

Boundary-Aware Periodicity-based Sparsification Strategy for Ultra-Long Time Series Forecasting

A EXPERIMENTAL DETAILS

A.1 Dataset

The benchmark datasets utilized in multivariate time series forecasting tasks predominantly consist of the following eight datasets: Traffic, Electricity, Weather, Exchange, ETT (ETTM1/2, ETTTh1/2), and ILI. These benchmark datasets are widely recognized and employed across the field of multivariate time series forecasting, serving as standard evaluation benchmarks for assessing the performance of proposed models.

In time series forecasting tasks, existing mainstream models can achieve good prediction performance when the number of variables is relatively small. However, the data sparsification strategy proposed in this paper primarily addresses the issue of computational costs when dealing with excessively long sequence lengths and a large number of variables in ULTSF tasks. Therefore, for the experimental datasets in this paper, we specifically selected the Traffic, Electricity, and Weather datasets, as indicated in Table 1, which have a higher number of variables, while excluding datasets with fewer variables, such as the Exchange dataset with only 9 variables and the ETT (ETTM1/2, ETTTh1/2), and ILI datasets with only 8 variables. The detailed information of the Traffic, Electricity, and Weather datasets is as follows:

- (1) **Weather:** The Weather dataset[5] comprises 21 meteorological factors collected by the Max Planck Institute for Biogeochemistry’s weather station. The data spans from January 2020 to December 2020, covering a duration of one year.
- (2) **Electricity:** The Electricity dataset[5] consists of electricity consumption data from 321 customers, recorded at an hourly frequency. The data spans from July 2016 to July 2019, covering a duration of three years.
- (3) **Traffic:** The Traffic dataset[5] consists of road occupancy measurements from 862 sensors deployed in the San Francisco Bay Area highways. The data spans from July 2016 to July 2018, covering a duration of two years.

When the input length is set to 720, the detailed information about the data partitioning for the three benchmark datasets is presented in Table 1. In this table, the column “VaS” denotes the num-

Table 1: Details of the three benchmark datasets.

Dataset	VaS	(Train,Val,Test)	Fre	Span
Traffic	862	(10526,3510,3508)	1 Hour	120Days
Electricity	321	(31617,10540,10539)	1 Hour	120Days
Weather	21	(15782,5262,5260)	10 MIN	20Days

ber of variables in each dataset. The columns “(Train, Val, Test)” represent the number of samples allocated to the training, validation, and testing sets, respectively. The column “Fre” indicates the sampling interval of the timestamps, specifying how frequently the data is collected or recorded. Lastly, the column “Span” indicates

the actual time duration covered by the predicted results when the prediction length is set to 2880, considering the dataset’s sampling frequency.

A.2 Experimental Settings

Baselines In this paper, we have selected six well-acknowledged models with different structures as our baseline models. These models include: TiDE[1] based on the Multilayer Perceptron (MLP) architecture, Transformer-based Model, TimesNet[4] based on CNN, DLinear[6] model based on a pure linear structure, iTransformer[3] based on Transformers, and PatchTST[2]. These models have been widely recognized in the literature for their superior performance in time series forecasting tasks.

Experimental Environment. All models were trained and tested in the same experimental environment, which consisted of an NVIDIA GeForce RTX 4090 GPU with 24GB of memory.

Training Parameters. The training parameters for all models were set as follows: a learning rate of 0.001, a batch size of 32, 10 training epochs, the Adam optimizer, and the mean squared error (MSE) loss function.

Data Allocation. To ensure an equal distribution of samples for the validation set, the training and testing datasets were partitioned in the same manner for all models. The training set accounted for 60% of the data, the validation set accounted for 20%, and the testing set accounted for the remaining 20%. The detailed information about the partitioning of each dataset is presented in Table 1.

The Length of Inputs and Outputs. In previous studies on LTSF tasks, a prediction length of 720 has already been achieved. However, this paper focuses on exploring ULTSF, where the prediction length is doubled compared to the maximum length of the LTSF. The length of the experimental backtracking window is set to 720, which is 7.5 times the original length, and the prediction sequence lengths are set to 1440, 2160, 2880, which are twice, three times, and four times the length of the original longest sequence, respectively.

Evaluation Metrics. The evaluation metrics employed for assessing the model performance were the mean squared error (MSE) and the mean absolute error (MAE).

Model Variants. This model possesses two hyperparameters, namely the number of periodic features k and the model dimension d_{model} . To conduct comparative experiments, we set k values to 2 and 3 across all datasets. Additionally, we varied the d_{model} parameter for different datasets, as outlined in the hyperparameter comparative experiments.

REFERENCES

- [1] Abhimanyu Das, Weihao Kong, Andrew Leach, Shaan Mathur, Rajat Sen, and Rose Yu. arXiv preprint arXiv:2304.08424,2023. Long-term Forecasting with TiDE: Time-series Dense Encoder. (arXiv preprint arXiv:2304.08424,2023).
- [2] Yuqi Nie1, Nam H. Nguyen2, Phanwadee Sinthong2, and Jayant Kalagnanam. 2023. A TIME SERIES IS WORTH 64 WORDS: LONG-TERM FORECASTING WITH TRANSFORMERS. In *International Conference on Learning Representations (ICLR)*.

- [3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. In *Advances in neural information processing systems*.
- [4] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. 2023. TIMESNET: TEMPORAL 2D-VARIATION MODELING FOR GENERAL TIME SERIES ANALYSIS. In *International Conference on Learning Representations(ICLR)*.
- [5] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. 2021. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In *Conference and Workshop on Neural Information Processing Systems(NeurIPS)*.
- [6] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. 2023. Are Transformers Effective for Time Series Forecasting?. In *Association for the Advancement of Artificial Intelligence(AAAI)*.