Unveiling the Potential of Text in High-Dimensional Time Series Forecasting

Xin Zhou*
Monash University
Melbourne, Victoria, Australia

Weiqing Wang
Monash University
Melbourne, Victoria, Australia

Shilin Qu
Monash University
Melbourne, Victoria, Australia

Zhiqiang Zhang
Southeast University
Nanjing, China

Christoph Bergmeir
University of Granada
Granada, Spain

Abstract

Time series forecasting has traditionally focused on univariate and multivariate numerical data, often overlooking the benefits of incorporating multimodal information, particularly textual data. In this paper, we propose a novel framework that integrates time series models with Large Language Models to improve high-dimensional time series forecasting. Inspired by multimodal models, our method combines time series and textual data in the dual-tower structure. This fusion of information creates a comprehensive representation, which is then processed through a linear layer to generate the final forecast. Extensive experiments demonstrate that incorporating text enhances high-dimensional time series forecasting performance. This work paves the way for further research in multimodal time series forecasting.

1 Introduction

Time series forecasting plays a vital role in various domains such as finance Wang et al. (2023); Zheng et al. (2023); Gabrielli et al. (2023), traffic Eldele et al. (2024); Zhao et al. (2023); Zhang et al. (2023), healthcare Angelopoulos et al. (2024); Shen et al. (2023); Saldanha et al. (2024), and weather prediction Cheng et al. (2024); Auer et al. (2024); Chatzigeorgakidis et al. (2024). Traditional deep learning-based methods focus on time series analysis, relying purely on the historical numerical values of individual series. However, they often overlook the potential of textual data, particularly in high-dimensional time series forecasting, which involves a large number of channels and complex, diverse temporal patternsDonoho et al. (2000); Barigozzi et al. (2024). In such a scenario, textual information can provide valuable supplementary insights.

The recent success of Large Language Models in Natural Language Processing and Computer Vision has sparked interest in using textual data to enhance time series forecasting Xi et al. (2023); Jia et al. (2024), however, their data are low-dimensional. Thus, incorporating textual information into high-dimensional time series forecasting remains unexplored. Our work aims to bridge this gap by being the first to integrate textual data into high-dimensional time series forecasting, introducing a novel challenge and paving the way for future research in this field. We propose that integrating numerical time series data with textual information allows models to develop a more comprehensive understanding of underlying dynamics, ultimately improving forecasting accuracy.

In our work, we propose a general framework - TextFusionHTS, integrating textual information into high-dimensional time series forecasting, focusing on the challenges of extracting and capably fusing textual data with time series inputs. Our key contributions are as follows: (1) Framework for Text-Integrated High-Dimensional Time Series Forecasting: We introduce a new framework TextFusionHTS that integrates textual data into high-dimensional time series analysis. By integrating text-based information, we open up new possibilities for high-dimensional time series forecasting

*xin.zhou@monash.edu

NeurIPS 2024 Workshop on Time Series in the Age of Large Models.

and pave the way for further research in multimodal time series analysis; (2) Experiments Validation: We conduct extensive experiments demonstrating the effectiveness of incorporating textual data into high-dimensional time series forecasting. Our results [2] reveal that textual information can notably enhance predictive performance, highlighting the untapped potential of multimodal inputs in this field.

2 Related Works

2.1 Time Series Forecasting

Time series forecasting methods have employed various deep learning techniques, including Linear Wang et al. (2024); Zeng et al. (2023); Darlow et al. (2023), Convolutional Neural Networks Cheng et al. (2024); Zhang et al. (2023); Luo and Wang (2024), and Transformer Piao et al. (2024); Liu et al. (2024b); Boussif et al. (2024); Chen et al. (2024). Most of them mainly solve the challenge of capturing short-term and long-term dependencies. PatchTST Nie et al. (2022) has proven to be a leading technique that divides the series window into patches, allowing it to capture the complex global and local relationships within individual series. Recently, LLM-based methods, e.g., TimeLLM Jin et al. (2023), TEMPO Cao et al. (2023), TEST Sun et al. (2023), and TimeCMA Liu et al. (2024a) have explored adapting LLMs for time series tasks by converting time series data into text formats that LLMs can effectively process.

2.2 Multimodal Time Series Forecasting

Several methods have explored integrating different data types. Fu et al. (2024) introduced a framework that incorporates video information into time series forecasting, while Multimodal Time-series Analysis Xi et al. (2023) fuses textual, visual, and audio data to forecast gifting behavior on live-streaming platforms. GPT4MTS Jia et al. (2024) proposes a news multimodal dataset based on GDELT and provides a pipeline for data extraction and processing. These efforts demonstrate the potential of multimodal forecasting, though they often encounter challenges due to the lack of published code or data. Additionally, MST-GAT Ding et al. (2023) utilizes graph attention mechanisms to manage multimodal inputs, specifically designed for anomaly detection in time series data. While these methods show promise in integrating textual data, the performance of fusing different modalities in high-dimensional time series forecasting remains unclear.

3 Method

Inspired by the dual-tower architecture commonly used in multimodal methods Koh et al. (2024); Sun et al. (2024), we propose a framework TextFusionHTS that integrates Large Language Models with time series models, as shown in Figure 1.

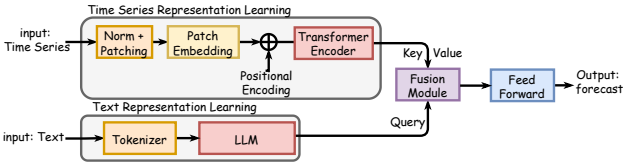


Figure 1: Dual-tower structure of TextFusionHTS.

3.1 Time Series Representation Learning

We first use a patch-based method, PatchTST, to extract the representation of the time series data. The input series $x \in \mathbb{N}^l$ with l length is divided into p patches. Using the patch-based method is crucial in dynamically using static textual data. Following PatchTST, dividing time series into patches allows the model to capture local temporal patterns within each patch. Since textual data is static, it's important to assign textual data to different interactions with various time series patches of the series. For example, the text may provide context that is more relevant to certain periods or patterns in the series. Therefore, using a patch-based method ensures that the model can apply the static text data in a context-sensitive manner for different time series windows.

[2] https://github.com/xinzzzhou/TextFusionHTS.git

PatchTST generates embedding for each series as follows,

$$z_{ts} = PatchTST(x) \in \mathbb{R}^{p \times d_{ts}}, \quad K = V = z_{ts} \tag{1}$$

where d_{ts} denotes the dimension of each patch's representation. This matrix serves as the key (K) and values (V) for the subsequent attention mechanism for the fusion step.

3.2 Text Representation Learning

To incorporate textual data, we use Meta Llama-3.1 8B, the state-of-the-art open-source LLM, to process the text description corresponding to the time series. The text is tokenized and converted into a set of toke embeddings. Let the set of token embeddings be $\{t_i \in \mathtt{R}^{d_{tx}} | i = 1, 2, ..., n\}$ where d_{tx} is the dimension of text representation, n is the number of tokens. We compute the average of all token embeddings to get full semantic meaning of text,

$$z_{tx} = \frac{1}{n} \sum_{i=1}^{n} t_i \in \mathbb{R}^{d_{tx}}, \quad Q = z_{tx} \tag{2}$$

This averaged embedding z_{tx} serves as the query (Q) in the cross-attention mechanism.

3.3 Fusion Module

We apply a cross-attention mechanism where the time series representations (K and V) are combined with the textual representation (Q). This operation results in a fused representation $z = CrossAttention(Q, K, V) \in \mathbb{R}^d$ that integrates information from both modalities.

Finally, we use a Feed Forward transformation to map the fused representation z to the forecast values $y \in \mathbb{R}^h$, resulting in the final predictions in h horizon.

4 Experiments

4.1 Experimental Settings

We collect two high-dimensional time series forecasting datasets with textural data. *Wiki-People* records daily traffic of Wikipedia webpages from July 2015 to December 2016. It was published in the Kaggle competition. We keep 3,857 full series published in the English language in the Wikipedia agent and crawl text data from the corresponding pages. *News* Moniz and Torgo (2018) is originally a large dataset of news items and their respective social feedback on Facebook. The collected data relates to a period of 8 months, between November 2015 and July 2016, accounting for 26,612 news items on the 'Obama' topic. The time interval is 20 minutes. Given that the time interval is 1 day for Wiki-People and 20 minutes for News, we set the input lengths to 7 and 9, corresponding to 1 week and 3 hours, respectively. The forecasting horizon h is set to $\{7, 14, 21, 28, 35\}$ for Wiki-People and $\{1, 3, 9, 12, 15\}$ for News.

All implementations are done in PyTorch and trained on a single NVIDIA A100 GPU. To ensure model convergence during training, we set the number of training epochs to 100 and employ an early stopping mechanism. Specifically, if the change in validation loss is less than 1×10^{-4}, the training stops, and testing begins.

4.2 Result Comparison

To assess the impact of incorporating textual information with our proposed framework, we conduct the following experiments for comparison: 1) using purely time series data as input, with PatchTST as the time series model; 2) using both time series data and textual data as input, with the framework introduced in Section 3.

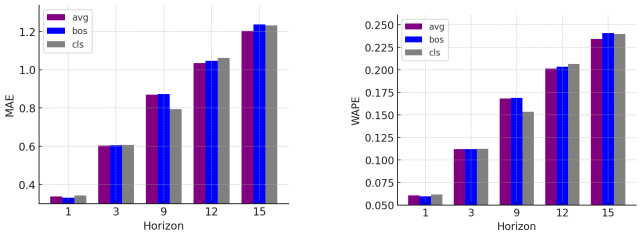
The Table 1 summarizes the overall results across two datasets, showing that integrating textual data into our framework consistently improves performance in both Mean Absolute Error (MAE) and Weighted Absolute Percentage Error (WAPE). While the only exception, is that when the forecasting horizon h is set to 21 on the Wiki-People dataset. When the time series data alone fails to provide adequate context, the additional textual data acts as a valuable complement, enhancing forecasting accuracy.

Table 1: Experimental results. $TextFusionHTS_{wo}$ is using purely time series as input, and $TextFusionHTS_{wt}$ is using both time series and text as input. The best results are in **bold**.

Dataset	News									
Metric	MAE					WAPE				
Horizon	1	3	9	12	15	1	3	9	12	15
$TextFusionHTS_{wo}$	0.3464	0.6072	0.8761	1.0431	1.2275	0.0625	0.1125	0.1695	0.2029	0.2390
$TextFusionHTS_{wt}$	**0.3367**	**0.6040**	**0.8700**	**1.0350**	**1.2031**	**0.0607**	**0.1119**	**0.1683**	**0.2013**	**0.2343**
Dataset	Wiki-People									
Metric	MAE					WAPE				
Horizon	7	14	21	28	35	7	14	21	28	35
$TextFusionHTS_{wo}$	0.5751	0.6509	0.7061	0.7506	0.7906	0.5209	0.5894	**0.6382**	0.6765	0.7108
$TextFusionHTS_{wt}$	**0.5734**	**0.6496**	**0.7053**	**0.7500**	**0.7873**	**0.5194**	**0.5883**	0.6474	**0.6759**	**0.7079**

4.3 Strategies for Text Feature Extraction

To investigate the relation between text feature extraction and horizon. We compare three strategies for extracting textual features: using the [cls] token, the [bos] token, and the average of all token embeddings. Following are the results on the News dataset.



(a) MAE (b) WAPE

Figure 2: Comparison of different strategies for text feature extraction. Here the average of all token embeddings, [bos] token embedding, and [cls] token embeddings are represented with purple, blue, and grey bars separately.

Figure 2 shows that the choice of text feature extraction method significantly affects the model's performance. For shorter horizons, the differences between the various strategies are minimal. However, when the horizon extends to 9, using the [cls] token appears to underperform compared to other strategies, possibly because [cls] fails to capture the full semantic meaning of the entire sentence. For longer horizons, the [bos] token outperforms the others, indicating that it may contain more valuable contextual information.

5 Conclusion

In this paper, we investigated the potential of incorporating textual data into high-dimensional time series forecasting. Our experiments demonstrate that using textual data can notably enhance forecasting performance, particularly for longer horizons where time series data alone may fall short in providing sufficient context. This approach effectively addresses the limitations of traditional time series models, significantly boosting their predictive capabilities. These findings underscore the importance of multimodal approaches in time series forecasting. For future work, we plan to develop more advanced models capable of fusing additional modalities, such as images and audio, to further enhance forecasting accuracy. Additionally, we aim to conduct more comprehensive experiments on diverse datasets to fully explore the benefits of integrating multiple sources of information.

Acknowledgement

This work contribution is part of the I+D+i project granted by C-ING-250-UGR23 co-funded by "Consejería de Universidad, Investigación e Innovación" and for the European Union related to the FEDER Andalucía Program 2021-2027". This work is supported by the Knowledge Generation Project PID2023-149128NB-I00, funded by the Ministry of Science, Innovation and Universities of Spain. Christoph Bergmeir is supported by a María Zambrano Fellowship that is funded by the Spanish Ministry of Universities and Next Generation funds from the European Union. Additionally, we gratefully acknowledge the support of Google for providing a travel grant, which enabled attendance at NeurIPS 2024. This funding has significantly contributed to our research efforts and facilitated essential academic exchange. We sincerely appreciate Google's generous support.

References

Angelopoulos, A., Candes, E., and Tibshirani, R. J. (2024). Conformal pid control for time series prediction. *Advances in neural information processing systems*, 36.

Auer, A., Gauch, M., Klotz, D., and Hochreiter, S. (2024). Conformal prediction for time series with modern hopfield networks. *Advances in Neural Information Processing Systems*, 36.

Barigozzi, M., Cho, H., and Owens, D. (2024). Fnets: Factor-adjusted network estimation and forecasting for high-dimensional time series. *Journal of Business & Economic Statistics*, 42(3):890–902.

Boussif, O., Boukachab, G., Assouline, D., Massaroli, S., Yuan, T., Benabbou, L., and Bengio, Y. (2024). Improving* day-ahead* solar irradiance time series forecasting by leveraging spatio-temporal context. *Advances in Neural Information Processing Systems*, 36.

Cao, D., Jia, F., Arik, S. O., Pfister, T., Zheng, Y., Ye, W., and Liu, Y. (2023). Tempo: Prompt-based generative pre-trained transformer for time series forecasting. *arXiv preprint arXiv:2310.04948*.

Chatzigeorgakidis, G., Lentzos, K., and Skoutas, D. (2024). Multicast: Zero-shot multivariate time series forecasting using llms. In *2024 IEEE 40th International Conference on Data Engineering Workshops (ICDEW)*, pages 119–127. IEEE.

Chen, P., Zhang, Y., Cheng, Y., Shu, Y., Wang, Y., Wen, Q., Yang, B., and Guo, C. (2024). Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting. *arXiv preprint arXiv:2402.05956*.

Cheng, M., Yang, J., Pan, T., Liu, Q., and Li, Z. (2024). Convtimenet: A deep hierarchical fully convolutional model for multivariate time series analysis. *arXiv preprint arXiv:2403.01493*.

Darlow, L. N., Joosen, A., Asenov, M., Deng, Q., Wang, J., and Barker, A. (2023). Tsmix: time series data augmentation by mixing sources. In *Proceedings of the 3rd Workshop on Machine Learning and Systems*, pages 109–114.

Ding, C., Sun, S., and Zhao, J. (2023). Mst-gat: A multimodal spatial–temporal graph attention network for time series anomaly detection. *Information Fusion*, 89:527–536.

Donoho, D. L. et al. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS math challenges lecture*, 1(2000):32.

Eldele, E., Ragab, M., Chen, Z., Wu, M., and Li, X. (2024). Tslanet: Rethinking transformers for time series representation learning. *arXiv preprint arXiv:2404.08472*.

Fu, Z., Wang, K., Wang, J., and Zhu, Y. (2024). Predicting sales lift of influencer-generated short video advertisements: A ladder attention-based multimodal time series forecasting framework. In *Proceedings of the Hawaii International Conference on System Sciences*.

Gabrielli, G., Melioli, A., and Bertini, F. (2023). High-dimensional data from financial statements for a bankruptcy prediction model. In *2023 IEEE 39th International Conference on Data Engineering Workshops (ICDEW)*, pages 1–7. IEEE.

Jia, F., Wang, K., Zheng, Y., Cao, D., and Liu, Y. (2024). Gpt4mts: Prompt-based large language model for multimodal time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(21):23343–23351.

Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J. Y., Shi, X., Chen, P.-Y., Liang, Y., Li, Y.-F., Pan, S., et al. (2023). Time-llm: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728*.

Koh, J. Y., Fried, D., and Salakhutdinov, R. R. (2024). Generating images with multimodal language models. *Advances in Neural Information Processing Systems*, 36.

Liu, C., Xu, Q., Miao, H., Yang, S., Zhang, L., Long, C., Li, Z., and Zhao, R. (2024a). Timecma: Towards llm-empowered time series forecasting via cross-modality alignment.

Liu, Z., Yang, J., Cheng, M., Luo, Y., and Li, Z. (2024b). Generative pretrained hierarchical transformer for time series forecasting. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2003–2013.

Luo, D. and Wang, X. (2024). Moderntcn: A modern pure convolution structure for general time series analysis. In *The Twelfth International Conference on Learning Representations*.

Moniz, N. and Torgo, L. (2018). Multi-source social feedback of online news feeds.

Nie, Y., Nguyen, N. H., Sinthong, P., and Kalagnanam, J. (2022). A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*.

Piao, X., Chen, Z., Murayama, T., Matsubara, Y., and Sakurai, Y. (2024). Fredformer: Frequency debiased transformer for time series forecasting. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2400–2410.

Saldanha, R., Ribeiro, V., Pena, E. H., Pedroso, M., Akbarinia, R., Valduriez, P., and Porto, F. (2024). Subset models for multivariate time series forecast. In *2024 IEEE 40th International Conference on Data Engineering Workshops (ICDEW)*, pages 86–90. IEEE.

Shen, T., Li, Y., and Moura, J. M. (2023). Forecasting covid-19 dynamics: Clustering, generalized spatiotemporal attention, and impacts of mobility and geographic proximity. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 2892–2904. IEEE.

Sun, C., Li, H., Li, Y., and Hong, S. (2023). Test: Text prototype aligned embedding to activate llm's ability for time series. *arXiv preprint arXiv:2308.08241*.

Sun, Q., Cui, Y., Zhang, X., Zhang, F., Yu, Q., Wang, Y., Rao, Y., Liu, J., Huang, T., and Wang, X. (2024). Generative multimodal models are in-context learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14398–14409.

Wang, D., Zhang, Z., Zhao, Y., Huang, K., Kang, Y., and Zhou, J. (2023). Financial default prediction via motif-preserving graph neural network with curriculum learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2233–2242.

Wang, S., Wu, H., Shi, X., Hu, T., Luo, H., Ma, L., Zhang, J. Y., and ZHOU, J. (2024). Timemixer: Decomposable multiscale mixing for time series forecasting. In *The Twelfth International Conference on Learning Representations*.

Xi, D., Tang, L., Chen, R., and Xu, W. (2023). A multimodal time-series method for gifting prediction in live streaming platforms. *Information Processing & Management*, 60(3):103254.

Zeng, A., Chen, M., Zhang, L., and Xu, Q. (2023). Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11121–11128.

Zhang, Z., Chen, Y., Zhang, D., Qian, Y., and Wang, H. (2023). Ctfnet: Long-sequence time-series forecasting based on convolution and time–frequency analysis. *IEEE Transactions on Neural Networks and Learning Systems*.

Zhao, Y., Luo, X., Ju, W., Chen, C., Hua, X.-S., and Zhang, M. (2023). Dynamic hypergraph structure learning for traffic flow forecasting. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 2303–2316. IEEE.

Zheng, Z., Shao, J., Zhu, J., and Shen, H. T. (2023). Relational temporal graph convolutional networks for ranking-based stock prediction. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 123–136. IEEE.