

GPT4MTS: Prompt-Based Large Language Model for Multimodal Time-Series Forecasting

Furong Jia, Kevin Wang, Yixiang Zheng, Defu Cao, Yan Liu

Department of Computer Science
Viterbi School of Engineering
University of Southern California
{florajia, kwang442, yzheng01, defucao, yanliu.cs}@usc.edu

Abstract

Time series forecasting is an essential area of machine learning with a wide range of real-world applications. Most of the previous forecasting models aim to capture dynamic characteristics from uni-modal numerical historical data. Although extra knowledge can boost the time series forecasting performance, it is hard to collect such information. In addition, how to fuse the multimodal information is non-trivial. In this paper, we first propose a general principle of collecting the corresponding textual information from different data sources with the help of modern large language models (LLM). Then, we propose a prompt-based LLM framework to utilize both the numerical data and the textual information simultaneously, named GPT4MTS. In practice, we propose a GDELTA-based multimodal time series dataset for news impact forecasting, which provides a concise and well-structured version of time series dataset with textual information for further research in communication. Through extensive experiments, we demonstrate the effectiveness of our proposed method on forecasting tasks with extra-textual information.

Introduction

Time series data has consistently played a pivotal role in diverse fields ranging from finance (Sezer, Gudelek, and Ozbayoglu 2020) and economics (Kalamara et al. 2022) to healthcare (Kaushik et al. 2020; Cao et al. 2023) and weather prediction (Nguyen et al. 2023), etc. The intricate patterns embedded within such data often reflect underlying mechanisms or behaviors, making time series forecasting an indispensable tool for decision-making processes. Meanwhile, in our current information-dense era, the influence of textual information spans from individual decisions to shaping national directives. However, due to insufficient data accumulation and limited resources, limited work has been done on designing multimodal time series datasets. With the emergence of Large-scale Language Models (LLMs), we are able to fill this gap by proposing an effective pipeline and a new paradigm of time series forecasting task as well as dataset as shown in Figure 1.

Conventional forecasting methods usually only focus on unimodal time series numerical information. Previous works (Wu et al. 2020; Jiang and Luo 2022; Cao et al. 2020)

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

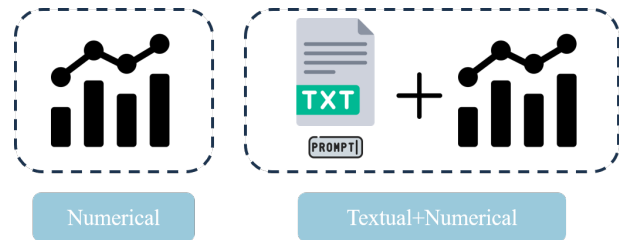


Figure 1: Difference between our datasets and conventional time series dataset. Conventional datasets contain only numerical information, while our dataset leverages textual information as well.

use Graph Neural Networks (GNNs) for short-term predictions, transformer variants (Zhou et al. 2021; Wu et al. 2021; Zhou et al. 2022; Nie et al. 2023) for long-term forecasting. Recently, linear models (Zeng et al. 2023) also exhibit strong performance for time series forecasting tasks. However, these works contain an intrinsic limitation: they overlook the rich contextual information provided by textual data.

While some recent works (Tian Zhou 2023; Sun et al. 2023; Xue and Salim 2022) have tried to apply language models for time series tasks, these works either treat time series data as text sequence inputs (Xue and Salim 2022), or align the time series input with the textual embedding of LLMs. Very few of these approaches utilized a multimodal input containing both time series information and textual information. METS (Li et al. 2023) is the only known work, based on our knowledge, that uses a multimodality input containing both ECG time series data and clinical reports texts. However, METS only focuses on the health-care domain and cannot be generalized to most of the other time series data.

In light of these challenges, we first introduce an innovative pipeline in Figure 2 that leverages the power of large language models (LLMs) to generate textual data along with the time series data: textual information collection, summarization, re-rank, efficient summary based on re-rank similarity. Note that, while not applicable to all domains, this generation pipeline can still be applied to many other fields through proper guidance to obtain extra-textual information,

such as financial forecasting and communication.

In this paper, based on the proposed pipeline, we created a GDEL¹-based multimodal time series forecasting dataset that contains both time series numerical values and textual summaries of events. GDEL database records global events and their associated media coverage, underpinning the profound impact that news can have in guiding our lives. It propels the development of computational communication research (Hopp et al. 2019; Lock 2020) under multiple domains, including social unrest (Galla and Burke 2018), government policies (Schintler and Kulkarni 2014), and financial market (Consoli, Pezzoli, and Tosetti 2020). The establishment of our dataset can enhance the accessibility of different fields to multimodal time series datasets, as well as foster further research in multimodal computational communication analysis.

Together with the established dataset, we propose a prompt tuning-based LLM, GPT4MTS, for time series forecasting with multimodal input, which contrasts with conventional approaches that rely on direct data alignment. For the numerical input, we first split the temporal input into different patches with reversible instance normalization (Nie et al. 2023) and then apply a linear layer to embed the patches into a hidden space as the time series input embedding. For the corresponding textual information, we propose to obtain the textual embeddings through pre-trained language model and treat them as trainable soft prompts prepended to the temporal input. In addition, we choose to freeze the attention layer inside the LLM to speed up the training and inference stage. Experimental results demonstrate the effectiveness of the proposed method and reveal a potential direction that combining the extra information and pure time series data can further enhance the forecasting performance.

Our main contributions can be summarized as follows:

- We propose a general pipeline to incorporate textual data into time series datasets. In addition, we propose the GDEL dataset following the proposed pipeline, which serves as a practical application of our innovative pipeline and methodology.
- Extensive experiments were used to illustrate the effectiveness of our model based on the multimodal time series dataset.

Related Works

In this section, we provide an overview of previous work and discuss differences between our method and related studies.

Large-Scale Language Model (LLM)

The introduction of Transformer architecture (Vaswani et al. 2017) has revolutionized several areas, including Natural Language Processing (NLP) and Computer Vision (CV). Coupled with the rise in computational resources, several Large-scale Language Models (LLMs) based on deep Transformer architecture have been proposed. Among them,

¹gdeltproject.org

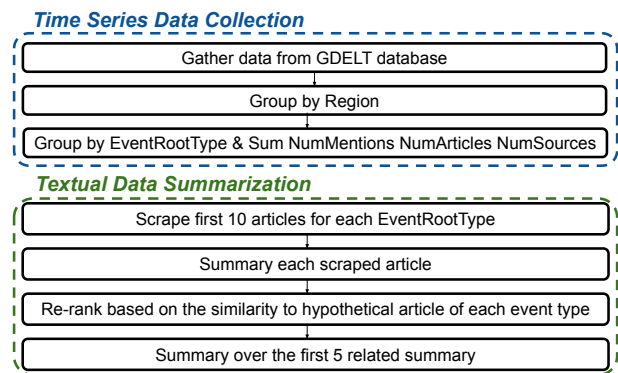


Figure 2: GDEL-based dataset data collection.

BERT is pre-trained on a large corpus and fine-tuned for specific NLP tasks (Devlin et al. 2019). T5 model adopts a unified text-to-text framework, converting every NLP problem into a text generation task (Raffel et al. 2020). GPT focuses on the autoregression generation of text, making it suitable for a variety of generative tasks (Radford et al. 2018, 2019; Brown et al. 2020). LLaMa models are trained on vast public datasets (Touvron et al. 2023). While LLMs have achieved substantial success in NLP, their application in areas such as time series forecasting remains underexplored.

Prompting

Prompting serves as a methodology for crafting queries or commands that steer models toward generating specific, targeted outputs. The CLIP model, for instance, uses textual prompts in the format of “a photo of a object” to perform image classification (Radford et al. 2021). T5 incorporates task descriptions into its text-to-text framework, effectively using these as implicit prompts to guide various NLP tasks (Raffel et al. 2020). Prompting techniques have also been extended to several tasks like object detection (Li et al. 2022), image captioning (Zhang et al. 2022), etc. Distinctively, our work leverages textual prompts to guide the model processing numerical information of time series.

Time Series Forecasting

Time series forecasting holds significance in diverse fields, such as anomaly detection, weather forecasting, and treatment effect modeling. Conventional methods, such as ARIMA (Box and Jenkins 1968), use statistical models to fit time series. As deep learning emerged as a game-changer, several neural architectures, such as CNN-based (Munir et al. 2018), RNN-based (Salinas et al. 2020), and Transformer-based (Liu et al. 2021; Wu et al. 2021; Zhou et al. 2021, 2022), demonstrated superiority. While Transformer-based methods have shown impressive performance, there are studies that juxtapose them with embarrassingly simple linear models, challenging their priority (Zeng et al. 2023). In contrast to these approaches, our study utilizes a pre-trained LLM to address the challenges of time series forecasting.

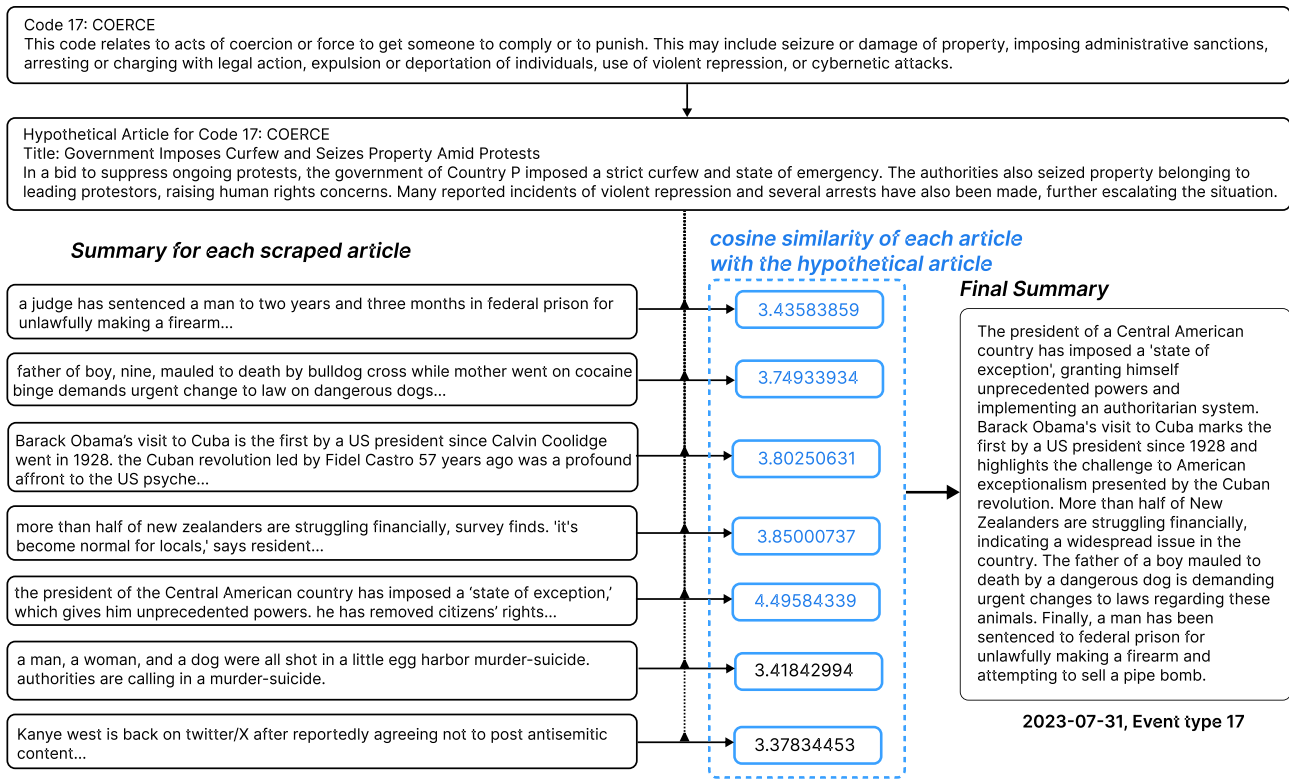


Figure 3: One example for textual information generation under our pipeline.

Furthermore, the prevalent benchmark datasets, like Weather and ETT (Zhou et al. 2021), predominantly focus on numeric series. Our introduction of the GDELT dataset diverges from this trend. By incorporating textual content into time series, we enrich the data, paving the way for LLMs to harness additional textual information when making predictions based on time series data.

LLM4TS

Several studies have ventured into using LLMs for time series forecasting. For instance, one work interprets time series as patches and employs pre-trained models for forecasting (Tian Zhou 2023). Another research direction involves aligning embeddings between time series and texts. The TEST method adopts contrastive learning to align time series embeddings with the original textual embeddings in LLMs (Sun et al. 2023). TimeLLM (Jin et al. 2023) adopts a distinct approach through reprogramming that aligns the embedding space of time series data with that of textual data. Other works convert time series into textual data via prompts (Xue and Salim 2022), while some rely exclusively on textual information (Yu et al. 2023). Our approach similarly harnesses a pre-trained LLM for time series forecasting but differentiates by using prompts to feed additional textual information. Combined with the capabilities of LLMs and the integration of data from both modalities, this approach boost the performance of time series forecasting beyond using either alone, emphasizing the importance of both modalities.

Methods

Dataset Collection

This dataset is derived from the Global Database of Events, Language, and Tone (GDELT) database², which is one of the largest publicly available databases that monitor news media from around the world in over 100 languages. The GDELT database covers various types of information primarily focused on events, thereby offering a rich set of variables for understanding global societal trends.

Time series data collection. The pipeline of our data collection process is shown in Figure 2. In our specific GDELT dataset, we focus on extracting key information related to the top 10 popular event types (EventRootCode) and their respective mentions and coverage in the news media. The relationship between EventRootCode and the name of the event can be looked up through Table 3. Specifically, we extract three key variables for forecasting: NumMentions, NumArticles, NumSources. These variables respectively represent the number of mentions, the number of related articles, and the number of sources, which are all relevant to the attention a particular event type receives within a given time frame and geographical region. We divided our dataset into 10 event root types, collecting data for 55 regions under the US and national data for the US. Currently, the data we used for training and evaluation spans from 2022-08-17 to 2023-07-31.

²gdeltproject.org

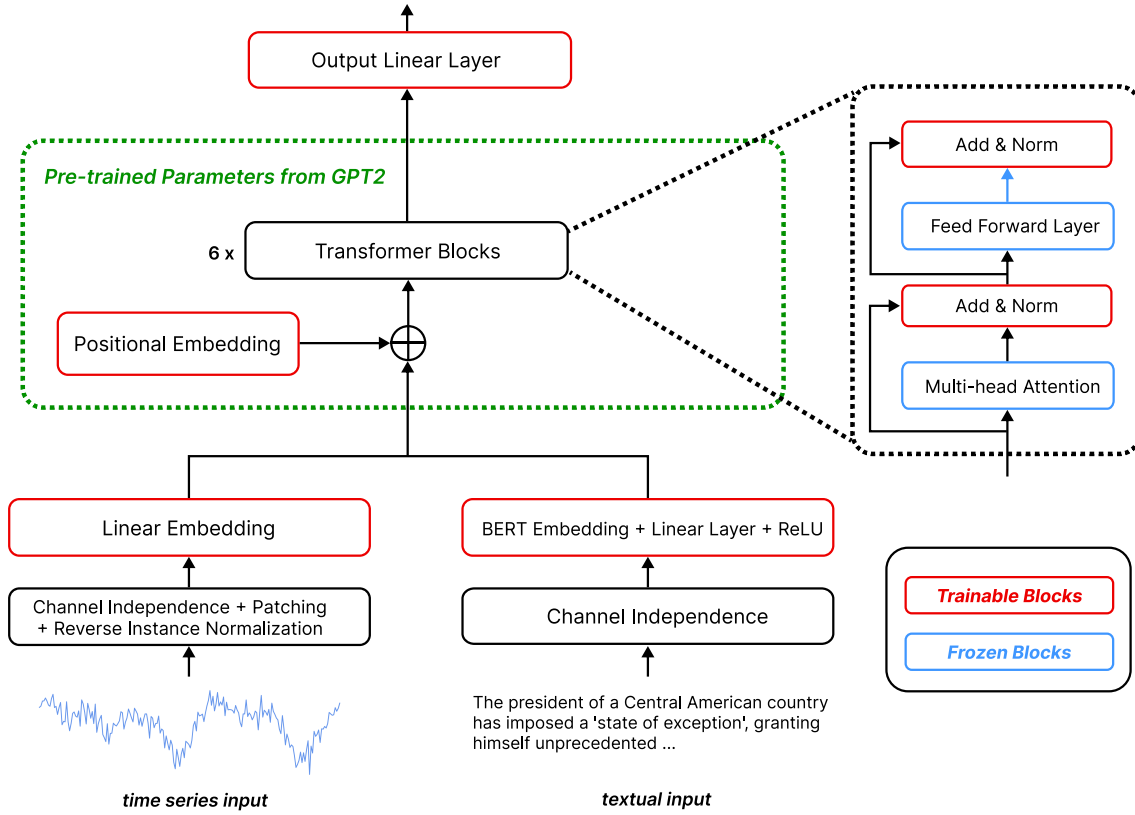


Figure 4: Overview of our proposed model.

Corresponding textual data collection. For textual data summarization, we first scrape 10 articles for each event type under the given region and given date. Then a summary of each scraped article is generated using T5 (Raffel et al. 2020). A hypothetical article summary is generated given the particular event type and its explanations, serving as a template for possible summaries under this event type. Summaries are then re-ranked based on the similarity to the hypothetical article, aiming to keep the summaries that are most relevant to the news under the event type. Finally, an overall summary is generated given the top 5 related article summaries using OpenAI ChatGPT3.5 API ³.

We demonstrate one example of how textual information is collected for our dataset in Figure 3. While the first round of summary for each scraped article may contain redundant information (such as the last summary in our example) as they are processed through T5, by utilizing similarity and re-ranking, they are less likely to be considered as top-related articles to be included into our final summary. Due to a limited budget, for some regional information, we use T5 instead of OpenAI ChatGPT3.5 API for the final summary of regional data. This can lead to certain mal-summarized texts due to the interference of useless information in the first round of summary. Therefore, to mitigate this issue, we replace these inferior final summaries with summaries of clean scraped titles.

³<https://platform.openai.com/docs/guides/gpt>

The primary goal of this dataset is to help individuals, researchers, and policymakers better predict and analyze the level of attention certain types of events are receiving and the impact they have in a specific region. This can facilitate more informed decision-making in both the public and private sectors as well as provide a concise and effective dataset for computational communication researchers.

This dataset also verifies the effectiveness of our multimodal time series dataset generation pipeline, enhancing the feasibility of establishing similar datasets for other domains.

Problem Definition

We consider the following problem, given a collection of multivariate time series samples and textual summaries with look-back window $L : ([x_1, s_1], \dots, [x_L, s_L])$ where x_t at time step t is a vector of dimension M and s_t at time step t is a textual summary, we would like to forecast T future values: $(x_{L+1}, \dots, x_{L+T})$.

Previous time series models learn from unimodal information: they learn a function f_{uni} such that

$$(x_{L+1}, \dots, x_{L+T}) = f_{uni}([x_1, \dots, x_L]|\theta), \quad (1)$$

where θ is the parameters of the model. On the contrary, our model leverages multimodal information: we learn a function f_{multi} such that

$$(x_{L+1}, \dots, x_{L+T}) = f_{multi}([x_1, s_1], \dots, [x_L, s_L]|\theta). \quad (2)$$

Event	DLinear	NLinear	PatchTST	Autoformer	Informer	Transformer	FEDformer	LLaMa	GPT4TS	Ours
01	0.601	0.597	0.649	0.655	2.187	1.699	<u>0.561</u>	0.590	0.582	0.552
02	0.197	0.196	0.209	0.208	0.354	0.275	<u>0.193</u>	0.194	0.197	0.189
03	0.217	0.215	0.226	0.222	0.302	0.293	0.208	0.216	0.223	<u>0.213</u>
04	1.436	1.430	1.542	1.700	7.558	6.518	1.483	1.405	<u>1.382</u>	1.342
05	0.238	<u>0.234</u>	0.253	0.244	0.362	0.296	0.234	0.236	<u>0.237</u>	0.230
07	0.127	<u>0.125</u>	0.127	0.138	0.180	0.321	0.135	0.125	0.129	0.121
08	0.085	0.083	0.086	0.100	0.201	0.405	0.103	<u>0.083</u>	0.085	0.080
11	0.179	0.177	0.188	0.182	0.213	0.271	<u>0.173</u>	<u>0.177</u>	0.177	0.165
17	0.204	0.202	0.213	0.201	0.258	0.298	<u>0.196</u>	0.199	0.200	0.190
19	0.423	0.421	0.435	0.391	0.439	0.449	0.396	0.411	0.403	<u>0.391</u>
Average	0.371	0.368	0.392	0.404	1.205	1.083	0.368	0.363	<u>0.362</u>	0.347

Table 1: Group by Event, MSE. Bold indicates the best result, while Underlines indicates the second best.

Event	DLinear	NLinear	PatchTST	Autoformer	Informer	Transformer	FEDformer	LLaMa	GPT4TS	Ours
01	0.566	0.564	0.601	0.615	1.207	1.028	0.562	0.551	<u>0.546</u>	0.544
02	0.332	0.330	0.347	0.349	0.456	0.413	0.342	0.323	0.325	<u>0.325</u>
03	0.353	<u>0.349</u>	0.366	0.366	0.424	0.434	0.358	0.348	0.353	0.352
04	0.910	0.911	0.959	1.017	2.379	2.162	0.942	0.889	0.879	<u>0.881</u>
05	0.370	0.365	0.388	0.379	0.445	0.412	0.378	<u>0.364</u>	0.363	0.367
07	0.256	<u>0.249</u>	0.255	0.276	0.340	0.493	0.282	0.250	0.255	0.249
08	0.213	<u>0.207</u>	0.214	0.240	0.387	0.585	0.255	<u>0.206</u>	0.210	0.206
11	0.311	0.306	0.323	0.321	0.359	0.430	0.323	<u>0.305</u>	0.305	0.298
17	0.344	0.340	0.357	0.348	0.398	0.452	0.349	0.337	<u>0.337</u>	0.331
19	0.416	0.411	0.425	0.407	0.425	0.479	0.420	0.407	<u>0.405</u>	0.390
Average	0.407	0.403	0.424	0.432	0.682	0.689	0.421	0.398	<u>0.398</u>	0.394

Table 2: Group by Event, MAE. Bold indicates the best result, while Underlines indicates the second best.

Proposed Method

The architecture of the model we employed is demonstrated in Figure 3. We utilize parameters from the pretrained GPT2 model (Radford et al. 2019). In order to understand information from both modalities, we add extra prompt layers to transform time series information and textual information to the input dimension of the pretrained model.

Frozen Pretrained Model Our model retains the positional embedding layers and transformer blocks from the pretrained GPT2 model, while we freeze the attention layers and the feed-forward layers and fine-tune the positional embeddings and layer normalization layers, following the standard practice (Houlsby et al. 2019; Lu et al. 2022).

Input Embedding In order to apply both textual input and time series input to pretrained LLM, we need to prepare the two modalities through separate embedding layers.

For the textual information, we apply the BERT (Devlin et al. 2019) embedding module as a feature extractor to obtain the representation for the summary texts over the look-back window.

For the time series information, following previous works (Tian Zhou 2023; Nie et al. 2023), we use the following operations to fit the time series input into the pre-trained GPT2 model:

- We apply Reversible Instance Normalization (Kim et al. 2021) to mitigate the distribution shift of the time series data over time, which performs normalization by extracting mean and variance from input time series and adding them back to the predicted output.

- We then apply patching (Nie et al. 2023) by aggregating adjacent timestamps. This enables time series stamps to gather context, similar to the continual impact of news over a period of time.
- Since channel-independence has demonstrated its effectiveness through previous works (Nie et al. 2023; Tian Zhou 2023), we treat each multivariate time series as multiple independent uni-variate series as well.

Output Layer Since the output from the frozen Pretrained Language Transformer contains hidden states of sequence length (for textual inputs) + patch number (for time series inputs), we apply a linear output layer, which takes the hidden state corresponding to the time series as input and transform it into the desired prediction length. As the ablation study in Table 5 and Table 4 show, this improves performance, as utilizing only the hidden states of the time series input can preserve a better representation of the numerical forecasting targets.

Experiments

Baselines and Experimental Settings

To evaluate the model’s performance on our generated dataset, we choose four SOTA *Transformer-based models*, including FEDformer (Zhou et al. 2022), Autoformer (Wu et al. 2021), Informer (Zhou et al. 2021), PatchTST (Nie et al. 2023), and two *Linear-based models*, including DLinear and NLinear (Zeng et al. 2023) as our baselines. We also include two *pre-trained large language models*, LLaMA (Touvron et al. 2023), and GPT(Radford et al. 2019) as

Event Number	Event Type Name
01	Make Public Statement
02	Appeal
03	Express Intent to Cooperate
04	Consult
05	Engage in Diplomatic Cooperation
07	Provide Aid
08	Yield
11	Disapprove
17	Coerce
19	Fight

Table 3: Event number to event type Name table

our baselines, with both adapted to time series forecasting tasks in a similar way as GPT4TS (Tian Zhou 2023). All of these models follow the same experimental setup with a prediction length $T = 7$, and a look-back window size $L = 15$ to forecast all three numerical time-series variables: NumMentions, NumArticles, NumSources. Under this setting, by dividing data to train, validate, and test with a ratio of 7 : 2 : 1, we gathered 343,200 training examples, 98,976 validation examples, and 41,904 testing examples.

Main Results

Table 1 and Table 2 show the multivariate forecasting results. The corresponding event type name for each event number is in Table 3. Overall, our model achieves the best average performance among the 10 event types for news and steadily exceeds the performance of other SOTA models in most event types. Quantitatively, while GPT4TS already achieves a SOTA performance, by using textual information as prompts to guide time series forecasting, our model achieves an overall 4.14% reduction on MSE and 1.0% reduction on MAE.

Ablation Study

We study the effects of only taking hidden representations from the time series inputs to the output layers. By comparing our methods with taking full hidden representations (from both textual inputs and time series inputs) in Table 4 and Table 5, one can observe that the performance is polluted by the hidden representations of textual inputs, though it is still better than without textual information (GPT4TS). This demonstrates the effectiveness of only taking the hidden representations from time series inputs into our model.

Analysis

The improvement in performance of our model is understood as a collaborative result of numerical data, textual information, and a prompt-based methodology for multimodal integration:

- Textual information provides contexts that numerical data alone cannot capture. For instance, in our dataset, news articles featuring prominent figures or with strong sentiments or characteristics can obtain a more eminent representation and therefore influence the fluctuation or trend in the numerical data.

Event Number	GPT4TS	Ours	full selection
01	0.582	0.552	0.575
02	0.197	0.189	0.197
03	0.223	0.213	0.224
04	1.382	1.342	1.400
05	0.237	0.230	0.239
07	0.129	0.121	0.122
08	0.085	0.080	0.084
11	0.177	0.165	0.169
17	0.200	0.190	0.196
19	0.403	0.391	0.389
Average	0.362	0.347	0.360

Table 4: Ablation studies for GPT4MTS models by MSE.

Therefore, our model is able to outperform other unimodality models. As shown in the ablation study, even GPT4MTS with full selection can achieve a better MSE score and a similar MAE score compared to GPT4TS, which already outperforms all other unimodal time series models. This demonstrates the benefits of having textual context for time series forecasting tasks.

- Numerical data offers precise quantitative values, which are indispensable for forecasting tasks. It uncovers the statistical patterns inherent in the dataset. Given that our forecasting targets are also represented numerically, this data primarily steers the predictions of our model. As observed from table 1 and table 2, when only utilizing numerical data, linear-based models (DLinear, NLinear) generally outperform transformer-based models (Informer, Autoformer, Transformer). This indicates that the transformer models may over-complicate the observed numerical information in our dataset with their complex architecture. However, pre-trained LLM models, (LLaMA and GPT4TS) tend to achieve a better performance than all other models, which demonstrates the effectiveness of cross-modality adaptation of these pre-trained Large Language models for time series forecasting tasks.
- The methodology of prompt-based integration for textual and numerical inputs is crucial. Comparative analyses between GPT4TS and our model highlight the efficacy of harnessing additional text representations as prompts. Concurrently, our ablation study demonstrates that these prompts should act as guiding factors, not as intrinsic components of the output representation. This approach facilitates the learning ability of LLMs in processing numerical data without being overshadowed by textual information, ensuring moderate assimilation of contextual information. Over-reliance on textual information, however, as demonstrated through the ablation study, could harm the predicting performance.

Relevance for Accessibility in Communication

In our current era of information abundance, accessibility to this wealth of data becomes crucial, especially for individuals in underdeveloped regions, those with reading disabilities, and non-English speakers. Our research directly

Event Number	GPT4TS	Ours	full selection
1	0.546	0.544	0.557
2	0.325	0.325	0.333
3	0.353	0.352	0.359
4	0.879	0.881	0.907
5	0.363	0.367	0.373
7	0.255	0.249	0.250
8	0.210	0.206	0.210
11	0.305	0.298	0.302
17	0.337	0.331	0.335
19	0.405	0.390	0.393
Average	0.398	0.394	0.402

Table 5: Ablation studies for GPT4MTS models by MAE.

contributes to enhancing communication accessibility by proposing a multimodal time series forecasting pipeline and dataset, which synergizes numerical data with summarized textual information. This integration enables individuals to comprehend complex global events and news in a more accessible and time-efficient manner.

Our approach effectively addresses language and textual barriers by providing summarized news and key information through numerical data. This enhances the global news reach, especially beneficial for non-English speakers and minority language users. It provides potential for translated summaries through LLMs and utilizes numerical data to highlight the impact and forecast of various events. This method not only broadens communication channels but also ensures that individuals without access to traditional news sources can stay informed through oral communication or general short summaries. It is particularly useful for those who are illiterate, offering a way to understand world events through numerical insights on event impacts as well as the forecasting results for each kind of event.

In fields like finance, our pipeline and methodology can empower individuals with limited time encountering accessibility challenges by providing concise and comprehensible summaries of complex financial reports, aiding in informed decision-making. Meanwhile, the forecasting results can provide insights for future trend of events, therefore, beneficial for the financial decisions.

Additionally, our dataset and pipeline are invaluable for computational communication research. They facilitate a deeper understanding of information dissemination and its impact on different populations. This understanding is critical for developing strategies to enhance news accessibility and comprehension across diverse groups.

This streamlined approach promises to make information more accessible and understandable, breaking down barriers and fostering a more inclusive information ecosystem.

Conclusion and Future Work

This paper introduces an effective pipeline for collecting textual information for time series forecasting tasks: a dataset is collected under this pipeline, and a corresponding multimodal time series forecasting model is proposed. Through extensive experiments and analysis, we show that

our dataset and pipeline are reliable and we prove the capability of improving prediction performance utilizing our model, GPT4MTS, based on multimodal inputs. This further testifies to the benefits of the extra textual information collected under our pipeline.

We provide a general pipeline for multimodal dataset generation and accordingly establish a dataset for the news domain. Our work also demonstrate the potential of both the dataset and our model as foundational elements for future multimodal time series forecasting research. A crucial next step in this field would be to further explore multimodal time series forecasting, particularly leveraging the capabilities of Large Language Models (LLMs). This direction promises significant advancements in the understanding and application of multimodal data in forecasting tasks.

References

- Box, G. E.; and Jenkins, G. M. 1968. Some recent advances in forecasting and control. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 17(2): 91–109.
- Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901.
- Cao, D.; Enouen, J.; Wang, Y.; Song, X.; Meng, C.; Niu, H.; and Liu, Y. 2023. Estimating Treatment Effects from Irregular Time Series Observations with Hidden Confounders. In Williams, B.; Chen, Y.; and Neville, J., eds., *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, 6897–6905. AAAI Press.
- Cao, D.; Wang, Y.; Duan, J.; Zhang, C.; Zhu, X.; Huang, C.; Tong, Y.; Xu, B.; Bai, J.; Tong, J.; and Zhang, Q. 2020. Spectral Temporal Graph Neural Network for Multivariate Time-series Forecasting. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.
- Consoli, S.; Pezzoli, L. T.; and Tosetti, E. 2020. Using the GDELT Dataset to Analyse the Italian Sovereign Bond Market. In *Machine Learning, Optimization, and Data Science: 6th International Conference, LOD 2020, Siena, Italy, July 19–23, 2020, Revised Selected Papers, Part I 6*, 190–202. Springer.
- Devlin, J.; Chang, M.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Burstein, J.; Doran, C.; and Solorio, T., eds., *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, 4171–4186. Association for Computational Linguistics.

- Galla, D.; and Burke, J. 2018. Predicting social unrest using GDELT. In *International conference on machine learning and data mining in pattern recognition*, 103–116. Springer.
- Hopp, F. R.; Schaffer, J.; Fisher, J. T.; and Weber, R. 2019. iCoRe: The GDELT interface for the advancement of communication research. *Computational Communication Research*, 1(1): 13–44.
- Houlsby, N.; Giurgiu, A.; Jastrzebski, S.; Morrone, B.; De Laroussilhe, Q.; Gesmundo, A.; Attariyan, M.; and Gelly, S. 2019. Parameter-efficient transfer learning for NLP. In *International Conference on Machine Learning*, 2790–2799. PMLR.
- Jiang, W.; and Luo, J. 2022. Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, 207: 117921.
- Jin, M.; Wang, S.; Ma, L.; Chu, Z.; Zhang, J. Y.; Shi, X.; Chen, P.; Liang, Y.; Li, Y.; Pan, S.; and Wen, Q. 2023. Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. *CoRR*, abs/2310.01728.
- Kalamara, E.; Turrell, A.; Redl, C.; Kapetanios, G.; and Kapadia, S. 2022. Making text count: economic forecasting using newspaper text. *Journal of Applied Econometrics*, 37(5): 896–919.
- Kaushik, S.; Choudhury, A.; Sheron, P. K.; Dasgupta, N.; Natarajan, S.; Pickett, L. A.; and Dutt, V. 2020. AI in healthcare: time-series forecasting using statistical, neural, and ensemble architectures. *Frontiers in big data*, 3: 4.
- Kim, T.; Kim, J.; Tae, Y.; Park, C.; Choi, J.-H.; and Choo, J. 2021. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*.
- Li, J.; Liu, C.; Cheng, S.; Arcucci, R.; and Hong, S. 2023. Frozen Language Model Helps ECG Zero-Shot Learning. *CoRR*, abs/2303.12311.
- Li, L. H.; Zhang, P.; Zhang, H.; Yang, J.; Li, C.; Zhong, Y.; Wang, L.; Yuan, L.; Zhang, L.; Hwang, J.-N.; et al. 2022. Grounded language-image pre-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10965–10975.
- Liu, S.; Yu, H.; Liao, C.; Li, J.; Lin, W.; Liu, A. X.; and Dustdar, S. 2021. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *International conference on learning representations*.
- Lock, I. 2020. Debating glyphosate: A macro perspective on the role of strategic communication in forming and monitoring a global issue arena using inductive topic modelling. *International Journal of Strategic Communication*, 14(4): 223–245.
- Lu, K.; Grover, A.; Abbeel, P.; and Mordatch, I. 2022. Frozen pretrained transformers as universal computation engines. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 7628–7636.
- Munir, M.; Siddiqui, S. A.; Dengel, A.; and Ahmed, S. 2018. DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. *Ieee Access*, 7: 1991–2005.
- Nguyen, T.; Brandstetter, J.; Kapoor, A.; Gupta, J. K.; and Grover, A. 2023. ClimaX: A foundation model for weather and climate. In Krause, A.; Brunskill, E.; Cho, K.; Engelhardt, B.; Sabato, S.; and Scarlett, J., eds., *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, 25904–25938. PMLR.
- Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2023. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.
- Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PMLR.
- Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I.; et al. 2018. Improving language understanding by generative pre-training.
- Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I.; et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8): 9.
- Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; and Liu, P. J. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1): 5485–5551.
- Salinas, D.; Flunkert, V.; Gasthaus, J.; and Januschowski, T. 2020. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3): 1181–1191.
- Schintler, L. A.; and Kulkarni, R. 2014. Big data for policy analysis: The good, the bad, and the ugly. *Review of Policy Research*, 31(4): 343–348.
- Sezer, O. B.; Gudelek, M. U.; and Ozbayoglu, A. M. 2020. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied soft computing*, 90: 106181.
- Sun, C.; Li, Y.; Li, H.; and Hong, S. 2023. TEST: Text Prototype Aligned Embedding to Activate LLM’s Ability for Time Series. *arXiv preprint arXiv:2308.08241*.
- Tian Zhou, X. W. L. S. R. J., Peisong Niu. 2023. One Fits All: Power General Time Series Analysis by Pretrained LM. In *NeurIPS*.
- Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; Rodriguez, A.; Joulin, A.; Grave, E.; and Lample, G. 2023. LLaMA: Open and Efficient Foundation Language Models. *CoRR*, abs/2302.13971.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

- Wu, H.; Xu, J.; Wang, J.; and Long, M. 2021. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, 34: 22419–22430.
- Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Chang, X.; and Zhang, C. 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 753–763.
- Xue, H.; and Salim, F. D. 2022. Prompt-Based Time Series Forecasting: A New Task and Dataset. *arXiv preprint arXiv:2210.08964*.
- Yu, X.; Chen, Z.; Ling, Y.; Dong, S.; Liu, Z.; and Lu, Y. 2023. Temporal Data Meets LLM—Explainable Financial Time Series Forecasting. *arXiv preprint arXiv:2306.11025*.
- 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, 11121–11128.
- Zhang, H.; Zhang, P.; Hu, X.; Chen, Y.-C.; Li, L.; Dai, X.; Wang, L.; Yuan, L.; Hwang, J.-N.; and Gao, J. 2022. Glipv2: Unifying localization and vision-language understanding. *Advances in Neural Information Processing Systems*, 35: 36067–36080.
- Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 11106–11115.
- Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; and Jin, R. 2022. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International Conference on Machine Learning*, 27268–27286. PMLR.