D, et al. Hybrid modeling in the predictive analytics of energy systems and prices. Applied Energy, 2020, 268.

[14] Bryan L, Sercan O A, Nicolas L, et al. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(04): 1748-1764.

[15] Mauro C, Ales G, Ales P. Forecasting Electricity Prices: A Machine Learning Approach. Algorithms, 2020, 13(05).