Ran TIAN, Xinmei LI, Zhongyu MA, Yanxing LIU, Jingxia WANG, Chu WANG, 2023. LDformer: a parallel neural network model for long-term power forecasting. *Frontiers of Information Technology & Electronic Engineering*, 24(9):1287-1301. https://doi.org/10.1631/FITEE.2200540

Front Inform Technol Electron Eng

LDformer: a parallel neural network model for long-term power forecasting

Key words: Long-term power forecasting; Long short-term memory (LSTM); UniDrop; Self-attention mechanism

Corresponding author: Ran TIAN
E-mail: tianran@nwnu.edu.cn
ORCID: https://orcid.org/0000-0003-4435-580X

Motivation

- Due to the highly complex and large-scale long-time-series data, traditional methods are limited in efficiently handling high dimensional large data and representing complex functions. It will forget some data and ignore the long-term dependency of the time-series data.
- When conducting long-term power forecasting, the model structure of traditional methods is not stable. Therefore, in order to improve the accuracy and reliability of long-term power forecasting, it is necessary to adopt more stable and accurate for prediction analysis.
- The traditional attention mechanism is prone to losing key connections between sequences when capturing the dependencies among long-time-series data, leading to the degradation of prediction performance.

Front Inform Technol Electron Eng

Method

- We propose a parallel neural network model called LDformer, which combines the advantages of Informer and long short-term memory (LSTM) to effectively solve the problem of long-term power forecasting. Using LSTM to learn the long-term correlation of time-series data, the model has good long-term forecasting performance.
- We propose a parallel encoder module that combines a convolutional layer and an attention mechanism to avoid value redundancy in the attention mechanism. Several experiments on different datasets validate the effectiveness and robustness of the parallel encoder and the convolutional layer.
- We propose a probabilistic sparse (ProbSparse) self-attention mechanism combined with UniDrop. UniDrop does not need additional computational overhead or external resources. Combining UniDrop, ProbSparse attention mechanism can mitigate the risk of losing some key connections in the sequence.

Front Inform Technol Electron Eng

Long-time-series prediction task

In the prediction setting with a fixed length, we have input

$$X^t=\left\{x_1^t, \cdots, x_{L_i}^t \,\middle|\, x_i^t \in \mathbb{R}^{\mathrm{dx}}, i = 1, 2, \cdots, L_\mathrm{i}\right\}$$

L_i is the number of input features

output at time t

$$Y^t=\left\{y_1^t, \dots, y_{L_o}^t \,\middle|\, y_j^t \in \mathbb{R}^{\mathrm{dy}}, j = 1, 2, \cdots, L_\mathrm{o}\right\}$$

L_o is the number of output features

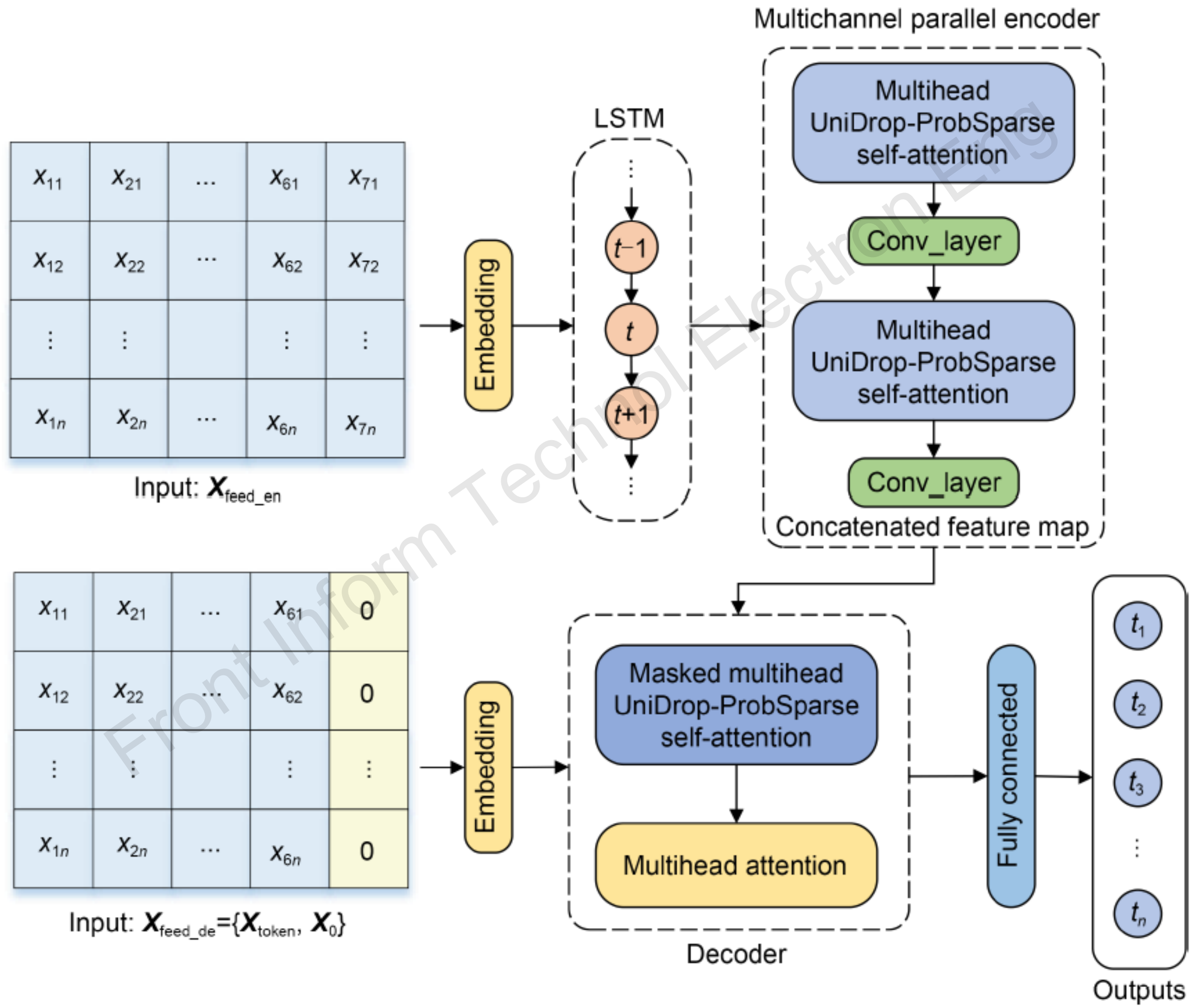
Front Inform Technol Electron Eng

The objective of long-term power forecasting is to map the historical power load profile at m time steps to the target feature at n time steps by the model $f\,()$

$$[X^{t-m+1}, X^{t-m+2}, \cdots, X^t] \xrightarrow{f(\cdot)} [y^{t+1}, y^{t+2}, \cdots, y^{t+n}].$$

Long-time-series prediction model

LDformer contains LSTM, encoder, and decoder structures.



Framework of LDformer

Embedding layers with multiple perspectives

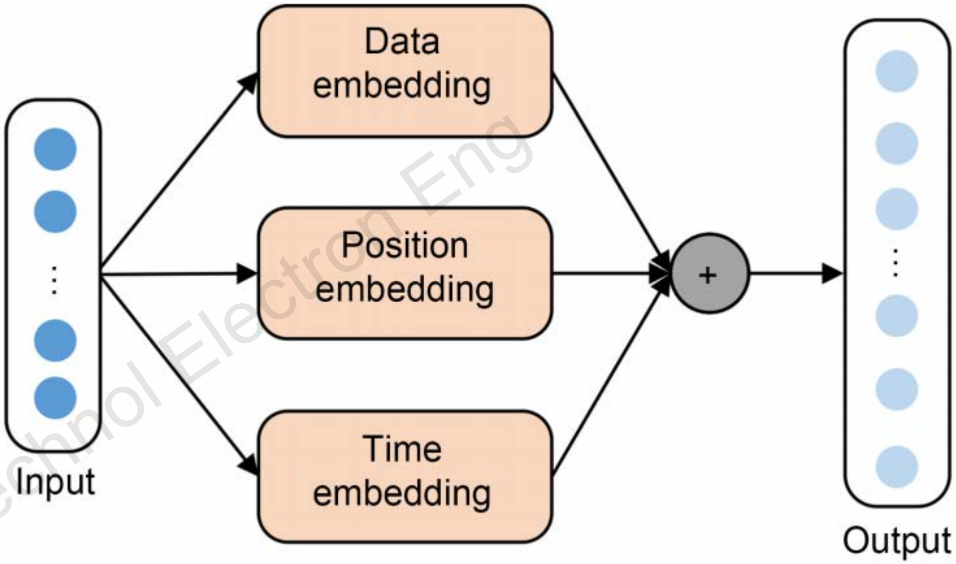
- Data embedding (DE)

$$\mathrm{DE}=\mathrm{conv1D}\left(x_{\mathrm{in}}, d_{\mathrm{model}}\right).$$

- Position embedding (PE) (Vaswani et al., 2017)

$$\mathrm{PE}(\mathrm{pos}, 2i) = \sin\left(\mathrm{pos}/10\,000^{\frac{2i}{d_{\mathrm{model}}}}\right),$$

$$\mathrm{PE}(\mathrm{pos}, 2i+1) = \cos\left(\mathrm{pos}/10\,000^{\frac{2i}{d_{\mathrm{model}}}}\right),$$



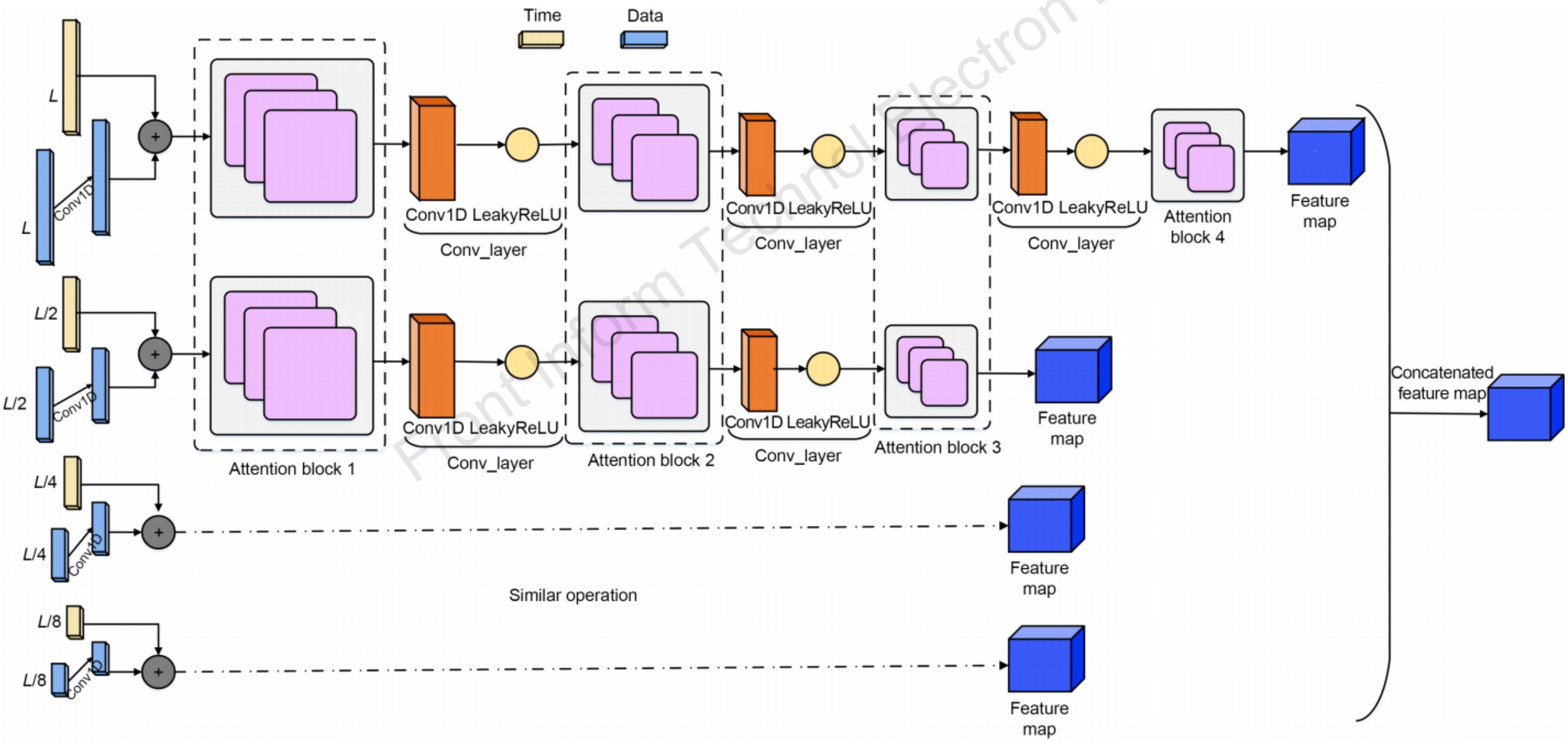
- Time embedding:

The time slice of the dataset used is mainly in hours and minutes, so hour_embed and minute_embed are chosen to obtain the timestamp encoding results.

Front Inform Technol Electron Eng

Multichannel parallel encoder module

- We build the encoder module by combining the attention mechanism with the convolutional layer. It takes four channels of length *L*, *L*/2, *L*/4, and *L*/8, and executes them in parallel.



Front Inform Technol Electron Eng

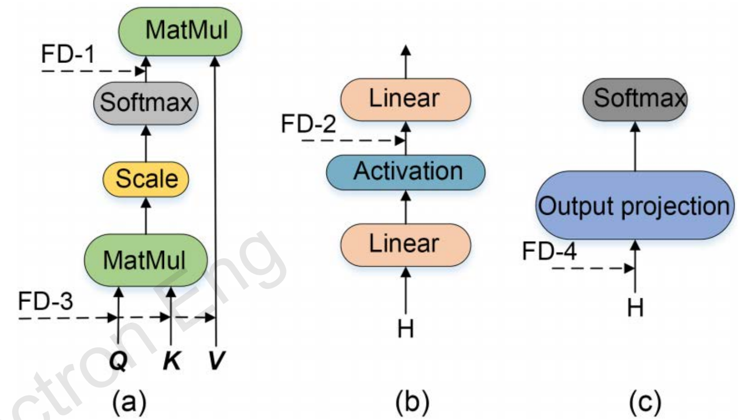
ProbSparse self-attention mechanism combined with UniDrop

- The canonical self-attention mechanism consists of a query and a set of key-value pairs. The formula is as follows (Vaswani et al., 2017):

$$A(Q, K, V) = \mathrm{Softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d}}\right)V$$

- Considering the time complexity and the risk of losing some key connections in sequences, we propose a ProbSparse self-attention mechanism combined with UniDrop.

$$\left.\begin{aligned}
&A(\boldsymbol{q}_i', \boldsymbol{K}', \boldsymbol{V}') = \sum_j \frac{k(\boldsymbol{q}_i', \boldsymbol{k}_j')}{\sum_l k(\boldsymbol{q}_i', \boldsymbol{k}_l')}\boldsymbol{v}_j' = E_{p(\boldsymbol{k}_j'|\boldsymbol{q}_i')}\left[\boldsymbol{v}_j'\right]. \\
&p(\boldsymbol{k}_j'|\boldsymbol{q}_i') = \frac{k(\boldsymbol{q}_i', \boldsymbol{k}_j')}{\sum_l k(\boldsymbol{q}_i', \boldsymbol{k}_l')}, \\
&\mathrm{KL}(q\|p) = \ln\sum_{l=1}^{L_{K'}} \mathrm{e}^{\frac{\boldsymbol{q}_i'(\boldsymbol{k}_l')^{\mathrm{T}}}{\sqrt{d}}} - \frac{1}{L_{K'}}\sum_{j=1}^{L_{K'}}\frac{\boldsymbol{q}_i'(\boldsymbol{k}_j')^{\mathrm{T}}}{\sqrt{d}} - \ln L_{K'}. \\
&M(\boldsymbol{q}_i', \boldsymbol{K}') = \ln\sum_{j=1}^{L_{K'}} \mathrm{e}^{\frac{\boldsymbol{q}_i'(\boldsymbol{k}_j')^{\mathrm{T}}}{\sqrt{d}}} - \frac{1}{L_{K'}}\sum_{j=1}^{L_{K'}}\frac{\boldsymbol{q}_i'(\boldsymbol{k}_j')^{\mathrm{T}}}{\sqrt{d}}.
\end{aligned}\right\} \quad A(\boldsymbol{Q}', \boldsymbol{K}', \boldsymbol{V}') = \mathrm{Softmax}\left(\frac{\bar{\boldsymbol{Q}}'(\boldsymbol{K}')^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V}',$$



UniDrop structure: (a) attention; (b) feed forward;(c) output prediction

Algorithm 1 ProbSparse self-attention mechanism combined with UniDrop

Input: tensors $\boldsymbol{Q}\in\mathbb{R}^{m\times d}$, $\boldsymbol{K}\in\mathbb{R}^{n\times d}$, $\boldsymbol{V}\in\mathbb{R}^{n\times d}$

1 Initialization: set hyperparameter c, u=$c$ln m, U=$m$ln n, and dropout parameter p

2 $\boldsymbol{Q}'$=dropout($\boldsymbol{Q}$), $\boldsymbol{K}'$=dropout($\boldsymbol{K}$), $\boldsymbol{V}'$=dropout($\boldsymbol{V}$)

3 Arbitrarily select U dot-product pairs from $\boldsymbol{K}'$ as $\bar{\boldsymbol{K}}'$

4 Sample $\boldsymbol{K}'$ and set the sample score $\bar{\boldsymbol{S}}$=$\boldsymbol{Q}'(\bar{\boldsymbol{K}}')^{\mathrm{T}}$

5 Each $\boldsymbol{q}_i$ on K_sample calculates the M value

6 Set top-u queries as $\bar{\boldsymbol{Q}}'$ based on formula M

7 Set $\boldsymbol{A}_0$=Softmax$\left(\bar{\boldsymbol{Q}}'(\boldsymbol{K}')^{\mathrm{T}}/\sqrt{d}\right)\boldsymbol{V}'$

8 For the value of the score of unchecked $\boldsymbol{q}_i'$ set $\boldsymbol{A}_1$=mean$(\boldsymbol{V}')$

9 Set self-attention formula $\boldsymbol{A}$=$\{\boldsymbol{A}_0, \boldsymbol{A}_1\}$

Output: self-attention feature map $\boldsymbol{A}$

Front Inform Technol Electron Eng

Long-time-series prediction tasks

- We extensively perform experiments on five datasets, namely, the ETTm1, ETTh1, ETTh2, PEMS03, and weather datasets. ETTm1, ETTh1, and ETTh2 are collectively referred to as the ETT dataset.

Table 1 Datasets and prediction task descriptions

Dataset	Data description	Time range	Time interval	Target feature	Target length	Length of time (h)
ETTm1	Key indicators for long-term power deployment collected from two different counties in China	07/01/2016–06/26/2018	15 min	Oil temperature	{24, 36, 48, 96, 168, 336}	{6, 9, 12, 24, 42, 84}
ETTh1		07/01/2016–06/26/2018	1 h	Oil temperature	{24, 36, 48, 96, 168, 336}	{24, 36, 48, 96, 168, 336}
ETTh2		07/01/2016–06/26/2018	1 h	Oil temperature	{24, 36, 48, 96, 168, 336}	{24, 36, 48, 96, 168, 336}
Weather	Local climate data for nearly 1600 U.S. locations	01/01/2010–12/31/2013	1 h	Wet bulb	{24, 48, 96, 168, 336, 720}	{24, 48, 96, 168, 336, 720}
PEMS03	Traffic flow dataset	09/01/2018–11/30/2018	5 min	313 512	{24, 48, 96, 228, 336, 720}	{2, 4, 8, 24, 28, 60}

Target length means number of time interval. Length of time=time interval×target length

- Assessment metrics: three evaluation indexes, mean square error (MSE), mean absolute error (MAE), and root mean square error (RMSE) are used.

Front Inform Technol Electron Eng

Major results

Table 2 Performance comparison of short-time-series prediction tasks in the ETT dataset at different prediction lengths (24, 36, 48)

Dataset	Model	MSE			MAE			RMSE		
		24	36	48	24	36	48	24	36	48
ETTm1	RNN	5.0206	2.7855	5.7508	1.9853	1.4088	2.1783	2.2407	1.6689	2.3981
	LSTM	2.6396	2.7410	2.7252	1.3441	1.3751	1.3704	1.6247	1.6555	1.6508
	GRU	1.9783	5.0422	3.0284	1.1282	1.9861	1.4052	1.4065	2.1454	1.7402
	Informer (np)	0.0904	0.0816	0.0739*	0.2279	0.2247	0.2117*	0.3006	0.2852	0.2719*
	Informer	**0.0432**	**0.0674**	0.1055	**0.1606**	**0.2021**	0.2571	**0.2079**	**0.2597**	0.3248
	LDformer (np)	0.0605*	0.0800	0.0999	0.1962*	0.2271	0.2574	0.2458*	0.2830	0.3161
	LDformer	0.0789	0.0754*	**0.0665**	0.2301	0.2136*	**0.1993**	0.2810	0.2747*	**0.2578**
ETTh1	RNN	3.6319	4.3593	4.6225	1.6149	1.8123	1.8864	1.9057	2.0879	2.1500
	LSTM	2.1395	2.7266	2.5198	1.1835	1.3697	1.3068	1.4627	1.6512	1.5874
	GRU	0.9840	0.9556	1.9024	0.7832	0.7625	1.1236	0.9920	0.9775	1.3792
	Informer (np)	0.3420*	0.3861*	0.5013	0.5325*	0.5532*	0.6441	0.5848*	0.6314*	0.7080
	Informer	0.4798	0.5079	0.3964*	0.6475	0.6685	**0.5406**	0.6926	0.7127	0.6296*
	LDformer (np)	0.4193	**0.3458**	0.5183	0.5937	**0.5314**	0.6545	0.6475	**0.5880**	0.7199
	LDformer	**0.2492**	0.4304	**0.3782**	**0.4361**	0.5859	0.5579*	**0.4992**	0.6560	**0.6150**
ETTh2	RNN	2.2385	2.0350	1.4774	1.2577	1.1824	1.0111	1.4961	1.4265	1.2155
	LSTM	1.0688	1.1154	**0.6622**	0.8114	0.8529	**0.6543**	1.0338	1.0561	**0.8138**
	GRU	0.7586*	0.7955	1.0504	0.7136*	0.7204*	0.8336	0.8710*	0.8919	1.0249
	Informer (np)	1.0555	1.2231	1.7355	0.9186	1.0052	1.2214	1.0273	1.1059	1.3173
	Informer	1.2559	1.1121	2.0611	1.0338	0.9529	1.3422	1.1206	1.0545	1.4356
	LDformer (np)	0.7810	**0.5988**	0.7493	0.7721	**0.6642**	0.9541	0.8837	**0.7738**	1.0720
	LDformer	**0.5713**	0.7553*	0.7224*	**0.6384**	0.7575	0.7296*	**0.7558**	0.8691*	0.8499*

The best results are in bold, and * denotes the second-best results. MSE: mean square error; MAE: mean absolute error; RMSE: root mean square error

Front Inform Technol Electron Eng

Major results

Table 3 Performance comparison of long-time-series prediction tasks in the ETT dataset at different prediction lengths (96, 168, 336)

Dataset	Model	MSE			MAE			RMSE		
		96	168	336	96	168	336	96	168	336
ETTm1	RNN	3.7047	4.9325	3.4728	1.6129	1.9187	1.5242	1.9247	2.2209	1.8635
	LSTM	2.7349	2.7401	2.7471	1.3728	1.3749	1.3763	1.6537	1.6553	1.6574
	GRU	3.6131	2.8525	3.1644	1.5948	1.4074	1.4829	1.9008	1.6889	1.7788
	Informer (np)	0.3387*	0.6584	1.0142	0.5361	0.7109	0.9140	0.5820*	0.8114	1.0070
	Informer	0.5603	0.6100	1.0266	0.6874	0.7097	0.9412	0.7485	0.7810	1.0132
	LDformer (np)	**0.3248**	0.5146*	**0.5357**	**0.5142**	0.6573*	**0.6519**	**0.5699**	0.7173*	**0.7319**
	LDformer	0.3523	**0.5096**	0.6346*	0.5277*	**0.6166**	0.7103*	0.5936	**0.7138**	0.7966*
ETTh1	RNN	3.5576	3.4802	2.8218	1.5451	1.5640	1.3619	1.8861	1.8655	1.6798
	LSTM	2.2462	2.2827	2.0860	1.2268	1.2244	1.1430	1.4987	1.5108	1.4443
	GRU	2.6474	1.8953	2.3861	1.3466	1.0810	1.2097	1.6271	1.3767	1.5447
	Informer (np)	**0.2577**	0.3024*	0.7343	**0.4236**	0.4721*	0.7743	**0.5076**	0.5499*	0.8569
	Informer	0.3014*	**0.2656**	0.9416	0.4639*	**0.4413**	0.8984	0.5490*	**0.5154**	0.9703
	LDformer (np)	0.4545	0.4046	0.5057*	0.6069	0.5617	**0.6421**	0.6741	0.5615	0.7111*
	LDformer	0.3106	0.4886	**0.5034**	0.4880	0.6322	0.6455*	0.5569	0.5224	**0.7095**
ETTh2	RNN	1.5415	2.0284	2.1122	1.0042	1.1726	1.1998	1.2416	1.4242	1.4533
	LSTM	**1.0823**	1.1039*	**1.1152**	**0.8412**	0.8520*	0.8568*	**1.0403**	1.0507*	**1.0560**
	GRU	1.1424*	**0.8537**	1.1582*	0.8764*	**0.7320**	**0.8390**	1.0688*	**0.9239**	1.0762*
	Informer (np)	2.2963	2.3733	2.2128	1.4216	1.4409	1.3732	1.5153	1.5405	1.4875
	Informer	2.7244	2.9010	2.1025	1.5558	1.6124	1.3379	1.6506	1.7032	1.4500
	LDformer (np)	2.1959	2.2004	2.8714	1.3798	1.3807	1.5847	1.4818	1.4834	1.6945
	LDformer	1.8590	2.5229	2.6768	1.2490	1.4819	1.5261	1.3634	1.5883	1.6361

Front Inform Technol Electron Eng

The best results are in bold, and * denotes the second-best results. MSE: mean square error; MAE: mean absolute error; RMSE: root mean square error

Major results

Table 4 Performance comparison of short-time-series prediction tasks in the weather and PEMS03 datasets at different prediction lengths (24, 48, 96)

Dataset	Model	MSE			MAE			RMSE		
		24	48	96	24	48	96	24	48	96
Weather	RNN	0.9135	1.0542	1.5387	0.5446	0.6193	0.8906	0.9558	1.0267	1.2404
	LSTM	0.6073	0.6203	1.0249	0.4037	0.5168	0.5648*	0.7793	0.7875	1.0123
	GRU	0.4992	0.5184	0.7345	0.3378	0.5661	0.6472	0.7066	0.7200	0.7965
	Informer (np)	0.1666	0.3414*	0.7039	0.2862	0.4285*	0.6236	0.4082	0.5843*	0.8390
	Informer	0.1594*	0.3802	0.6459	0.2809*	0.4552	0.5981	0.3993*	0.6166	0.8037
	LDformer (np)	0.1650	**0.3393**	**0.5638**	0.2905	**0.4279**	**0.5594**	0.4062	**0.5825**	**0.7508**
	LDformer	**0.1574**	0.3677	0.5653*	**0.2918**	0.4484	0.5672	**0.3968**	0.6064	0.7519*
PEMS03	RNN	0.4110	0.1498	0.8322	0.5089	0.3166	0.7091	0.6411	0.3871	0.9122
	LSTM	0.0776	0.0835*	0.2132	0.2191	0.2291	0.2644	0.2786	0.3089	0.4364
	GRU	0.0779	0.0842	0.2098	0.2206	0.2309	0.3635	0.2791	0.3002	0.4313
	Informer (np)	0.0582	0.0970	**0.1145**	0.1772	0.2158	**0.2377**	0.2413	0.3115	**0.3384**
	Informer	0.0553*	0.0891	0.1195*	0.1705*	0.2070	0.2426*	0.2352*	0.2986*	0.3457*
	LDformer (np)	**0.0551**	0.0915	0.1507	**0.1688**	**0.2054**	0.2568	**0.2348**	0.3024	0.3882
	LDformer	0.0650	**0.0865**	0.1279	0.1788	0.2055*	0.2400	0.2550	**0.2942**	0.3576

The best results are in bold, and * denotes the second-best results. MSE: mean square error; MAE: mean absolute error; RMSE: root mean square error

Front Inform Technol Electron Eng

Major results

Table 5 Performance comparison of long-time-series prediction tasks in the weather dataset at prediction lengths of 168, 336, and 720 and the PEMS03 dataset at prediction lengths of 288, 336, and 720

Dataset	Model	MSE			MAE			RMSE		
		168	336	720	168	336	720	168	336	720
Weather	RNN	1.6785	1.9695	1.9957	1.0261	1.1659	1.1651	1.2955	1.4034	1.4126
	LSTM	0.9975	1.0505	1.0395	0.6133*	0.7713	0.7682	0.9987	1.0249	1.0195
	GRU	0.7594	0.8735	1.6688	0.6360	0.6890	0.9437	0.8714	0.9346	1.2918
	Informer (np)	0.8772	0.8803	0.8857	0.7056	0.7128	0.7244	0.9366	0.9382	0.9411
	Informer	0.7835	0.8791	0.8743	0.6681	0.7196	0.7194	0.8851	0.9376	0.9350
	LDformer (np)	0.6598*	**0.7277**	**0.7751**	0.6134	0.6485*	**0.6699**	0.8122*	**0.8531**	**0.8804**
	LDformer	**0.6194**	0.7452*	0.7881*	**0.6040**	**0.6477**	0.6807*	**0.7870**	0.8632*	0.8877*

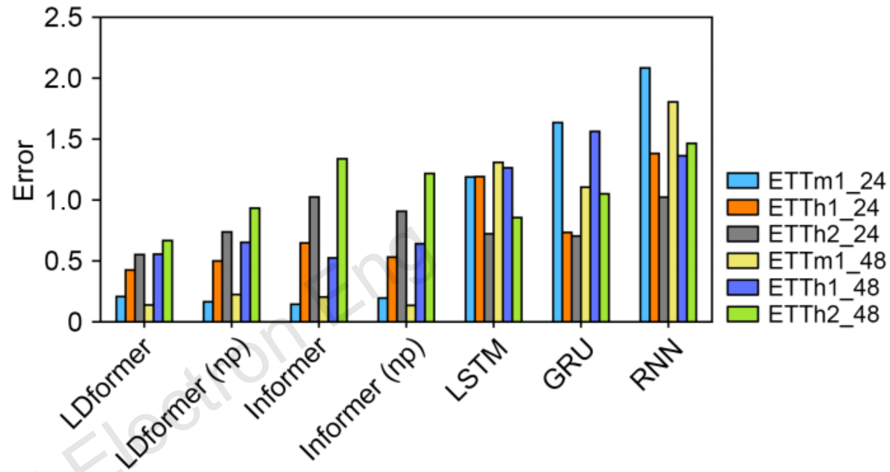
Dataset	Model	MSE			MAE			RMSE		
		288	336	720	288	336	720	288	336	720
PEMS03	RNN	1.1370	1.0338	1.3234	0.7934	0.7569	0.8594	1.0663	1.0167	1.1504
	LSTM	0.8874	1.0287	0.3207	0.8013	0.8595	0.4365	0.9420	1.0142	0.5663
	GRU	0.2669	0.2323	0.2395	0.4190	0.3918	0.4963	0.5085	0.4637	0.5735
	Informer (np)	0.2264	0.1917*	0.2252	0.3100	0.2941	0.3112	0.4758	0.4378*	0.4746
	Informer	0.2039	0.2090	0.2264	0.2967	0.3091	0.3200	0.4515	0.4571	0.4758
	LDformer (np)	0.1759*	0.1957	**0.2184**	0.2690*	0.2885*	**0.3046**	0.4195*	0.4424	**0.4674**
	LDformer	**0.1603**	**0.1826**	0.2251*	**0.2617**	**0.2788**	0.3087*	**0.4004**	**0.4273**	0.4744*

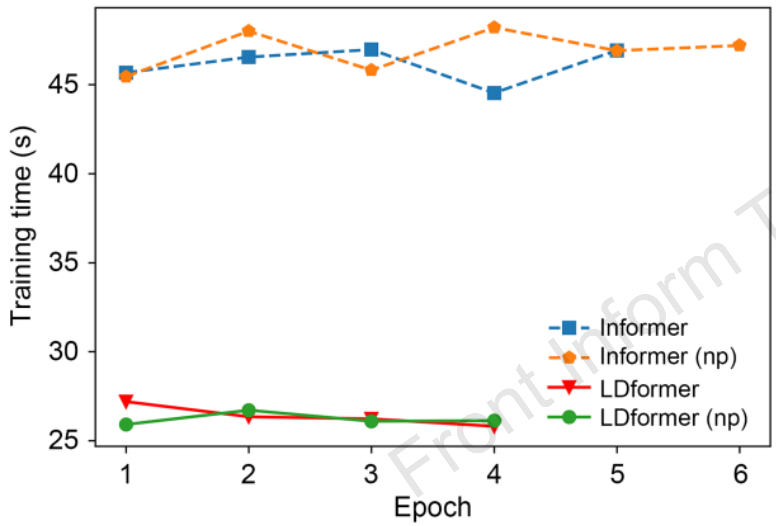
The best results are in bold, and * denotes the second-best results. MSE: mean square error; MAE: mean absolute error; RMSE: root mean square error

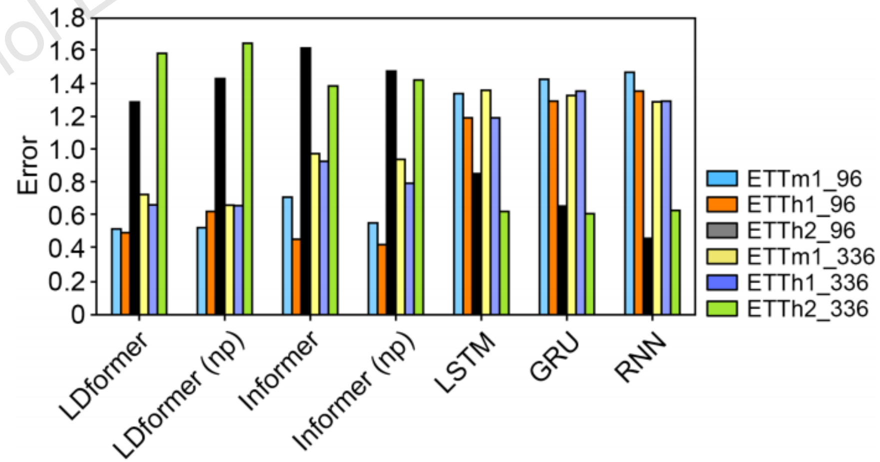
Front Inform Technol Electron Eng

Major results

- The average error between the true and prediction values for different datasets.







Front Inform Technol Electron Eng

- Under the early stop mechanism, both LDformer and LDformer (np) stop training at the fourth epoch, Informer stops training at the fifth epoch, and Informer (np) stops training at the sixth epoch. Informer (np) obtains the optimal result at the sixth epoch.

Conclusions

- Compared with traditional long-term power prediction models, LDformer and LDformer (np) can better capture long-term dependencies in long-time-series data without increasing computational complexity, thereby improving prediction accuracy.
- In long-time-series prediction tasks conducted on multiple short-term interval datasets, the LDformer model performs better than traditional long-term power prediction models.
- ProbSparse self-attention mechanism combined with UniDrop effectively avoids the risk of losing key connections between sequences without increasing complexity.

Front Inform Technol Electron Eng

Ran TIAN obtained a PhD in computer application technology from the School of Computer Science, Southwest Jiaotong University in 2015. Since July 2018, he has served as an associate professor at the College of Computer Science & Engineering in Northwest Normal University. His main research interests include digital supply chain, time-series prediction methods, and intelligent decision-making methods.

Xinmei LI received a Master's degree in computer science from Northwest Normal University in 2023. Her research direction is the prediction problem of power load time-series data.

Zhongyu MA obtained a PhD in Engineering from Northwest Polytechnical University in 2020, majoring in information and communication engineering. He serves as an associate professor at the College of Computer Science & Engineering in Northwest Normal University, mainly focusing on the theory and technology of B5G/6G wireless communication networks, and algorithms and protocols for wireless communication networks based on artificial intelligence.

Front Inform Technol Electron Eng

Yanxing LIU received a Master's degree in computer science from Northwest Normal University in 2017. His research interests include time-series data analysis, knowledge mining, application of sensor networks, Blockchain, and information systems.

Jingxia WANG received a Master's degree in computer science from Northwest Normal University in 2023. Her research direction is the prediction problem of time-series data of commodity prices.

Chu WANG obtained a Master's degree in computer science from Northwest Normal University in 2023. His research direction is the time-series prediction problem of traffic flow.

Front Inform Technol Electron Eng