Comparative Study of Prediction Models based on Modified Large Language Model for South Korea C2X Energy Demand Forecasting

Nuh Enola, Rahma Gantassi, Yonghoon Choi

Chonnam National University

{nuhenola, rahmag, yh.choi}@jnu.ac.kr

한국 C2X 에너지 수요 예측을 위한 수정 대언어 모델 기반 예측 모델 비교 연구

에놀라 누, 간타씨 라흐마, 최용훈

전남대학교

Abstract

This study evaluates time-series prediction algorithms, focusing on Modified Generative Pre-trained Transformer 2 (GPT-2), Long Short Term Memory (LSTM), Random Forest (RF), and Time Series Transformer (TST) for energy demand forecasting under two scenarios: same-dataset training and cross-dataset generalization. The results demonstrate that the Modified GPT-2 model consistently outperforms all other models across various scenarios, including both stable and unpredictable datasets. Its superior ability to handle high variability, accurately capture anomalies, and generalize effectively makes it as potential robust and reliable solution for dynamic applications like energy trading.

I. Introduction

Prediction algorithms play a critical role in optimizing decision-making in energy trading, particularly in Peer-to-Peer (P2P) systems where dynamic pricing reflects local energy production and consumption [1]. Unlike images or videos, which typically have consistent input scales and sampling rates, aggregated time series datasets often consist of sequences from highly diverse sources. These datasets frequently contain missing values and require extrapolation from observations that represent only a small fraction of the overall information, making accurate point predictions challenging and uncertainty estimation essential [2],[3]. Recently, large language models (LLMs) like Generative Pre-trained Transformer 2 (GPT-2), initially designed for natural language processing, have demonstrated significant potential in time series forecasting. Their ability to capture intricate sequential patterns and long-term dependencies, combined with their flexibility for generalizable forecasting without requiring retraining from scratch, makes them well-suited to the complexities of P2P energy trading [4],[5]. This study seeks to implement LLM-based prediction using real-world energy demand data and evaluate its performance against other forecasting methods like Long Short Term Memory (LSTM), Random Forest (RF), and Time Series Transformer (TST).

II. Method

This study adopts a simulation-based approach using Python, utilizing real-world demand and supply datasets spanning January 2023 to December 2023, obtained from Grida Energy's Community to-X (C2X) Project. The simulations are conducted under two experimental scenarios:

1. **Trained and tested on the same dataset**: The model is evaluating real-world demand data of Shinyeocheon Road 1 No. 10, provided by Grida Energy, using a traditional data split (70% training, 20% validation, and 10% testing).
2. **Trained on one dataset and tested on another**: the model is trained on data from Shinyeocheon Road 1 No. 10 and tested on data from a different dataset, Shinyeocheon Road 2 No. 11. Both datasets are sourced from the same location, ensuring that they share similar dynamic patterns, but the test data represents new, unseen data.

The simulation flow is illustrated in Fig. 1, which provides a comprehensive overview of the prediction process. The pipeline includes data preprocessing, model training, prediction generation, and evaluation. Python libraries such as NumPy, pandas, and PyTorch are employed for data manipulation and modeling, ensuring flexibility and scalability.

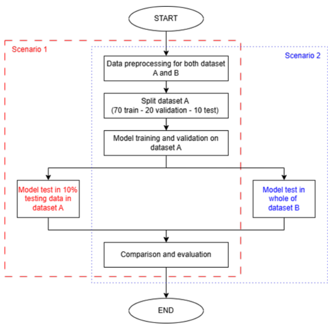

Fig. 1. Simulation workflow

In this study, the key modification is adapting and modifying the GPT-2 model by adjusting its attention mechanism to better accommodate numerical sequences and ensure temporal alignment. These enhancements enable the model to focus on both short-term and long-term temporal dependencies as well as manage variable-length input sequences, a common feature of real-world time-series data. The architecture of the proposed approach is summarized in Fig. 2, which outlines the preprocessing and model adaptation.

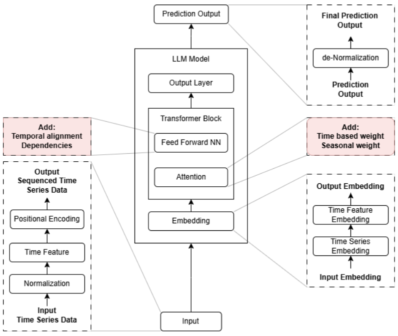

Fig. 2. Proposed model architecture.

To ensure consistency across all experiments, the same hyperparameters are used for every model: 30 sequence length, 0.0001 learning rate, 50 epoch, and 32 batch size.

Ⅲ. Result and Analysis

The simulation is run as explained in method section and resulting a real and prediction comparison graph as shown in Figure 3.

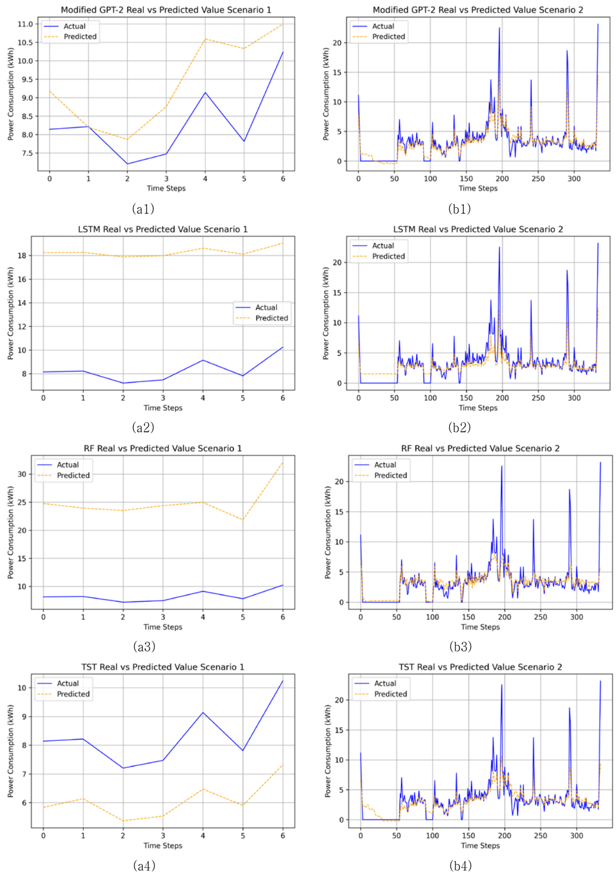

Fig. 3. Actual and prediction comparison plot for (1) Modified GPT-2, (2) LSTM, (3) RF, and (4) TST in the (a) first scenario and (b) second scenario.

From Figure 3, all models capture the general trend of the actual data, but with varies behavior. Modified GPT-2 excels with strong trend-following behavior and minimal error gaps, robustly capturing rapid changes and effectively handling anomalies like sudden peaks or troughs, making it ideal for energy trading demand modeling. TST also shows a decent trend-following but has noticeable error gaps and struggles with abrupt anomalies, performing better in less dynamic scenarios. LSTM moderately captures overall trends but exhibits larger error gaps and slower responses to rapid changes, with poor anomaly handling. RF performs the weakest, struggling to adapt to dynamics changes, showing significant error gaps, and failing to capture anomalies.

To further support these findings, performance metrics such as MSE, RMSE, and MAE are calculated for each model and scenario to assess their accuracy, and these results are summarized in Table 1.

Table 1. Evaluation Metrics

Model	Scenario 1			Scenario 2		
	MSE (kWh^2)	RMSE (kWh)	MAE (kWh)	MSE (kWh^2)	RMSE (kWh)	MAE (kWh)
Modified GPT-2	1.74	1.32	1.11	1.53	1.23	0.73
RF	199.90	14.14	14.00	4.02	2.00	0.89
LSTM	72.46	8.51	8.48	5.07	2.25	0.14
TST	5.14	2.27	2.24	3.13	1.77	0.95

The results show that the Modified GPT-2 outperformed others with the lowest RMSE 1.32 and MAE 1.11, showcasing high precision. TST follows with RMSE 2.27 and MAE 2.24, indicating decent but less accurate predictions. LSTM and RF lagged, with RF performing the worst due to its inability to handle temporal data complexities. In the second scenario, Modified GPT-2 maintained superiority with RMSE 1.23 and MAE 0.73, excelling in generalizing unseen data and managing unpredictable dynamics. TST trailed with RMSE 1.77 and MAE 0.95, struggling with data variability. LSTM and RF show higher errors, reflecting their challenges in capturing the complexities of the new dataset.

IV. Conclusion

This study demonstrated time-series forecasting across two scenarios for several models. The simulation results show that Modified GPT-2 outperformed in both scenarios, underscoring its ability to do prediction normal cases and generalization tasks as well as showcasing its versatility in dynamic applications like energy trading. These findings underline the potential of Modified GPT-2 for complex temporal tasks while acknowledging opportunities for enhancement. Future work will compare with focus on improving Modified GPT-2 or combining its strengths with time-series-specific algorithms, such as TST or XG-Booster, to achieve both adaptability and precision for more robust predictions.

ACKNOWLEDGMENT

This paper (author) is a study supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea in 2023 (NRF-2023S1A5C2A07096111).

REFERENCES

[1] Q. A. Sias, R. Gantassi, and Y. Choi, “Multivariate bidirectional gate recurrent unit for improving accuracy of energy prediction,” ICT Express, vol. 11, no. 1, pp. 80–86, Feb. 2025, doi: 10.1016/j.icte.2024.10.002.

[2] N. Gruver, M. Finzi, S. Qiu, and A. G. Wilson, “Large Language Models Are Zero-Shot Time Series Forecasters.” [Online]. Available: https://github.com/ngruver/llmtime.

[3] M. Jin et al., “Time-LLM: Time Series Forecasting by Reprogramming Large Language Models,” Oct. 2023, [Online]. Available: http://arxiv.org/abs/2310.01728

[4] R. Gantassi, Z. Masood, and Y. Choi, “Machine Learning for QoS Optimization and Energy-Efficient in Routing Clustering Wireless Sensors,” Computers, Materials & Continua, vol. 82, no. 1, pp. 327-343, Jan. 2025, doi: 10.32604/cmc.2024.058143.

[5] Y. Jiang et al., “Empowering Time Series Analysis with Large Language Models: A Survey,” Feb. 2024, [Online]. Available: http://arxiv.org/abs/2402.03182