Exploiting Language Power for Time Series Forecasting with Exogenous Variables

Qihe Huang
University of Science and Technology of China
Hefei, China
hqh@mail.ustc.edu.cn

Zhengyang Zhou*
University of Science and Technology of China
Hefei, China
Suzhou Institute for Advanced Research, USTC
Suzhou, China
zzy0929@ustc.edu.cn

Kuo Yang
University of Science and Technology of China
Hefei, China
yangkuo@mail.ustc.edu.cn

Yang Wang*
University of Science and Technology of China
Hefei, China
Suzhou Institute for Advanced Research, USTC
Suzhou, China
angyan@ustc.edu.cn

Abstract

The World Wide Web thrives on intelligent services that depend heavily on accurate time series forecasting to navigate dynamic and evolving environments. Due to the partially-observed nature of real world, exclusively focusing on the target of interest, so-called *endogenous variables*, is insufficient for accurate forecasting, especially in web systems that are susceptible to external influences. Thus, utilizing *exogenous variables* to harness external information, i.e., forecasting with exogenous variable (FEV), is imperative. Nevertheless, as the external environment is complex and ever-evolving, inadequately capturing external influences can even lead to learning spurious correlations and invalid prediction. Fortunately, recent studies have demonstrated that large language models (LLMs) exhibit exceptional recognition capabilities across open real-world systems, including a deep understanding of exogenous environments. However, it is difficult to directly apply LLMs for FEV due to challenges of task activation, exogenous knowledge extraction, and feature space alignment. In this work, we devise ExoLLM, an LLM-driven method to sufficiently utilize Exogenous variables for time series forecasting. We begin by Meta-task Instruction to activate the knowledge transfer of LLM from natural language processing to FEV. To comprehensively understand the intricate and hierarchical influences of exogenous variables, we propose Multi-grained Prompts, encompassing diverse external influences, including natural attributes, trend correlations, and period relationships between two types of variables. Additionally, a Dual TS-Text Attention is devised to bridge the feature gap between text and numeric data in LLM. Evaluation on real-world datasets demonstrates ExoLLM's superiority in exploiting exogenous information for forecasting with open-world language knowledge.

CCS Concepts

• **Mathematics of computing → Time series analysis.**

Keywords

Time Series Forecasting, Language Power, Modality Alignment, Forecasting with Exogenous Variables

ACM Reference Format:
Qihe Huang, Zhengyang Zhou, Kuo Yang, and Yang Wang. 2025. Exploiting Language Power for Time Series Forecasting with Exogenous Variables. In *Proceedings of the ACM Web Conference 2025 (WWW '25), April 28–May 2, 2025, Sydney, NSW, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3696410.3714793

1 Introduction

The World Wide Web, as a continuously and ever-changing physical system, heavily depends on the ability to forecast and respond to shifting patterns and user behaviors [16, 17, 20, 37]. Time series forecasting is essential to modern web technologies, utilizing historical data to anticipate future web patterns and trends [29, 34, 59]. Its predictive accuracy not only enhances user experience but also drives the development of intelligent web services, ranging from personalized content recommendations [36] and web economics modeling [5] to microservice log analysis [11]. These capabilities position time series forecasting as a cornerstone in creating adaptive, data-driven web platforms [32, 55].

Recently, deep models have achieved promising progress in time series forecasting [3, 13, 14, 21, 24–27, 40–42, 50, 63, 64, 70], with most of them focusing exclusively on the target of interest, known as *endogenous variables*, to make predictions [30, 35, 49, 51, 54, 66]. This approach often ignores the influence of *exogenous variables* from the external environment. **Exogenous variables refer to observable data within a system that are not the target variable being predicted.** As shown in Figure 1 (a), the variations within

*Zhengyang Zhou and Yang Wang are corresponding authors.

WWW '25, April 28–May 2, 2025, Sydney, NSW, Australia.

ACM ISBN 979-8-4007-1274-6/25/04
https://doi.org/10.1145/3696410.3714793

web page views (endogenous variable) are often influenced by exogenous variables, such as traffic flow, hospitalization rate, and societal events [47, 53, 62]. Thus, given the complex and changing physical system [23, 46, 48, 58, 61, 71], incorporating exogenous factors, i.e., forecasting with exogenous variables (FEV) is becoming prevalent and indispensable [33]. Generally, the core of FEV is to effectively model the influence of exogenous variables on endogenous variable [4, 22, 28, 39]. Recent research in FEV proposes using attention among observed numerical exogenous series and endogenous series to capture this inherent relationship [33, 53]. Nevertheless, due to the *Intricate influences and interactions* from external environment, relying solely on time series modality is insufficient for capturing these external influences: **(1) Multi-grained temporal dependencies** [19]. The external influences and interactions from exogenous variables is multi-grained, such as periodicity and trends, which can be reflected by various aspects including complex human behaviors and living habits [68]. It is difficult to model such changing and diverse impact only by observed numeric [29], highlighting the necessity of thoroughly learning multi-grained temporal features to effectively model these intricate patterns [19]. **(2) Spurious correlation** [12]. Noise and interventions in current data can lead to learning biased external influence, thereby affecting the accuracy of forecasting results [45]. For example, traffic flows are positively correlated with exogenous weather variables, but mandatory controls can lead to less traffic even when the weather is good, resulting in spurious association that may be learned by models. Without any external knowledge from real world, a high prediction uncertainty tends to be inevitable [69].

Consequently, designing more intelligent and robust FEV framework that enable models to effectively understand the intricate external influence and avoid spurious correlation is in demand. Fortunately, with rapid development of large language models (LLMs) [7–9, 15, 52], there have been more opportunities to leverage the vast language knowledge to comprehend external influence on endogenous variables. Through extensive training on large-scale text corpora, pre-trained LLMs have extensively acquired knowledge of multi-grained correlation between two types of variables. Intuitively, empowering FEV with these full-scale external knowledge can significantly enhance forecasting accuracy [67]. Nevertheless, as shown in Figure 1 (b), considering distinct task differences between NLP and time series forecasting [2, 67], and distant data gap between discrete text and continues numeric [19], employing LLMs to FEV faces several urgent challenges: **(1) Task activation.** How to construct task instruction to fully activate the potential of LLMs in FEV, enabling the knowledge transfer across tasks. **(2) Full-scale language-driven knowledge acquirement.** Given an LLM-based solution, how to devise effective and comprehensive prompts to acquire hierarchical and sufficient knowledge from exogenous variables. **(3) Feature space alignment.** Given the solution is concerned with two data modalities of both numerical and text data, how to construct a feasible encoding-decoding strategy to ensure the alignment between text space and time series space.

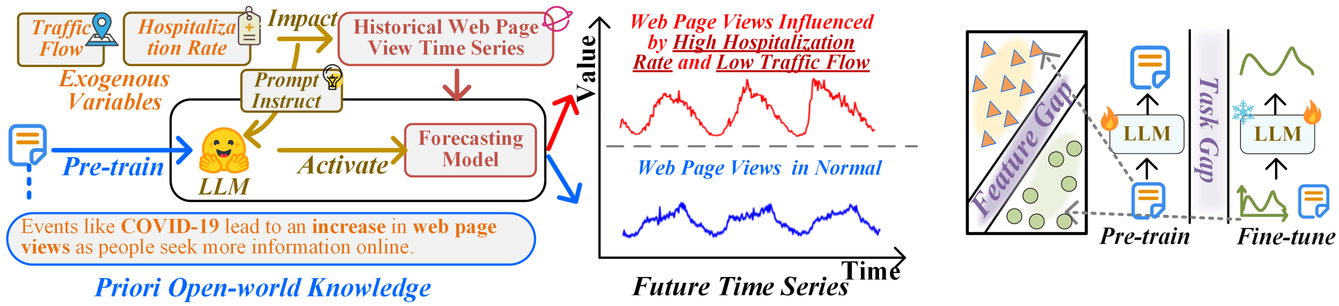
In this work, we devise ExoLLM to forecast with Exogenous variables using LLM, capturing diverse and changing external influences from exogenous variables with language-based knowledge. Technically, we elaborately craft domain-specific Meta-task Instructions to guide LLMs to process FEV tasks in different data domains. Subsequently, we establish Multi-grained Prompts to dynamically capture the natural attributes, periodic associations, trend correlations, and other granular external influence of exogenous variables, thereby adaptive transferring the dynamic auxiliary information into knowledge that can be understood by ExoLLM. Additionally, we design the Dual Time series-Text Attention Attention (DT^{2}Attention) to mitigate data discrepancies during time series encoding and feature decoding, respectively. Comprehensive evaluation demonstrates that LLM can even act as an effective few-shot and zero-shot FEV learners when adopted through our elaborate design, outperforming specialized forecasting models. Our meticulous design enables LLMs to function even as a proficient few-shot and zero-shot FEV learner, surpassing specialized forecasting models in terms of effectiveness, as demonstrated by the comprehensive evaluation. Our contributions can be summarized as follows:

- Given the complex and evolving external environment of real-world system, i.e., web service, traffic, electricity and weather, we introduce LLMs to maximally explore the auxiliary information of exogenous variables.
- We propose ExoLLM, the first LLM-based forecasting model to accomplish FEV:
  1) To fully exploit the potential of LLM in FEV, we elaborately design Meta-task Instruction and Multi-grained Prompt, realizing the pre-trained knowledge transfer from NLP to FEV and integrate dynamic context information into knowledge of time-series domain.
  2) To deal with the distant data gap between discrete text and continues numeric, we design modality-aware encoding and decoding mechanisms, i.e., DT^{2}Attention, to achieve aligned feature before and after LLM encoding.
- ExoLLM demonstrates outstanding predictive performance across various real scenarios, including long-term, short-term, few-shot, and zero-shot forecasting. Quantitatively, ExoLLM outperforms 10 state-of-the-art models for long-term forecasting, achieving top-1 performance in 51 settings and top-2 in 5 settings out of a total of 56 settings. In addition, ExoLLM reduces MAE by an average of 4.1%, 5.2%, and 4.5% in short-term, few-shot, and zero-shot forecasting tasks, respectively.

2 Related Work

2.1 Forecasting with Exogenous Variables

In practical forecasting scenarios, the utilization of exogenous variables as auxiliary information for forecasting endogenous variables is more prevalent. Previous research has explored statistical methods such as ARIMAX [56] and SARIMAX [44], which understand relationships between exogenous and endogenous series along with auto-regression. Additionally, deep learning models like NBEATSx [38] and TiDE [6] argue that forecasting models can leverage future values of exogenous variables during the forecasing endogenous variables. Notably, TimeXer [53] introduces external information into transformer architectures through well-designed embedding strategies to effectively incorporate external information into segmented representations of endogenous variables, accommodating temporal lags or missing data records. However, these approaches rely on establishing auxiliary information only based on



(a)LLM-empowered Forecasting with Exogenous Variables (b) Gap in Feature and Task

Figure 1: (a) Illustration of Knowledge Reserve from Pre-trained LLM: The extensive pre-trained text data endows LLMs with the potential to understand intricate influence of exogenous variables on web page views. (b) Huge Gaps in Feature Space and Tasks: Text embeddings and time series features are usually mapped to different feature spaces, and it is challenging to fine-tune text-generation pre-trained LLM for FEV.

Table 1: Comparison between prior LLM-based time series forecasting models and ExoLLM.

Method	**ExoLLM (Ours)**	AutoTimes [2024]	TimeLLM [2023]	LLM4TS [2023]	UniTime [2024]	LLMTime [2023]	TEST [2023]	TEMPO [2023]	GPT4TS [2023]
Exogenous Variables	✓	✗	✗	✗	✗	✗	✗	✗	✗
Multimodal	✓	✓	✓	✗	✗	✗	✓	✓	✗
Feature Alignment	✓	✗	✓	✗	✗	✗	✗	✗	✗

numeric correlation between exogenous and endogenous series. In contrast, ExoLLM has the capability to extract multi-grained effects of exogenous variables on endogenous ones as auxiliary information from extensive world knowledge, thereby holding significant potential for enhancing accuracy and generalization in FEV.

2.2 LLM-based Forecasting

The recent emergence of LLMs has opened up new possibilities for time series forecasting [29, 31]. GPT4TS [67] utilizes a pre-trained language model without updating its self-attention and feedforward layers. The model undergoes fine-tuning and evaluation across various time series analysis tasks, demonstrating comparable or state-of-the-art performance by leveraging knowledge transfer from natural language pre-training. LLM4TS [2] adopts a two-stage fine-tuning approach on the LLM to fully leverage time series data. TimeLLM [19] introduces the concept of text prototypes and reprograms time series based on these prototypes to align them more naturally with language models. Tempo [1] decomposes the trend, seasonality, and residual components of time series while dynamically selecting prompts to address comprehension challenges for LLMs. UniTime [29] proposes a language-empowered unified model to efficiently capture knowledge from cross-domain time series data. With their extensive knowledge base, LLMs exhibit tremendous potential in time series forecasting. However, as shown in Table 1, there has been no prior research exploiting LLM for forecasting with exogenous variables (FEV) to enhance the prediction accuracy. To address this gap, we propose ExoLLM which harnesses the power of language to capture the influence of exogenous variables on endogenous variables.

3 Problem Definition

In forecasting with exogenous variables, there is a historical endogenous series $\mathbf{X} \in \mathbb{R}^{1\times L}$ and its associated exogenous information $\mathbf{E}$, where L is look-back window size. Concretely, $\mathbf{E} \in \mathbb{R}^{M\times L}$ comprises multiple exogenous variables $\{\mathbf{E}^{(1)}, \mathbf{E}^{(2)}, \ldots, \mathbf{E}^{(M)}\}$, where M is the variable num and $\mathbf{E}^{(m)} \in \mathbb{R}^{1\times L}$ is the m-th exogenous series. Our goal is to learn a forecasting model $f(\cdot)$, which predicts the future T time steps of endogenous series $\widehat{\mathbf{X}} \in \mathbb{R}^{1\times T}$, based on its historical observation $\mathbf{X}$ and the exogenous variables $\mathbf{E}$.

4 Methodology

The detailed framework of ExoLLM is illustrated in Figure 2. Firstly, the Meta-task Instruction (MTI) and Multi-grained Prompt (MGP) text are embedded using frozen large language model to get uniform size embedding. Then, exogenous and endogenous series will be tokenized by shared Temporal-property preserved Tokenizer (TPT) to preserve temporal properties. Furthermore, a mainly frozen pre-trained LLM is utilized to integrate exogenous knowledge into endogenous token. It's worth noting that a Dual TS-Text Attention (DT^2Attention) is devised to align TS-Text feature space before and after LLM encoding, which enables the model to aware of specific modality. The output endogenous token wil be finally mapped to the future time series by a lightweight forecasting head.

4.1 Language-driven Exogenous Knowledge Utilization

Meta-task Instruction. To activate the knowledge transfer of LLM from nature language processing (NLP) to FEV, it is necessary to construct task instructions as guidance. As illustrated in Figure 3,

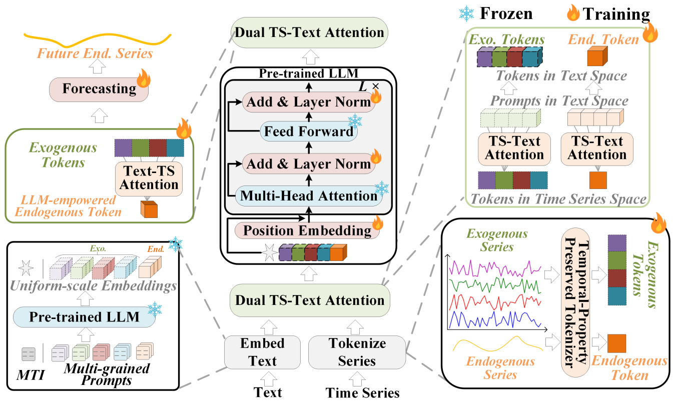


Figure 2: Overall architecture of ExoLLM, which consists of Dual TS-Text Attention and pre-trained LLM to sufficiently exploit exogenous variables in FEV.

the meta-task instruction comprises three key elements: (1) Overall description and analysis of dataset, offering explicit domain identification information to the model. (2) Brief summary of endogenous and exogenous variables, facilitating model to discern the source of each variables. (3) Introduction to the FEV task, fully activating LLM to accomplishing forecasting task with exogenous variables. We aim to activate the LLM's FEV capability in different domains through carefully designed meta-task instructions.

[Meta-task Instruction]
This dataset is ETTh1, containing the data collected from electricity transformers. Exogenous variables are high useful load, high useless load, middle useful load, middle useless load, low useful load, low useless load. It is necessary to utilize these exogenous variables sequentially to predict the endogenous variable oil temperature.

[Multi-grained Prompt]
① Natural attribute ② Trend ③ Periodicity ④ Stability ⑤ Noise

Figure 3: Example of Meta-task Instruction and headlines of Multi-grained Prompt in ETTh1.

Multi-grained Prompt. To comprehensively understand the external environment of Entire systems [60], we need to consider not only the apparent data correlation between numerical exogenous and endogenous variables, but also the natural properties, constant relationships, sequential trends, period influences, stability, and other multilevel factors. Therefore, we design multi-grained prompts (MGP) to exploit the LLM's comprehensive knowledge of the world to a diversified understanding of a specific environment. As shown in Table 2, the multi-grained prompt mainly consists of two elements: (1) Revealing the natural attribute of exogenous variables and their essential correlation with endogenous variables, endowing the model with prior knowledge of external environment. (2) Describing the dynamic characteristics of exogenous/endogenous series in term of trends, period, stability, and noise intensity, enabling the model to consider dynamic external influences. Intuitively, MGP not only deepens the LLM's understanding of exogenous variables, but also enhances the LLM's perception of the external invisible environment.

Uniform-scale Text Encoding. After constructing the meta-task instruction and multi-grained prompt, the next step involves encoding the text to obtain embeddings of uniform dimensions. To integrate these text with adequate language knowledge, we use a pre-trained LLM to encode these text descriptions. Since the text length of each prompt is different, we design an ingenious method to obtain the same embedding size. Particularly, we add a special token *<EOS>* at the end of the prompt. Since all the previous tokens are visible to *<EOS>* throughout the causal attention in LLM, the embedding of *<EOS>* could represent the entire text. The text encoding process is given by:

$$\mathbf{PT} = \mathrm{SelectLast}(\mathrm{LLM}(\mathbf{TD}; \langle EOS \rangle)), \tag{1}$$

where $\mathrm{SelectLast}(\cdot)$ denotes selecting the embedding of the last *<EOS>* token, $\mathrm{LLM}(\cdot)$ represents encoding part of large language model, $\mathbf{TD} = \{td_{\text{task}}, td^{(1)}_{\text{exo}}, td^{(2)}_{\text{exo}}, ..., td^{(M)}_{\text{exo}}, td_{\text{end}}\}$ is text description set of Meta-task Instruction and Multi-grained Prompt. $\mathbf{PT} = \{pt_{\text{task}}, pt^{(1)}_{\text{exo}}, pt^{(2)}_{\text{exo}}, ..., pt^{(M)}_{\text{exo}}, pt_{\text{end}}\}$ represents the uniform-scale text embeddings of $\mathbf{TD}$, where $pt_{\text{task}} \in \mathbb{R}^{1\times D}$ is the embedding of meta-task instruction, $pt^{(i)}_{\text{exo}} \in \mathbb{R}^{k\times D}$ is the i-th exogenous prompt embedding set, $pt_{\text{end}} \in \mathbb{R}^{k\times D}$ is endogenous prompt embedding set, k is multi-grained prompt number of one-type variable and D is the uniform hidden dimension.

4.2 Temporal-property Preserved Tokenizer

To facilitate LLM's understanding of the different types of variable series, we need to compress each series into a single token. Recent studies [30] use a linear layer to embed the entire time series as a token. However, this embedding approach neglects the temporal properties of data, resulting in the model's incomplete understanding of the relationships between exogenous and endogenous series. Therefore, we devise a Temporal-property Preserved Tokenizer (TPT) to obtain tokens reserving the temporal characteristics. Firstly, we partition the exogenous variables $\mathbf{E}$ and endogenous variables $\mathbf{X}$ into non-overlapping patches to enhance the local semantics at each time step [35], resulting in $\mathbf{P}_{end} \in \mathbb{R}^{1\times N\times P}$ and $\mathbf{P}_{exo} \in \mathbb{R}^{M\times N\times P}$, where P is patch length, and $N = \frac{L}{P}$ is the corresponding numbers of patches. To compress the temporal representations, TPT employs Self-Attention to learn temporal interactions among patches and selects the the last patch as the output:

$$\mathbf{TK}^{time}_{*} = \mathrm{SelectLast}(\mathrm{Self\text{-}Attn}(\mathrm{PE} + \mathbf{P}_{*})), \tag{2}$$

where $\mathrm{Self\text{-}Attn}(\cdot)$ denotes self-attention applied to time series, PE represents the position embedding, $\mathrm{SelectLast}(\cdot)$ denotes the operation of selecting the last patch, $\mathbf{P}_{*}$ is patched exogenous or endogenous series and $\mathbf{TK}^{time}_{*}$ is the corresponding token. Selecting the last patch as the token representation of the entire series is justified by two reasons: (1) It interacts with all preceding patches through attention, thus possessing sequence-level temporal information; (2) It is closest to the future sequence, providing crucial near-term information. Finally, we obtain exogenous tokens $\mathbf{TK}^{time}_{exo} \in \mathbb{R}^{M\times D}$ and endogenous token $\mathbf{TK}^{time}_{end} \in \mathbb{R}^{1\times D}$ in the time series feature space.

Table 2: An example of Multi-grained Prompt of one variable in ETTh1. Orange is chosen from exogenous and endogenous. Green is the variable name. Blue is prior knowledge about the variable's nature attribute. Black is the fixed template.

Characteristics	Prompts
Nature Attribute	① This Exogenous variable is High UseLess Load, representing external load that is inefficiently utilized. ② Exogenous High UseLess Load indicates a potential inefficiency in the system's external load handling. ③ Exogenous High UseLess Load can lead to increased energy consumption without corresponding output. ④ Exogenous High UseLess Load might suggest that the system is operating under suboptimal external conditions.
Trend	⑤ Exogenous High UseLess Load series shows an overall upward trend . ⑥ Exogenous High UseLess Load series initially rises and then declines . ⑦ Exogenous High UseLess Load series exhibits an overall declining trend . ⑧ Exogenous High UseLess Load series initially declines and then rises .
Period	⑨ Exogenous High UseLess Load series has no apparent periodicity . ⑩ Exogenous High UseLess Load series exhibits shorter periodicity and higher frequency . ⑪ Exogenous High UseLess Load series displays clear periodicity . ⑫ Exogenous High UseLess Load series exhibits relatively longer periodicity .
Stability	⑬ Exogenous High UseLess Load series undergoes significant instability over all the time. ⑭ Exogenous High UseLess Load series remains relatively stable with minimal fluctuations . ⑮ Exogenous High UseLess Load series experiences occasional bouts of volatility, interspersed with periods of relative calm. ⑯ Exogenous High UseLess Load series shows consistent stability, with values remaining close to a steady mean.
Noise Intensity	⑰ Exogenous High UseLess Load series is subject to very strong noise interference . ⑱ Exogenous High UseLess Load series has a low signal-to-noise ratio, where noise significantly affects the clarity of the underlying data. ⑲ Exogenous High UseLess Load series experiences moderate noise, partially obscuring the underlying pattern. ⑳ Exogenous High UseLess Load series is not influenced by any noise interference .

4.3 Knowledge-retained LLM Encoder

Understanding the exogenous impact on endogenous variables is crucial for time series forecasting. We utilize meta-task instruction along with tokenized exogenous and endogenous variables to LLMs to fully exploit the prior knowledge in LLMs, thereby forming enhanced representations of endogenous token:

$$\mathbf{TK}_{end}^{llm} = \mathrm{LLM}(\{pt_{task}, \mathbf{TK}_{exo}, \mathbf{TK}_{end}\}), \tag{3}$$

where $\mathrm{LLM}(\cdot)$ denotes the encoder part of LLM. Each variable is treated as a token, and exogenous and endogenous variables are concatenated in a fixed order to form a "sentence" in a fixed order, like $[pt_{task}, \mathbf{TK}_{exo}^{(1)}, \mathbf{TK}_{exo}^{(2)}, ..., \mathbf{TK}_{exo}^{(M)}, \mathbf{TK}_{end}]$. Following [67], we freeze the positional embedding layers and self-attention blocks in LLM to retain majority of learned knowledge from language pre-training. Ultimately, we obtain an exogenous variable enhanced representation of endogenous token, $\mathbf{TK}_{end}^{llm} \in \mathbb{R}^{1\times D}$, which encapsulates rich information from prior exogenous knowledge.

4.4 Feature Alignment with Dual TS-Text Attention

Given that LLM is pre-trained on discrete textual data and lack exposure to continuous numerical values, directly inputting tokens $\mathbf{TK}_{exo}^{time}$ and $\mathbf{TK}_{end}^{time}$ in time series featrue space into LLMs would increase the difficulties in understanding never-seen modality, thus resulting in degraded predictive performance. Besides, the output token from LLM in text space is difficult to decode into future series. Thus, a Dual TS-Text Attention (DT^{2}Attention) is devised to align ts-text feature space before and after LLM encoder, respectively.

TS-Text Attention. Intuitively, there should be a certain distinction between exogenous and endogenous tokens to avoid over-smoothing representation among different types of tokens, and absorb certain prior external knowledge to enhance the LLM's encoding ability. Thus, a TS-Text Attention is designed to achieve: 1) Mapping tokens from time series feature space to text feature space; 2) Distinguishing between endogenous and exogenous Tokens. Specifically, for any type of token, TS-Text Attention designs its Query as token in time series space, while the Key and Value are its corresponding multi-grained prompt. Then, we perform Cross Attention to align tokens:

$$\mathbf{TK}_{*}^{text} = \mathrm{Cross\text{-}Attn}(\mathbf{TK}_{*}^{time}, \mathbf{PT}_{*}, \mathbf{PT}_{*}), \tag{4}$$

where $\mathbf{TK}_{*}^{time} \in \mathbb{R}^{D}$ is the exogenous/endogenous token in time series space, $\mathbf{PT}_{*} \in \mathbb{R}^{k\times D}$ is this variable's corresponding multi-grained prompt, $\mathbf{TK}_{*}^{text} \in \mathbb{R}^{D}$ is the mapped token in text space and will be input into LLM in Eq (3).

Text-TS Attention. Denote $\mathbf{TK}_{end}^{llm} \in \mathbb{R}^{1\times D}$ as the endogenous token encoded by LLM encoder in Eq (3). Since $\mathbf{TK}_{end}^{llm}$ remains in the text space, directly decode $\mathbf{TK}_{end}^{llm}$ for forecasting faces the challenge of converting textual semantics into time series. Thus, we use Text-TS Attention to alleviate such problem, decoding $\mathbf{TK}_{end}^{llm}$ into time series space based on the temporal information of exogenous series. This can be expressed as:

$$\mathbf{TK}_{dec}^{llm} = \mathrm{Cross\text{-}Attn}(\mathbf{TK}_{end}^{llm}, \mathbf{TK}_{exo}^{time}, \mathbf{TK}_{exo}^{time}), \tag{5}$$

where $\mathbf{TK}_{exo}^{time}$ represents exogenous variables in time series space, $\mathbf{TK}_{dec}^{llm} \in \mathbb{R}^{1\times D}$ is the decoded endogenous token. Through this approach, we can better utilize the representation capability of LLMs and combine exogenous series to enhance the endogenous forecasting.

4.5 Lightweight Forecasting Head

Considering the richness of the encoded token and maximumly preserving exogenous information by LLMs, a simple linear layer is employed to transform $\mathbf{TK}_{dec}^{llm}$ for forecasting:

$$\widehat{\mathbf{X}} = \mathrm{Linear}(\mathbf{TK}_{dec}^{llm}). \tag{6}$$

Table 3: Full results of the long-term FEV. The input sequence length is set to 96 for all baselines. Results are averaged from all prediction lengths.

Models	ExoLLM (Ours)		TimeXer [2024]		ITrans. [2024]		PatchTST [2023]		Cross. [2023]		TiDE [2023]		SCINet [2022]		Auto. 2021		GPT4TS [2023]		TimeLLM [2024]		LLM4TS [2023]	
Dataset	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ECL	**0.330**	**0.404**	0.336	0.415	0.365	0.486	0.394	0.446	0.344	0.412	0.419	0.468	0.428	0.450	0.495	0.528	0.392	0.442	0.365	0.413	0.378	0.427
Weather	**0.001**	**0.027**	0.002	0.031	0.002	0.029	0.002	0.031	0.005	0.055	0.002	0.029	0.007	0.030	0.007	0.061	0.005	0.056	0.003	0.036	0.004	0.046
ETTh1	**0.069**	**0.205**	0.074	0.211	0.075	0.224	0.078	0.216	0.285	0.447	0.084	0.223	0.437	0.256	0.130	0.282	0.126	0.305	0.104	0.277	0.115	0.304
ETTh2	**0.175**	**0.327**	0.183	0.337	0.200	0.357	0.192	0.345	1.027	0.873	0.205	0.356	1.154	0.406	0.243	0.386	0.277	0.443	0.226	0.388	0.251	0.415
ETTm1	**0.049**	**0.165**	0.051	0.168	0.053	0.173	0.054	0.173	0.412	0.548	0.053	0.173	0.099	0.204	0.086	0.231	0.106	0.264	0.080	0.233	0.093	0.248
ETTm2	**0.113**	**0.249**	0.116	0.252	0.127	0.261	0.120	0.258	0.976	0.769	0.122	0.261	0.685	0.334	0.154	0.304	0.196	0.349	0.162	0.311	0.179	0.330
Traffic	**0.145**	**0.220**	0.150	0.227	0.161	0.412	0.173	0.253	0.182	0.268	0.319	0.408	0.447	0.362	0.303	0.353	0.166	0.247	0.186	0.271	0.177	0.260

Table 4: Full results of the short-term FEV. The input length and prediction length are set to 168 and 24 respectively for all baselines. Avg means the average results from all five datasets.

Models	ExoLLM (Ours)		TimeXer [2024]		ITrans. [2024]		PatchTST [2023]		Cross. [2023]		TiDE [2023]		SCINet [2022]		Auto. 2021		GPT4TS [2023]		TimeLLM [2024]		LLM4TS [2023]	
Dataset	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	**0.216**	**0.234**	0.238	0.268	0.265	0.300	0.267	0.284	0.245	0.289	0.335	0.340	0.373	0.368	0.402	0.398	0.275	0.303	0.255	0.293	0.265	0.315
PJM	**0.076**	**0.175**	0.088	0.188	0.097	0.197	0.106	0.209	0.149	0.198	0.124	0.228	0.143	0.259	0.168	0.267	0.118	0.207	0.210	0.283	0.255	0.308
BE	**0.358**	**0.225**	0.374	0.241	0.394	0.270	0.403	0.264	0.436	0.294	0.523	0.336	0.731	0.412	0.500	0.333	0.502	0.288	0.384	0.230	0.426	0.258
FR	**0.365**	**0.203**	0.381	0.211	0.439	0.233	0.411	0.220	0.440	0.216	0.510	0.290	0.855	0.384	0.519	0.295	0.570	0.497	0.501	0.443	0.519	0.459
DE	**0.422**	**0.401**	0.440	0.418	0.479	0.443	0.461	0.432	0.540	0.423	0.568	0.496	0.565	0.497	0.674	0.544	0.569	0.490	0.498	0.438	0.517	0.460
AVG	**0.288**	**0.251**	0.304	0.265	0.335	0.289	0.330	0.282	0.362	0.284	0.412	0.338	0.533	0.384	0.453	0.368	0.325	0.326	0.338	0.378	0.399	0.408

where $\widehat{\mathbf{X}} \in \mathbb{R}^{1\times T}$ is the future endogenous series.

5 Experiments

5.1 Dataset and Experimental Settings

Datasets and Experimental Setups. To completely evaluate the FEV capability of ExoLLM, we conduct experiments on 12 real-world datasets. These datasets are collected from web and especially the exogenous factors retrieved from are in the formation of language. In particular, seven well-established public long-term datasets from different domains, and five short-term datasets in electricity price are involved in our FEV experiments. For short-term forecasting datasets, the input length is set as 168 and prediction length is 24. For long-term forecasting datasets, the input length is set as 96 and prediction length varies $\{96, 192, 336, 720\}$.

Baselines. We compare ExoLLM with 10 baselines, which comprise the state-of-the-art forecasting methods, including LLM-based model: LLM4TS [2], GPT4TS [67], TimeLLM [19], Transformer-based model: TimeXer [53] PatchTST [35], ITransformer [30], Crossformer [65], Autoformer [57], CNN-based model: SCINet [28], and Linear-based model: TiDE [6] . Among these models, TimeXer and TiDE are advanced recent forecaster elaborated for exogenous variables.

5.2 Main Results

Long-term FEV. Long-term forecasting results are presented in Table 3, where ExoLLM demonstrates superior performance across different prediction length again all baselines. In contrast to the cutting-edge LLM-based model TimeLLM, ExoLLM achieves performance gains of 31.2% and 19.8% in MSE and MAE metrics. Compared to the state-of-the-art (SOTA) FEV model TimeXer, ExoLLM exhibits a relative reduction of 9.1% and 4.1% in MSE and MAE metrics, respectively. These results highlight ExoLLM's exceptional FEV capability in long-term scenario.

Short-term FEV. In Table 4, ExoLLM consistently maintains a leading predictive performance. Compared to SOTA short-term forecasting model SCINet, ExoLLM achieves significant reductions in 35.5% MAE and 46.1% MSE respectively. Besides, ExoLLM outperforms the FEV-designed model TimeXer in all short-term datasets.The comprehensive experimental results underscore the FEV efficacy of ExoLLM in short-term forecasting.

Table 5: Results of few-shot FEV. Results are averaged from all prediction lengths.

Models	ExoLLM		TimeXer		ITrans.		PatchTST		Cross.		TiDE		SCINet		Auto.		GPT4TS		TimeLLM		LLM4TS	
Dataset	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.084**	**0.230**	0.094	0.248	0.091	0.251	0.153	0.344	0.346	0.506	0.126	0.312	0.533	0.288	0.159	0.316	0.095	0.242	0.101	0.251	0.140	0.342
ETTh2	**0.253**	**0.403**	0.279	0.435	0.290	0.439	0.401	0.546	1.501	1.080	0.327	0.478	1.681	0.500	0.352	0.475	0.278	0.425	0.298	0.439	0.364	0.512
ETTm1	**0.057**	**0.181**	0.062	0.194	0.062	0.190	0.124	0.290	0.475	0.601	0.094	0.256	0.115	0.224	0.102	0.255	0.062	0.190	0.062	0.190	0.109	0.273
ETTm2	**0.144**	**0.291**	0.156	0.310	0.163	0.306	0.253	0.410	1.187	0.882	0.209	0.365	0.869	0.389	0.204	0.360	0.155	0.301	0.158	0.306	0.231	0.388

Table 6: Results of zero-shot FEV. Results are averaged from all prediction lengths.

Models		ExoLLM		TimeXer		ITrans.		PatchTST		Cross.		TiDE		SCINet		Auto.		GPT4TS		TimeLLM		LLM4TS	
Source	Target	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	ETTh2	**0.204**	**0.359**	0.228	0.390	0.221	0.395	0.380	0.544	0.875	0.796	0.308	0.490	1.309	0.453	0.384	0.497	0.232	0.381	0.248	0.394	0.344	0.538
ETTh2	ETTh1	**0.074**	**0.212**	0.082	0.228	0.085	0.230	0.118	0.287	0.429	0.562	0.096	0.251	0.489	0.262	0.103	0.250	0.082	0.223	0.087	0.230	0.107	0.269
ETTm1	ETTm2	**0.162**	**0.309**	0.177	0.332	0.178	0.324	0.353	0.495	1.348	1.025	0.267	0.437	0.328	0.382	0.299	0.438	0.178	0.324	0.176	0.324	0.310	0.466
ETTm2	ETTm1	**0.054**	**0.176**	0.058	0.187	0.061	0.185	0.094	0.248	0.455	0.538	0.078	0.220	0.326	0.236	0.075	0.217	0.058	0.182	0.059	0.185	0.086	0.234

Few-shot FEV. In few-shot learning, only 10% of the training data are utilized, and the outcomes are presented in Table 5. Quantitatively, ExoLLM achieves an average 8.9% reduction in MSE and 4.5% reduction in MAE compared to the top-performing GPT4TS.

Zero-shot FEV. This task is to evaluate how effectively a model can perform on *target dataset* when it has been trained on *source dataset*, and the results are presented in Table 6. ExoLLM outperforms all SOTA models, achieving a performance improvement of over 5% compared to other models in zero-shot FEV. This demonstrates ExoLLM's powerful FEV generalization capabilities with pre-trained knowledge.

5.3 Ablation Study

We conduct comprehensive ablation studies by systematically removing each key module from the ExoLLM framework and evaluating its performance across six distinct datasets. This allows us to assess the individual contributions of each component to the overall effectiveness of the model. **w/o MGP** removes Multi-grained Prompt (MGP). **w/o MTI** removes Meta-task Instruction (MTI). **w/o DT^{2}A** removes Dual TS-text Attention (DT^{2}Attention) for feature space alignment. **w/o TPT** replaces Temporal-property Preserved Tokenizer (TPT), which could preserve temporal properties for each token, with a linear layerr. We analyze the results shown in Table 7. The obervations are listed as follows: Obs.1) Removing MGP and MTI results in the most significant decrease in prediction metrics, emphasizing their strong ability in activating LLM in FEV. Obs.2) DT^{2}Attention also significantly improves the model performance, demonstrating the importance of featrure space alignment. Obs.3) TST constantly promotes the forecasting accuracy, suggesting that reserving temporal properties in each token is needed.

Table 7: Ablation of each module on ECL, Weather, ETTh1, Traffic, PJM and NP.

Dataset	ECL		Weather		ETTh1		Traffic		PJM		NP	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ExoLLM	**0.330**	**0.404**	**0.001**	**0.027**	**0.069**	**0.205**	**0.145**	**0.220**	**0.076**	**0.175**	**0.216**	**0.234**
w/o MGP	0.359	0.413	0.003	0.037	0.083	0.239	0.152	0.227	0.110	0.185	0.238	0.281
w/o MTI	0.354	0.418	0.003	0.035	0.096	0.209	0.153	0.222	0.099	0.187	0.238	0.277
w/o DT^{2}A	0.348	0.418	0.002	0.034	0.074	0.212	0.157	0.226	0.090	0.176	0.235	0.262
w/o TPT	0.332	0.405	0.002	0.031	0.079	0.210	0.152	0.223	0.086	0.182	0.225	0.241

5.4 Exogenous Scale Analysis

Real-world time series often encounter challenges such as the absence of crucial exogenous data. In this section, we employ random masking to simulate these scenarios and further investigate the forecasting performance. As illustrated in Figure 4 (a) and (b), We vary **prompt number** from 0 to 16 and report the MSE and MAE results on ETTh1 and ETTh2. For instance, when the number of prompts in Figure 4 is 16, we respectively remove prompt of Natural Attribute, Trend, Period, Stability, and Noise Intensity, and report the average results. We observe that the performance improvement is positive to the prompt size, indicating that more prompts extracting more auxiliary information from LLM. As shown in Figure 4 (c) and (d), we vary **exogenous variable number** in {0%, 25%, 50%, 75%, 100%} and find that more exogenous variables improve the model performance, indicating that ExoLLM is able to sufficiently understand complex and evolving environment. To further identify which prompts and exogenous variables are most important, we individually remove each exogenous prompt and variable, and results are in Table 9 and Table 8. For ETTh1, the

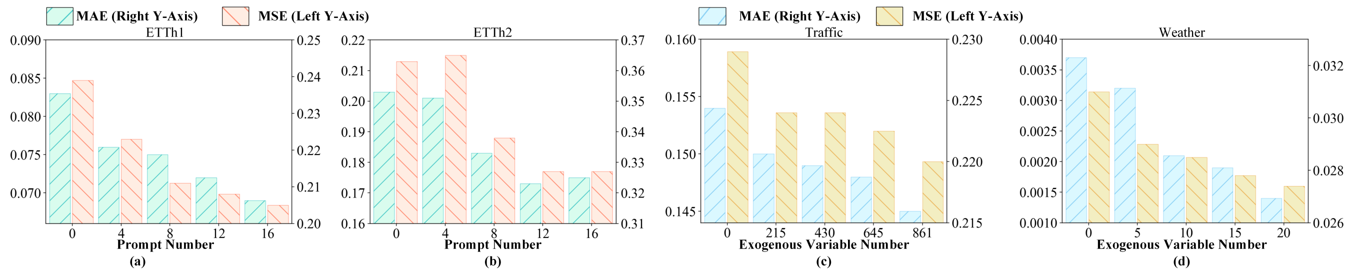


Figure 4: The MAE (left Y-axis) and MSE results (right Y-axis) of ExoLLM on ETTh, Traffic and Weather. (a) and (b) display the performance on different prompt number. (c) and (d) demonstrates the performance on different number of exogenous variables.

critical exogenous prompt is Trend, and the most important exogenous variable is HUFT. This result aligns with intuition, as trend information often plays a pivotal role in long-term forecasting, and HUFT captures crucial temporal dynamics. It further demonstrates ExoLLM's ability to effectively leverage exogenous knowledge, enhancing prediction accuracy by identifying and utilizing the most relevant external factors.

Table 8: Results of removing different type of MGP.

	w/o Attribute	w/o Trend	w/o Period	w/o Stability	w/o Noise
MSE	0.071	0.071	0.076	0.070	0.070
MAE	0.206	0.206	0.210	0.207	0.206

Table 9: Results of removing different exogenous variables.

	w/o HUFT	w/o HULL	w/o MUFL	w/o MULL	w/o LUFL	w/o LULL
MSE	0.074	0.070	0.072	0.071	0.072	0.072
MAE	0.212	0.207	0.206	0.206	0.206	0.207

5.5 Case Study

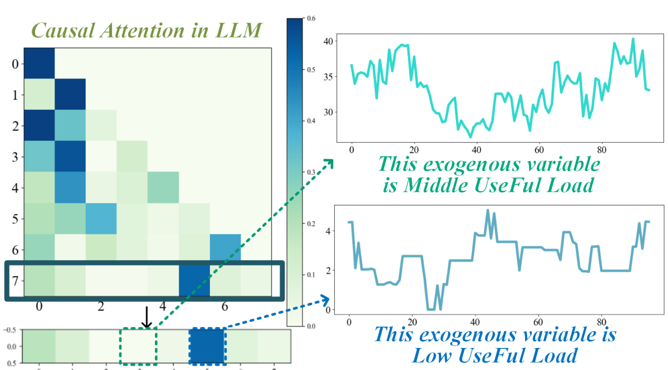


Figure 5: Case study of causal attention on different tokens.

Figure 5 illustrates the attention map of ETTh2 (comprising 6 exogenous variables and 1 endogenous variable) during causal LLM encoding. Here, tokens 0 through 7 represent the inputs to the LLM: token 0 denotes the MTI, tokens 1 through 6 represent the token sequence for the exogenous variables, and token 7 represents the endogenous variable. These tokens are arranged in a fixed sequence, forming an input structure akin to a sentence. Encoding through a LLM allows the endogenous variable token to be enriched with open-world knowledge conveyed through language, thereby enhancing prediction accuracy. The case study on the ETTh2 dataset demonstrates that: (1) Meta-task Instruction receive extensive attention for each variable, demonstrating that the guidance we design perfectly activates the LLM's ability to transition from NLP to FEV. (2) ExoLLM is able to distinguish between exogenous variables that exhibit strong association with the endogenous variable, resulting in a more focused and interpretable attention map.

6 Conclusion

To align with the evolving needs of web-related technologies, which require handling external influences from dynamic and shifting environments, we propose an LLM-based approach, ExoLLM, for time series forecasting with exogenous variables (FEV). By incorporating Meta-task Instruction, Multi-grained Prompt, and Dual TS-Text Attention, ExoLLM enables large language models (LLMs) to excel in multiple forecasting scenarios, including long-term, short-term, few-shot, and zero-shot tasks. This framework introduces a novel paradigm that taps into the textualized knowledge embedded in LLMs to enhance the understanding of structural time-series data. It allows ExoLLM to interpret temporal dynamics in a way that is enriched by the contextual, text-like representation of the data, thereby improving its forecasting accuracy and robustness. The versatility of ExoLLM also opens new possibilities for structural and tabular data learning across various domains, driving innovations in real-world applications central to the web ecosystem.

Acknowledgment

This paper is partially supported by the National Natural Science Foundation of China (No.12227901, No.62072427), Natural Science Foundation of Jiangsu Province (BK20240460), the grant from State Key Laboratory of Resources and Environmental Information System.

References

[1] Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. Tempo: Prompt-based generative pre-trained transformer for time series forecasting. *arXiv preprint arXiv:2310.04948*, 2023.

[2] Ching Chang, Wen-Chih Peng, and Tien-Fu Chen. Llm4ts: Two-stage fine-tuning for time-series forecasting with pre-trained llms. *arXiv preprint arXiv:2308.08469*, 2023.

[3] Feiyi Chen, Zhen Qin, Mengchu Zhou, Yingying Zhang, Shuiguang Deng, Lunting Fan, Guansong Pang, and Qingsong Wen. Lara: A light and anti-overfitting retraining approach for unsupervised time series anomaly detection. In *Proceedings of the ACM on Web Conference 2024*, pages 4138–4149, 2024.

[4] Sheng Chen and Steve A Billings. Representations of non-linear systems: the narmax model. *International journal of control*, 49(3):1013–1032, 1989.

[5] Ernesto Colacrai, Federico Cinus, Gianmarco De Francisci Morales, and Michele Starnini. Navigating multidimensional ideologies with reddit's political compass: Economic conflict and social affinity. In *Proceedings of the ACM on Web Conference 2024*, pages 2582–2593, 2024.

[6] Abhimanyu Das, Weihao Kong, Andrew Leach, Shaan K Mathur, Rajat Sen, and Rose Yu. Long-term forecasting with tide: Time-series dense encoder. *Transactions on Machine Learning Research*, 2023.

[7] Junfeng Fang, Shuai Zhang, Chang Wu, Zhengyi Yang, Zhiyuan Liu, Sihang Li, Kun Wang, Wenjie Du, and Xiang Wang. Moltc: Towards molecular relational modeling in language models. *arXiv preprint arXiv:2402.03781*, 2024.

[8] Jie Feng, Yuwei Du, Tianhui Liu, Siqi Guo, Yuming Lin, and Yong Li. Citygpt: Empowering urban spatial cognition of large language models. *arXiv preprint arXiv:2406.13948*, 2024.

[9] Chen Gao, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanities and Social Sciences Communications*, 11(1):1–24, 2024.

[10] Nate Gruver, Marc Finzi, Shikai Qiu, and Andrew Gordon Wilson. Large language models are zero-shot time series forecasters. *arXiv preprint arXiv:2310.07820*, 2023.

[11] Yunda Guo, Jiake Ge, Panfeng Guo, Yunpeng Chai, Tao Li, Mengnan Shi, Yang Tu, and Jian Ouyang. Pass: Predictive auto-scaling system for large-scale enterprise web applications. In *Proceedings of the ACM on Web Conference 2024*, pages 2747–2758, 2024.

[12] Brian D Haig. What is a spurious correlation? *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, 2(2):125–132, 2003.

[13] Qihe Huang, Lei Shen, Ruixin Zhang, Jiahuan Cheng, Shouhong Ding, Zhengyang Zhou, and Yang Wang. Hdmixer: Hierarchical dependency with extendable patch for multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 12608–12616, 2024.

[14] Qihe Huang, Lei Shen, Ruixin Zhang, Shouhong Ding, Binwu Wang, Zhengyang Zhou, and Yang Wang. Crossgnn: Confronting noisy multivariate time series via cross interaction refinement. *Advances in Neural Information Processing Systems*, 36, 2024.

[15] Qihe Huang, Zhengyang Zhou, Kuo Yang, Gengyu Lin, Zhongchao Yi, and Yang Wang. Leret: Language-empowered retentive network for time series forecasting. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, 2024.

[16] Sheo Yon Jhin, Jaehoon Lee, Minju Jo, Seungji Kook, Jinsung Jeon, Jihyeon Hyeong, Jayoung Kim, and Noseong Park. Exit: Extrapolation and interpolation-based neural controlled differential equations for time-series classification and forecasting. In *Proceedings of the ACM Web Conference 2022*, pages 3102–3112, 2022.

[17] Renhe Jiang, Zhaonan Wang, Yudong Tao, Chuang Yang, Xuan Song, Ryosuke Shibasaki, Shu-Ching Chen, and Mei-Ling Shyu. Learning social meta-knowledge for nowcasting human mobility in disaster. In *Proceedings of the ACM Web Conference 2023*, pages 2655–2665, 2023.

[18] Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, et al. Time-llm: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728*, 2023.

[19] Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. Time-LLM: Time series forecasting by reprogramming large language models. In *International Conference on Learning Representations (ICLR)*, 2024.

[20] Harshavardhan Kamarthi, Lingkai Kong, Alexander Rodríguez, Chao Zhang, and B Aditya Prakash. Camul: Calibrated and accurate multi-view time-series forecasting. In *Proceedings of the ACM Web Conference 2022*, pages 3174–3185, 2022.

[21] Zhichen Lai, Huan Li, Dalin Zhang, Yan Zhao, Weizhu Qian, and Christian S Jensen. E2usd: Efficient-yet-effective unsupervised state detection for multivariate time series. In *Proceedings of the ACM on Web Conference 2024*, pages 3010–3021, 2024.

[22] Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021.

[23] Gengyu Lin, Zhengyang Zhou, Qihe Huang, Kuo Yang, Shifen Cheng, and Yang Wang. Fairstg: Countering performance heterogeneity via collaborative sample-level optimization. *arXiv preprint arXiv:2403.12391*, 2024.

[24] Shengsheng Lin, Weiwei Lin, Wentai Wu, Feiyu Zhao, Ruichao Mo, and Haotong Zhang. Segrnn: Segment recurrent neural network for long-term time series forecasting. *arXiv preprint arXiv:2308.11200*, 2023.

[25] Shengsheng Lin, Weiwei Lin, Wentai Wu, Haojun Chen, and Junjie Yang. Sparsetsf: Modeling long-term time series forecasting with* 1k* parameters. In *Forty-first International Conference on Machine Learning*, 2024.

[26] Shengsheng Lin, Weiwei Lin, Wentai Wu, Songbo Wang, and Yongxiang Wang. Petformer: Long-term time series forecasting via placeholder-enhanced transformer. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024.

[27] Shengsheng Lin, Weiwei Lin, HU Xinyi, Wentai Wu, Ruichao Mo, and Haocheng Zhong. Cyclenet: Enhancing time series forecasting through modeling periodic patterns. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[28] Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia Lai, Lingna Ma, and Qiang Xu. Scinet: Time series modeling and forecasting with sample convolution and interaction. *Advances in Neural Information Processing Systems*, 35:5816–5828, 2022.

[29] Xu Liu, Junfeng Hu, Yuan Li, Shizhe Diao, Yuxuan Liang, Bryan Hooi, and Roger Zimmermann. Unitime: A language-empowered unified model for cross-domain time series forecasting. In *Proceedings of the ACM on Web Conference 2024*, pages 4095–4106, 2024.

[30] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024.

[31] Yong Liu, Guo Qin, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Autotimes: Autoregressive time series forecasters via large language models. *arXiv preprint arXiv:2402.02370*, 2024.

[32] Qingqing Long, Zheng Fang, Chen Fang, Chong Chen, Pengfei Wang, and Yuanchun Zhou. Unveiling delay effects in traffic forecasting: A perspective from spatial-temporal delay differential equations. In *Proceedings of the ACM on Web Conference 2024*, pages 1035–1044, 2024.

[33] Jiecheng Lu, Xu Han, Yan Sun, and Shihao Yang. Cats: Enhancing multivariate time series forecasting by constructing auxiliary time series as exogenous variables. *arXiv preprint arXiv:2403.01673*, 2024.

[34] Kentaro Miyake, Hiroyoshi Ito, Christos Faloutsos, Hirotomo Matsumoto, and Atsuyuki Morishima. Netevolve: Social network forecasting using multi-agent reinforcement learning with interpretable features. In *Proceedings of the ACM on Web Conference 2024*, pages 2542–2551, 2024.

[35] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

[36] Wentao Ning, Reynold Cheng, Xiao Yan, Ben Kao, Nan Huo, Nur Al Hasan Haldar, and Bo Tang. Debiasing recommendation with personal popularity. In *Proceedings of the ACM on Web Conference 2024*, pages 3400–3409, 2024.

[37] Kohei Obata, Koki Kawabata, Yasuko Matsubara, and Yasushi Sakurai. Dynamic multi-network mining of tensor time series. In *Proceedings of the ACM on Web Conference 2024*, pages 4117–4127, 2024.

[38] Kin G Olivares, Cristian Challu, Grzegorz Marcjasz, Rafał Weron, and Artur Dubrawski. Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with nbeatsx. *International Journal of Forecasting*, 39(2): 884–900, 2023.

[39] Boris N Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*, 2019.

[40] Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S Jensen, Zhenli Sheng, et al. Tfb: Towards comprehensive and fair benchmarking of time series forecasting methods. *arXiv preprint arXiv:2403.20150*, 2024.

[41] Xiangfei Qiu, Xiuwen Li, Ruiyang Pang, Zhicheng Pan, Xingjian Wu, Liu Yang, Jilin Hu, Yang Shu, Xuesong Lu, Chengcheng Yang, et al. Easytime: Time series forecasting made easy. *arXiv preprint arXiv:2412.17603*, 2024.

[42] Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. Duet: Dual clustering enhanced multivariate time series forecasting. *arXiv preprint arXiv:2412.10859*, 2024.

[43] Chenxi Sun, Yaliang Li, Hongyan Li, and Shenda Hong. Test: Text prototype aligned embedding to activate llm's ability for time series. *arXiv preprint arXiv:2308.08241*, 2023.

[44] Stylianos I Vagropoulos, GI Chouliaras, Evaggelos G Kardakos, Christos K Simoglou, and Anastasios G Bakirtzis. Comparison of sarimax, sarima, modified sarima and ann-based models for short-term pv generation forecasting. In *2016 IEEE international energy conference (ENERGYCON)*, pages 1–6. IEEE, 2016.

[45] Binwu Wang, Yudong Zhang, Jiahao Shi, Pengkun Wang, Xu Wang, Lei Bai, and Yang Wang. Knowledge expansion and consolidation for continual traffic

prediction with expanding graphs. *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[46] Binwu Wang, Jiaming Ma, Pengkun Wang, Xu Wang, Yudong Zhang, Zhengyang Zhou, and Yang Wang. Stone: A spatio-temporal ood learning framework kills both spatial and temporal shifts. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2948–2959, 2024.

[47] Binwu Wang, Pengkun Wang, Yudong Zhang, Xu Wang, Zhengyang Zhou, Lei Bai, and Yang Wang. Towards dynamic spatial-temporal graph learning: A decoupled perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 9089–9097, 2024.

[48] Binwu Wang, Pengkun Wang, Zhengyang Zhou, Zhe Zhao, Wei Xu, and Yang Wang. Make bricks with a little straw: Large-scale spatio-temporal graph learning with restricted gpu-memory capacity. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 2388–2396, 2024.

[49] Chengsen Wang, Qi Qi, Jingyu Wang, Haifeng Sun, Zirui Zhuang, Jinming Wu, and Jianxin Liao. Rethinking the power of timestamps for robust time series forecasting: A global-local fusion perspective. *arXiv preprint arXiv:2409.18696*, 2024.

[50] Chengsen Wang, Qi Qi, Jingyu Wang, Haifeng Sun, Zirui Zhuang, Jinming Wu, Lei Zhang, and Jianxin Liao. Chattime: A unified multimodal time series foundation model bridging numerical and textual data. *arXiv preprint arXiv:2412.11376*, 2024.

[51] Chengsen Wang, Zirui Zhuang, Qi Qi, Jingyu Wang, Xingyu Wang, Haifeng Sun, and Jianxin Liao. Drift doesn't matter: dynamic decomposition with diffusion reconstruction for unstable multivariate time series anomaly detection. *Advances in Neural Information Processing Systems*, 36, 2024.

[52] Pengkun Wang, Zhe Zhao, HaiBin Wen, Fanfu Wang, Binwu Wang, Qingfu Zhang, and Yang Wang. Llm-autoda: Large language model-driven automatic data augmentation for long-tailed problems. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

[53] Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Yong Liu, Yunzhong Qiu, Haoran Zhang, Jianmin Wang, and Mingsheng Long. Timexer: Empowering transformers for time series forecasting with exogenous variables. *arXiv preprint arXiv:2402.19072*, 2024.

[54] Zexin Wang, Changhua Pei, Minghua Ma, Xin Wang, Zhihan Li, Dan Pei, Saravan Rajmohan, Dongmei Zhang, Qingwei Lin, Haiming Zhang, et al. Revisiting vae for unsupervised time series anomaly detection: A frequency perspective. In *Proceedings of the ACM on Web Conference 2024*, pages 3096–3105, 2024.

[55] Wei Wei, Chao Huang, Lianghao Xia, and Chuxu Zhang. Multi-modal self-supervised learning for recommendation. In *Proceedings of the ACM Web Conference 2023*, pages 790–800, 2023.

[56] Billy M Williams. Multivariate vehicular traffic flow prediction: evaluation of arimax modeling. *Transportation Research Record*, 1776(1):194–200, 2001.

[57] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, 34:22419–22430, 2021.

[58] Wei Xu, Pengkun Wang, Zhe Zhao, Binwu Wang, Xu Wang, and Yang Wang. When imbalance meets imbalance: Structure-driven learning for imbalanced graph classification. In *Proceedings of the ACM on Web Conference 2024*, pages 905–913, 2024.

[59] Wentao Xu, Weiqing Liu, Chang Xu, Jiang Bian, Jian Yin, and Tie-Yan Liu. Rest: Relational event-driven stock trend forecasting. In *Proceedings of the web conference 2021*, pages 1–10, 2021.

[60] Kuo Yang, Zhengyang Zhou, Qihe Huang, Limin Li, Yuxuan Liang, and Yang Wang. Improving generalization of dynamic graph learning via environment prompt. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

[61] Kuo Yang, Zhengyang Zhou, Wei Sun, Pengkun Wang, Xu Wang, and Yang Wang. Extract and refine: Finding a support subgraph set for graph representation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2953–2964, 2023.

[62] Zhongchao Yi, Zhengyang Zhou, Qihe Huang, Yanjiang Chen, Liheng Yu, Xu Wang, and Yang Wang. Get rid of isolation: A continuous multi-task spatio-temporal learning framework. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[63] Liheng Yu, Pengkun Wang, Zhe Zhao, Zhongchao Yi, Sun Nan, Di Wu, and Yang Wang. Xrdmamba: Large-scale crystal material space group identification with selective state space model. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 4233–4237, 2024.

[64] Yuan Yuan, Jingtao Ding, Jie Feng, Depeng Jin, and Yong Li. Unist: A prompt-empowered universal model for urban spatio-temporal prediction. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4095–4106, 2024.

[65] Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *International Conference on Learning Representations*, 2023.

[66] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.

[67] Tian Zhou, Peisong Niu, Xue Wang, Liang Sun, and Rong Jin. One fits all: Power general time series analysis by pretrained lm. 2023.

[68] Zhengyang Zhou, Yang Wang, Xike Xie, Lianliang Chen, and Hengchang Liu. Riskoracle: A minute-level citywide traffic accident forecasting framework. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 1258–1265, 2020.

[69] Zhengyang Zhou, Yang Wang, Xike Xie, Lei Qiao, and Yuantao Li. Stuanet: Understanding uncertainty in spatiotemporal collective human mobility. In *Proceedings of the Web Conference 2021*, pages 1868–1879, 2021.

[70] Zhengyang Zhou, Qihe Huang, Gengyu Lin, Kuo Yang, LEI BAI, and Yang Wang. GReto: Remedying dynamic graph topology-task discordance via target homophily. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=8duT3mi_5n.

[71] Zhengyang Zhou, Qihe Huang, Kuo Yang, Kun Wang, Xu Wang, Yudong Zhang, Yuxuan Liang, and Yang Wang. Maintaining the status quo: Capturing invariant relations for ood spatiotemporal learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, page 3603–3614, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701030. doi: 10.1145/3580305.3599421. URL https://doi.org/10.1145/3580305.3599421.