BALM-TSF: Balanced Multimodal Alignment for LLM-Based Time Series Forecasting

Shiqiao Zhou[1], Holger Schöner[2], Huanbo Lyu[1], Edouard Fouché[2], Shuo Wang[1]*

1. University of Birmingham, Birmingham, UK
2. Siemens AG, Munich, Germany

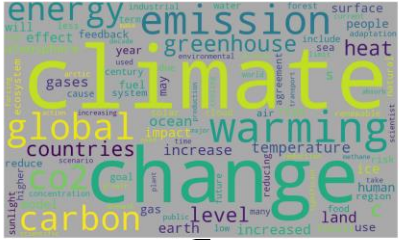
Background

Weather

Sensors

Energy

Text

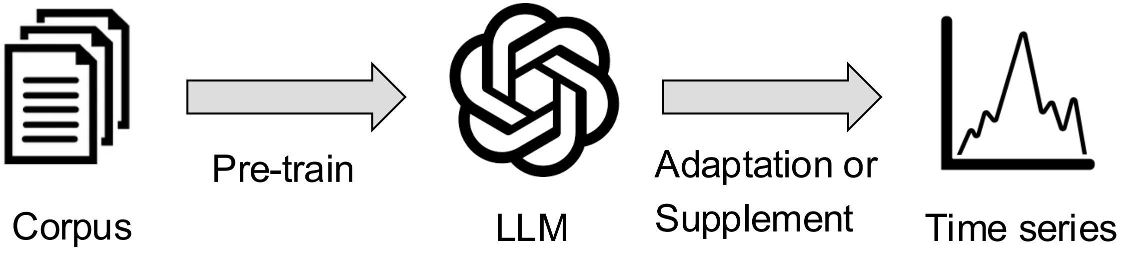


Time series data is often accompanied by rich textual information.

☺ **[What is goal of LLM-based forecasting?]**

Leveraging textual information for
Time series forecasting task with LLM

Background: LLM-based forecaster



- **There are some good LLM-based forecasters: TimeLLM (ICLR 2024), GPT4TS (ICLR 2024), FSCA (ICLR 2025), TimeCMA (AAAI 2025)…**

Issue: Modality Imbalance

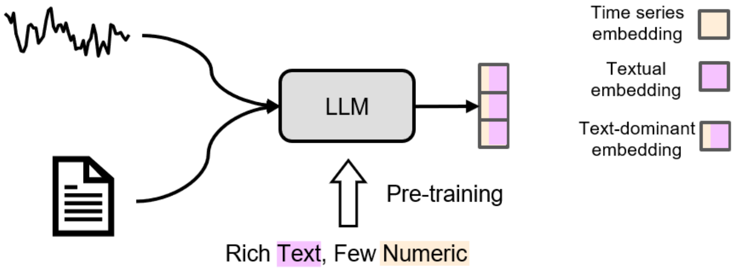
For the LLM-based time series forecaster, we found “modality imbalance issue”:

- **Semantics Imbalance**:

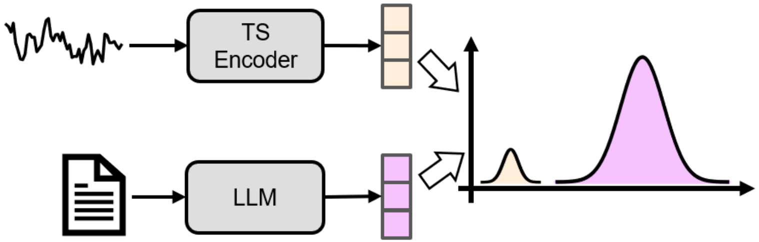
LLMs pretrained on rich text but sparse numeric data, resulting in “text-dominant” embeddings.

- **Distribution Imbalance**:

Separately encoded time series and text embeddings often have mismatched value ranges, resulting a “text-driven” training.



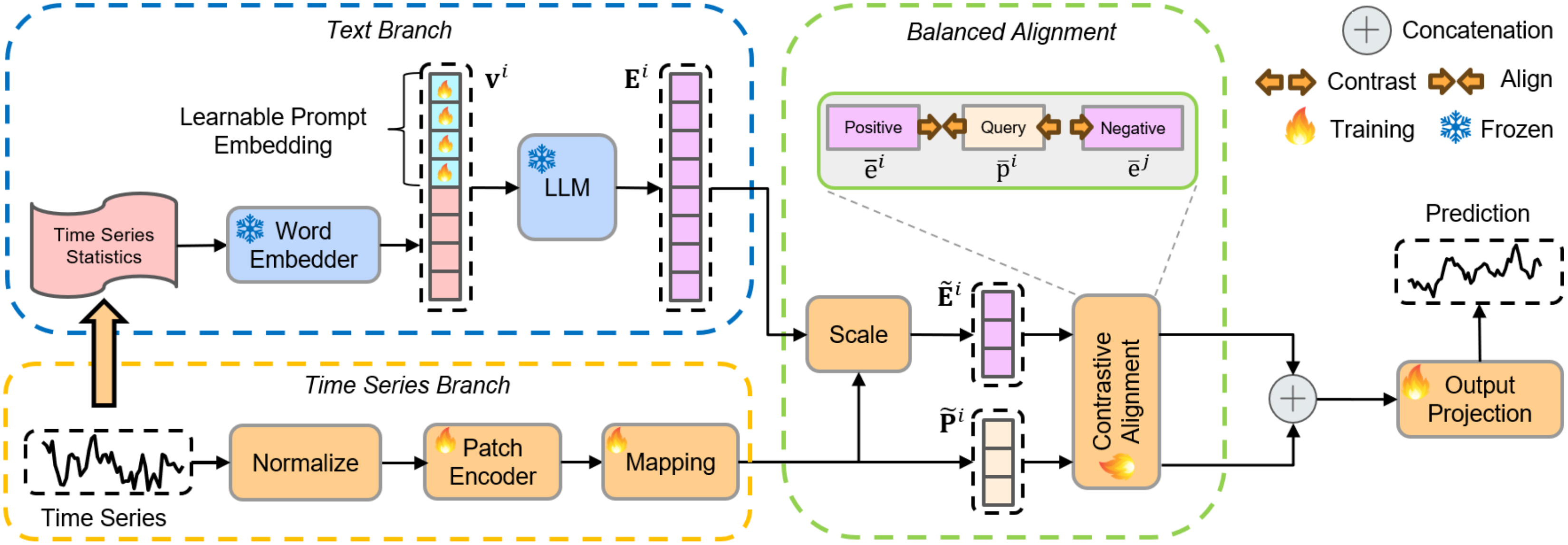
(a) Semantics Imbalance.



(b) Distribution Imbalance.

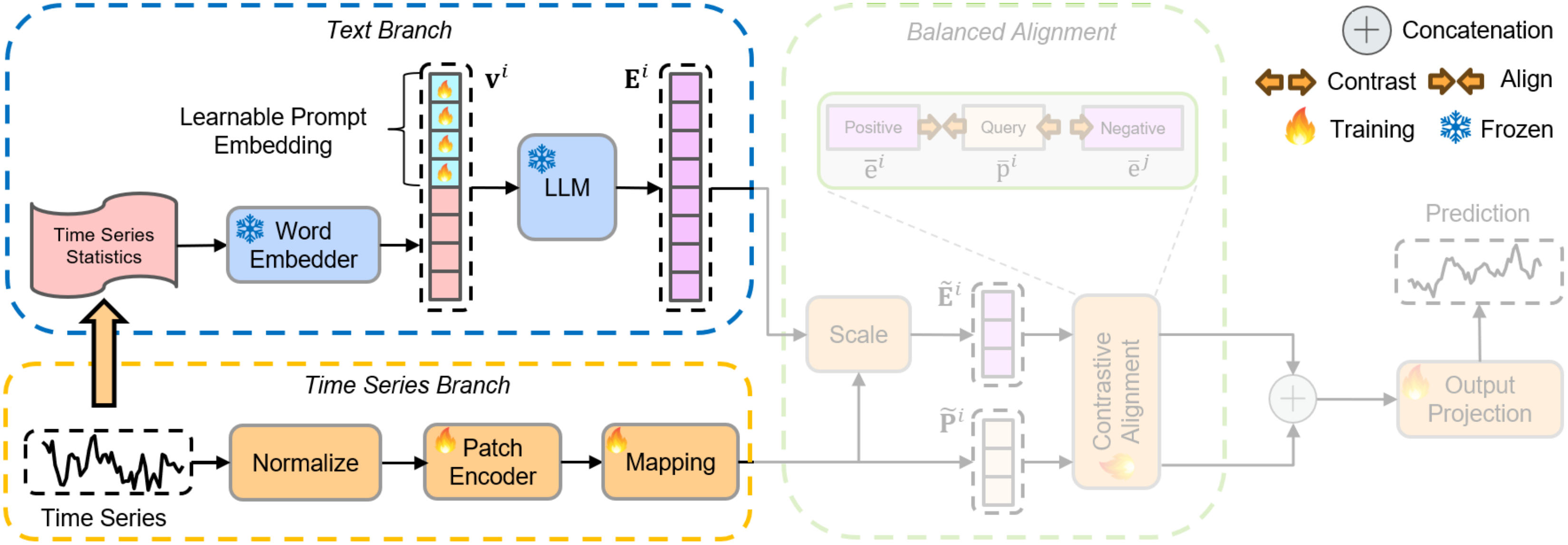
Method: BALM-TSF framework

- **Overall Structure: use “BALM” to smooth multiple modalities!**



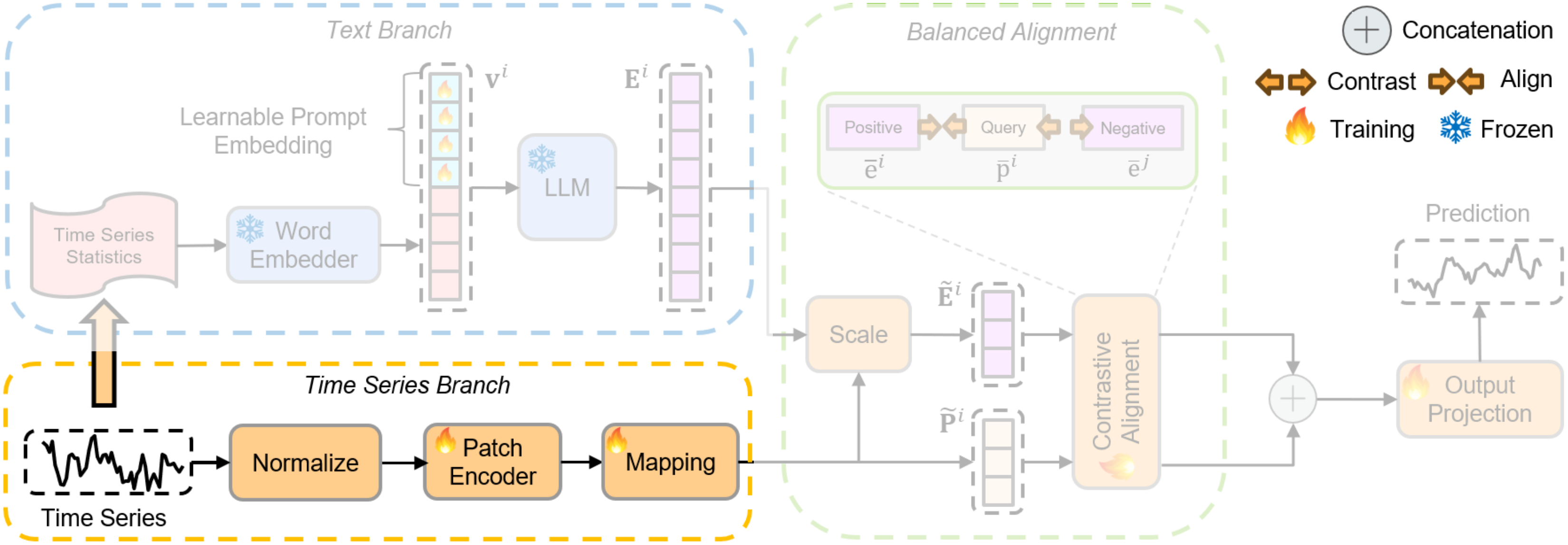
Method: BALM-TSF framework

- **Semantic Imbalance -> Solution: Text Branch & Time Series Branch**



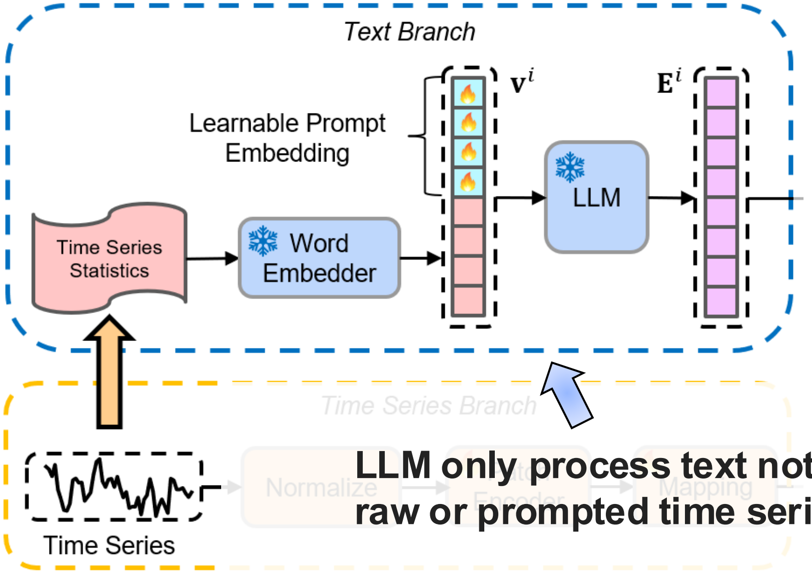
Method: BALM-TSF framework

- **Semantic Imbalance -> Solution: Text Branch & Time Series Branch**

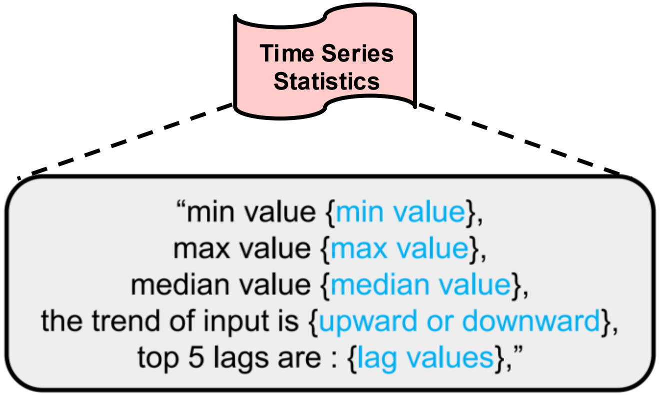


Method: BALM-TSF framework

- **Semantic Imbalance -> Solution: Text Branch & Time Series Branch**



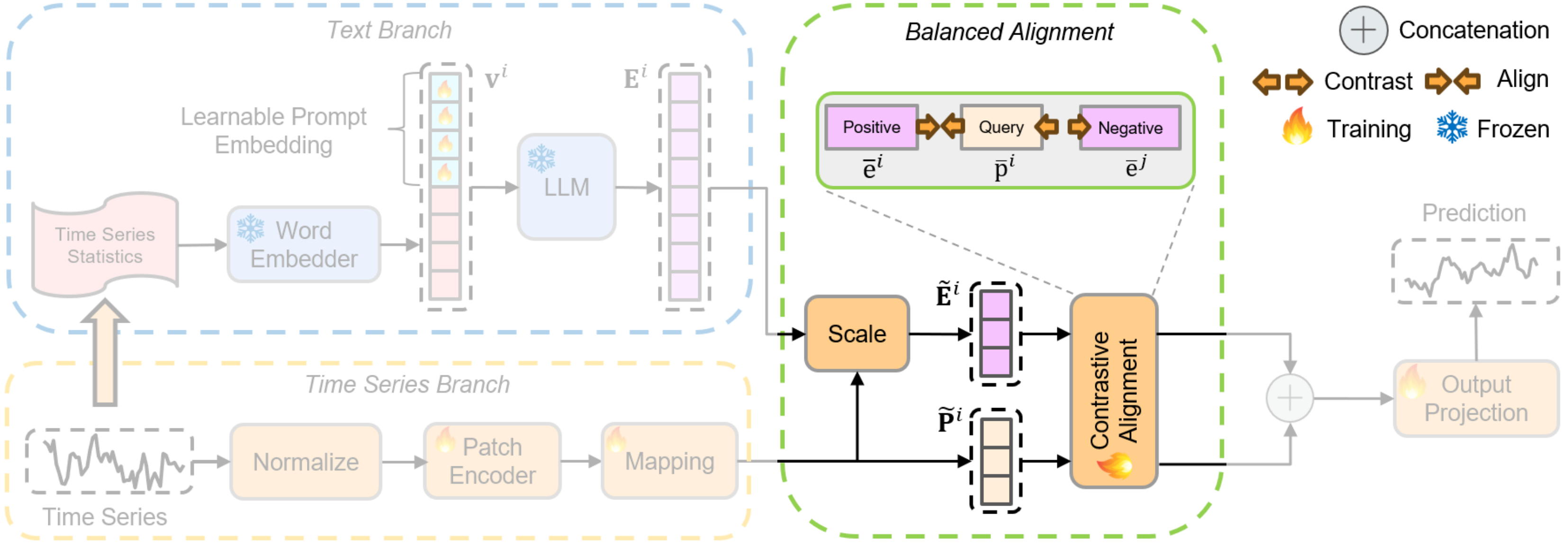
LLM only process text not raw or prompted time series.



Statistical prompt template for time series.

Method: BALM-TSF framework

- **Distribution Imbalance -> Solution :Balanced Multimodal Alignment**



Method: BALM-TSF framework

- **Distribution Imbalance -> Solution :Balanced Multimodal Alignment**

◆ **Two-step Scale:**

1. Truncate textual embeddings:

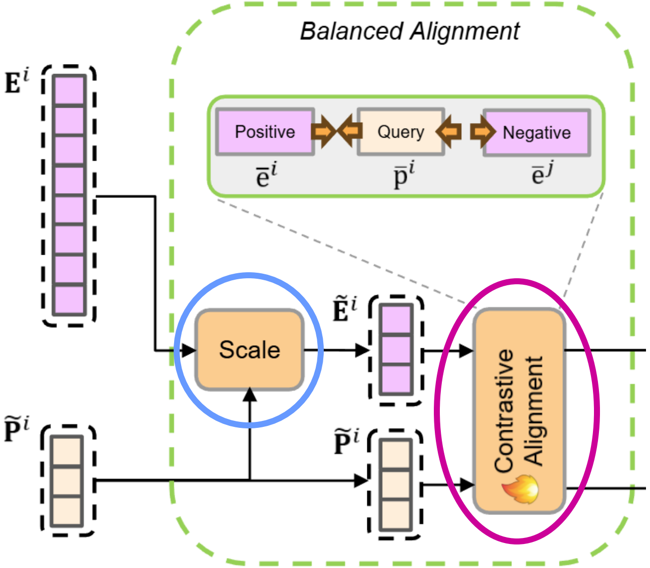
$$N_E = \min\left(N_P, \left\lfloor \frac{N_P \cdot H}{L} \right\rfloor\right)$$

$$\mathrm{E}^i_{\text{trunc}} = \mathrm{E}^i_{[N_T - N_E : N_T]} \in \mathbb{R}^{N_E \times d_{\text{LLM}}}$$

2. Rescale textual embeddings by STD:

$$\alpha = \frac{\text{STD}_{\text{time}}}{\text{STD}_{\text{text}}}$$

$$\tilde{\mathrm{E}}^i = \alpha\, \mathrm{E}^i_{\text{trunc}} \in \mathbb{R}^{N_E \times d_{\text{LLM}}}$$

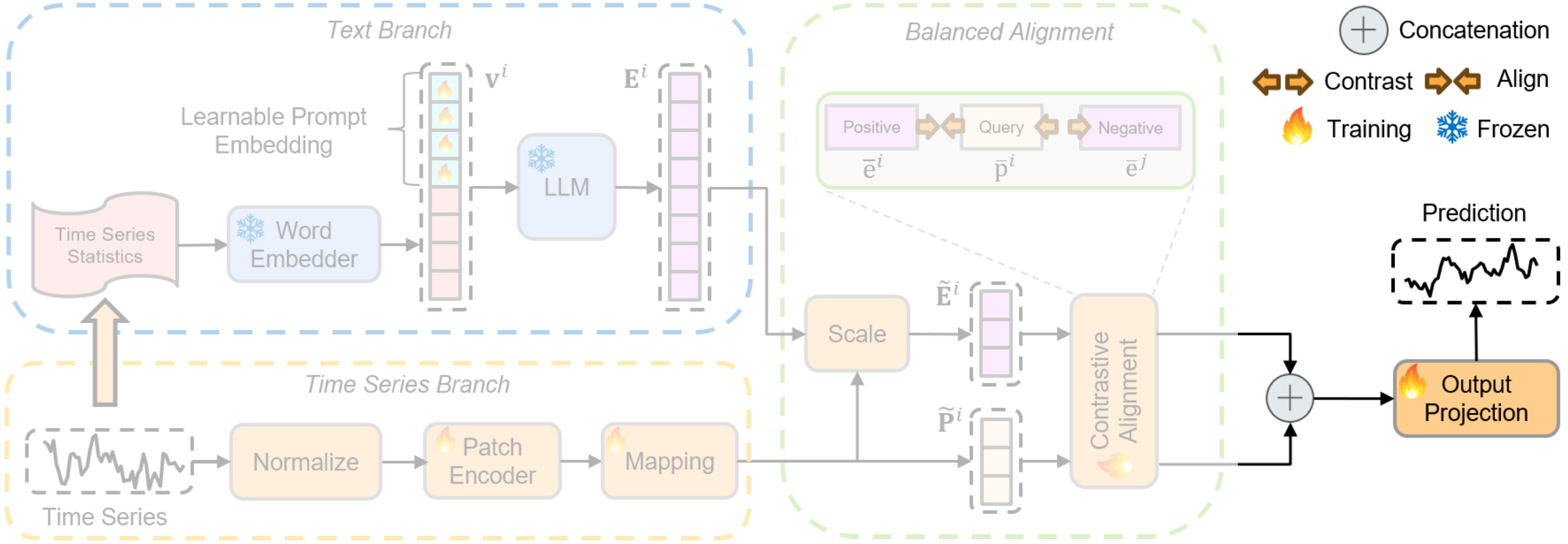


◆ **Semantic Alignment:**

$$\mathcal{L}_{align} = -\frac{1}{K}\sum_{i=1}^{K} \log \frac{\exp(\bar{\mathrm{p}}^i \cdot \bar{\mathrm{e}}^i / \tau)}{\sum_{j=1}^{K} \exp(\bar{\mathrm{p}}^i \cdot \bar{\mathrm{e}}^j / \tau)}$$

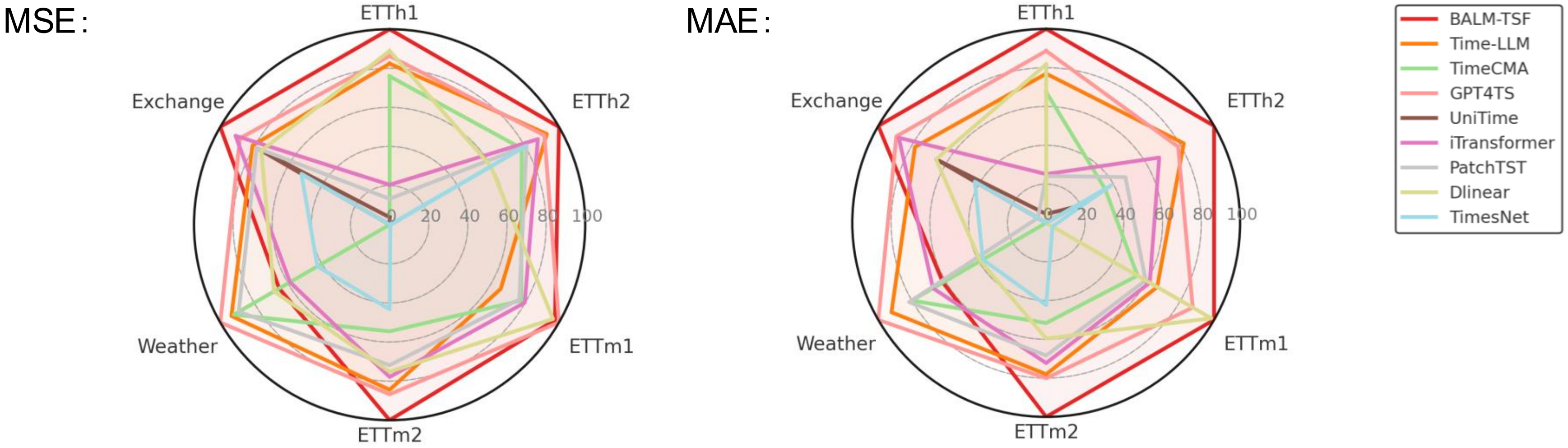
Method: BALM-TSF framework

- **Final loss function:** $\mathcal{L} = \mathcal{L}_{\text{task}} + \lambda \mathcal{L}_{\text{align}}$ **, where** $\mathcal{L}_{\text{task}}$ **is MSE loss.**



Experiments: Long-term forecasting

- **Average performance over four horizons (96, 192, 336, 720) on six datasets.**
- **We compared eight baselines including four LLM-based and four ML models.**

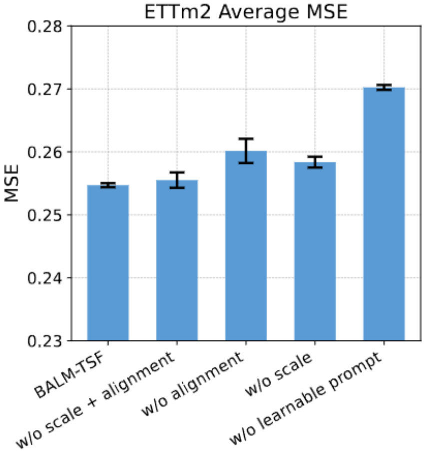


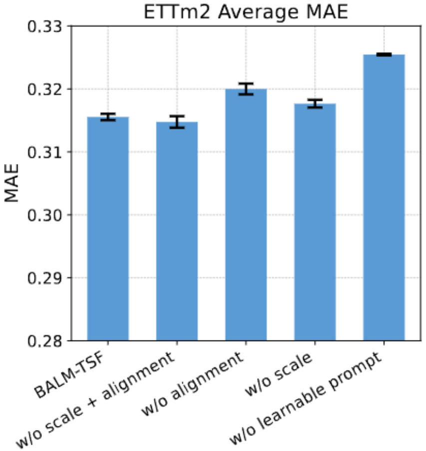
- Results show improvements of **8.9 %** in MSE and **5.8 %** in MAE on six real-world datasets.
- We also did few-shot forecasting experiments (with 10% training data), please check our paper.

Ablation Study

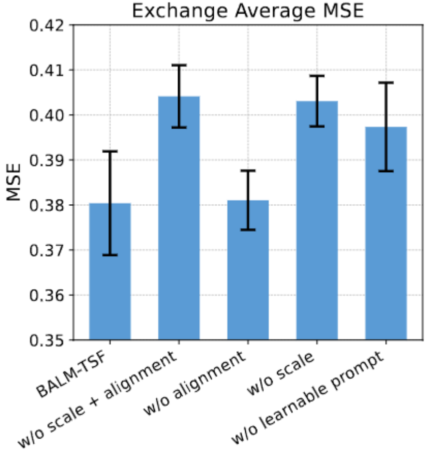
- **w/o alignment, w/o scale, w/o scale + alignment, and w/o learnable prompt**

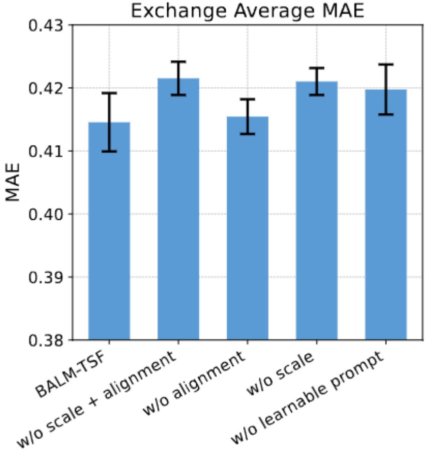
ETTm2





Exchange

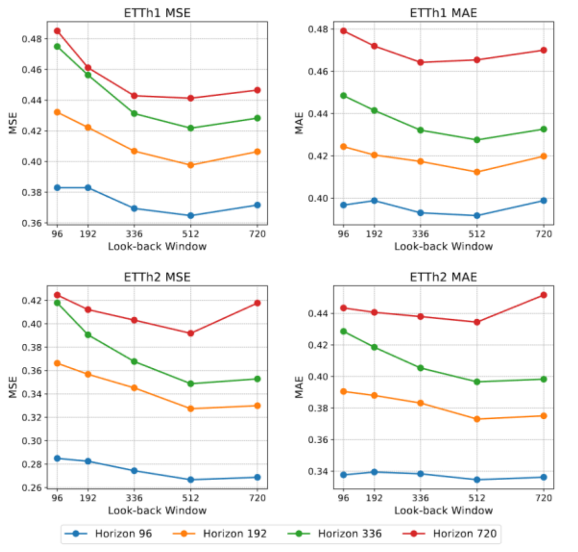




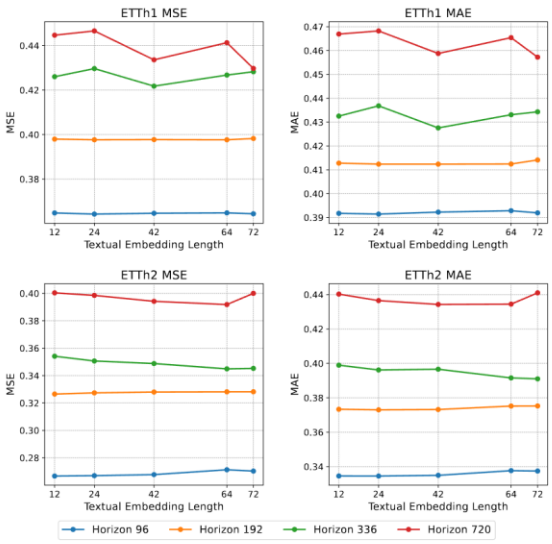
➢ Nearly all components offer measurable gains, yet their effectiveness varies across datasets.

Hyperparameter Sensitivity

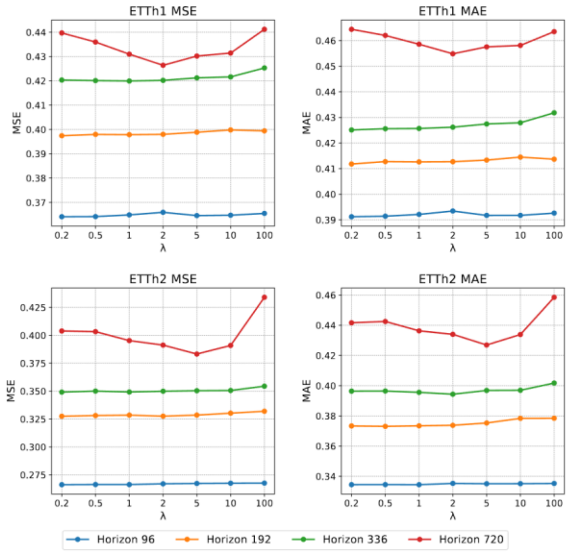
Look-back Window



Textual Embedding Length



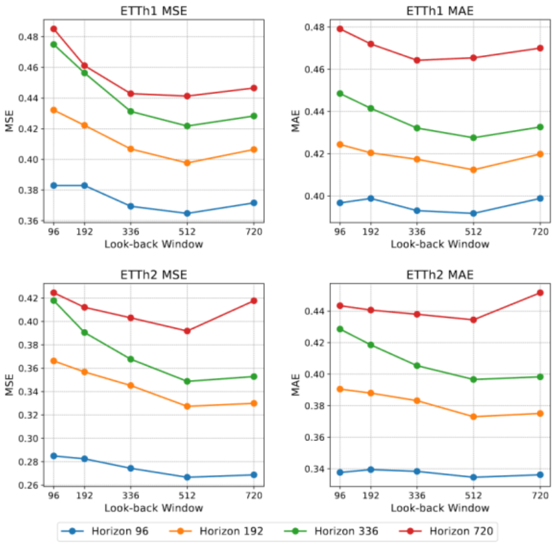
λ in Loss Function



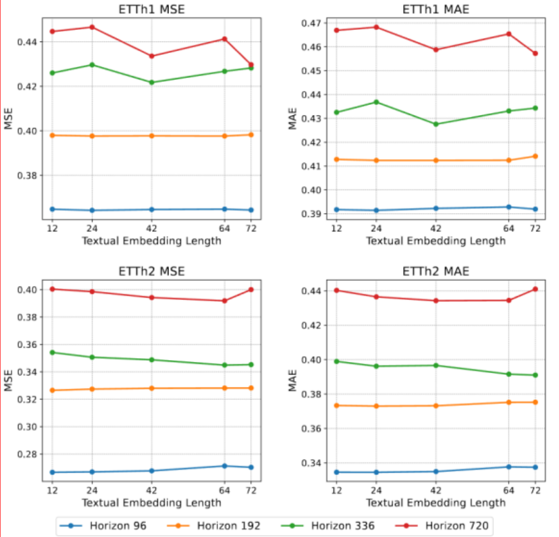
- There may be an optimal look-back window and λ, which indicates potential room for further optimization.

Hyperparameter Sensitivity

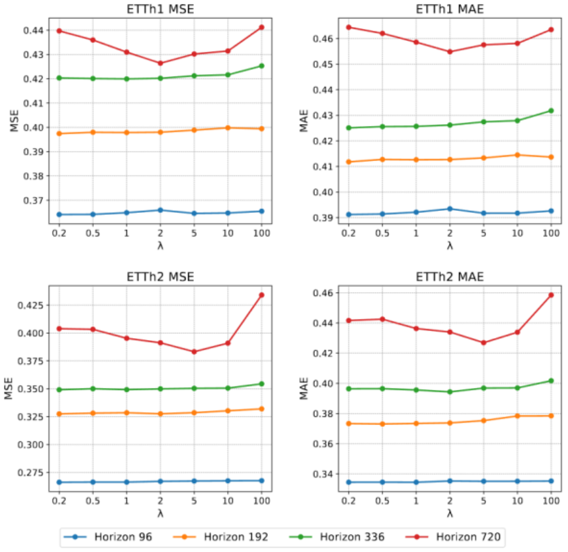
Look-back Window



Textual Embedding Length



λ in Loss Function



- ➢ Shorter forecast rely mainly on intrinsic temporal signals, while longer forecasts may benefit from richer statistical context in the textual prompts.

Efficiency Analysis

- **Efficiency Analysis between BALM-TSF with two LLM-based forecasters**

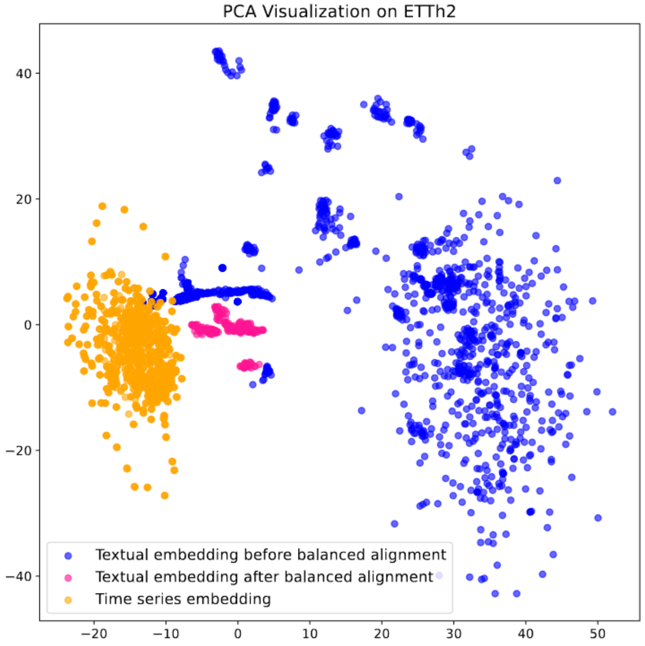
Table 3: Efficiency analysis of LLM-based models tested on the ETT-m1 dataset with a forecasting horizon of 96.

Models	LLM in model?	Total Params.	Trainable Params.	% Trained	Inference GPU memory	Inference speed
Time-LLM	yes	135.35M	53.44M	39.49%	895.65 MiB	0.0222s/iter
TimeCMA	no	**18.32M**	18.32M	100.00%	361.93 MiB	**0.0094s/iter**
BALM-TSF	yes	82.88M	**0.97M**	**1.17%**	**231.30 MiB**	0.0218s/iter

➢ By training less parameters (**1.17%** of the total), BALM-TSF delivers SOTA forecasting performance.

Visualization

- **Textual embedding before (blue) and after (pink) balanced multimodal alignment**
- **Time series embedding (orange)**



➢ Finding: Balanced Multimodal Alignment encourages textual embeddings to align more closely with the underlying time series representations

Summary

1. **Identifies the modality imbalance problem in LLM-based time-series forecasting, encompassing both semantic imbalance and embedding distribution imbalance.**

2. **Introduces BALM-TSF, a dual-branch architecture with learnable prompts and a balanced alignment strategy to mitigate modality imbalance.**

3. **Demonstrates that BALM-TSF achieves state-of-the-art performance with fewer parameters, providing a lightweight and strong method for time series forecasting.**

Thank you!

Source code: https://github.com/ShiqiaoZhou/BALM-TSF

Email me anytime: sxz363@student.bham.ac.uk