Interpretable Sparse System Identification: Beyond Recent Deep Learning Techniques on Time-Series Prediction

Xiaoyi Liu[1], Duxin Chen[1*], Wenjia Wei[2], Xia Zhu[2] & Wenwu Yu[1]

1. School of Mathematics, Southeast University, Nanjing 210096, China 2. Network Laboratory, 2012 Lab, Huawei Technologies Co. Ltd, Shenzhen 518129, China

{xiaoyiliu,chendx,wwyu}@seu.edu.cn {weiwenjia,zhuxia1}@huawei.com *: Corresponding Author

Introduction

We propose the **Global-Local Identification and Prediction (GLIP)** model, which combines the framework of identification models with insights from deep learning training methods. The primary contributions can be summarized as follows:

- ✓ A novel identification model for long-term prediction is introduced, which **operates efficiently on CPUs** instead of relying on GPUs.
- ✓ The utilization of **global identification and local prediction** allows for precise capture of both the knowledge of long-term trends and short-term fluctuations in time series data.
- ✓ **Storage basis** enables accurate extraction of potential future trends even when the input-output ratio is very small.
- ✓ Remarkable prediction performance was achieved on the four benchmark datasets, surpassing the current **state-of-the-art results**. The proposed model exhibited a notable 23.64%.

Methods

In the stage of the entire prediction framework, we first construct **global basis** and **storage basis** in the training set for global prediction. Secondly, we evaluate the performance of global prediction in the validation set to determine its suitability for local rolling prediction. Finally, in the test set, the amalgamation of all preceding information will be leveraged to execute local rolling prediction with **local basis**. The Sparse Identification optimization formula is as follows:

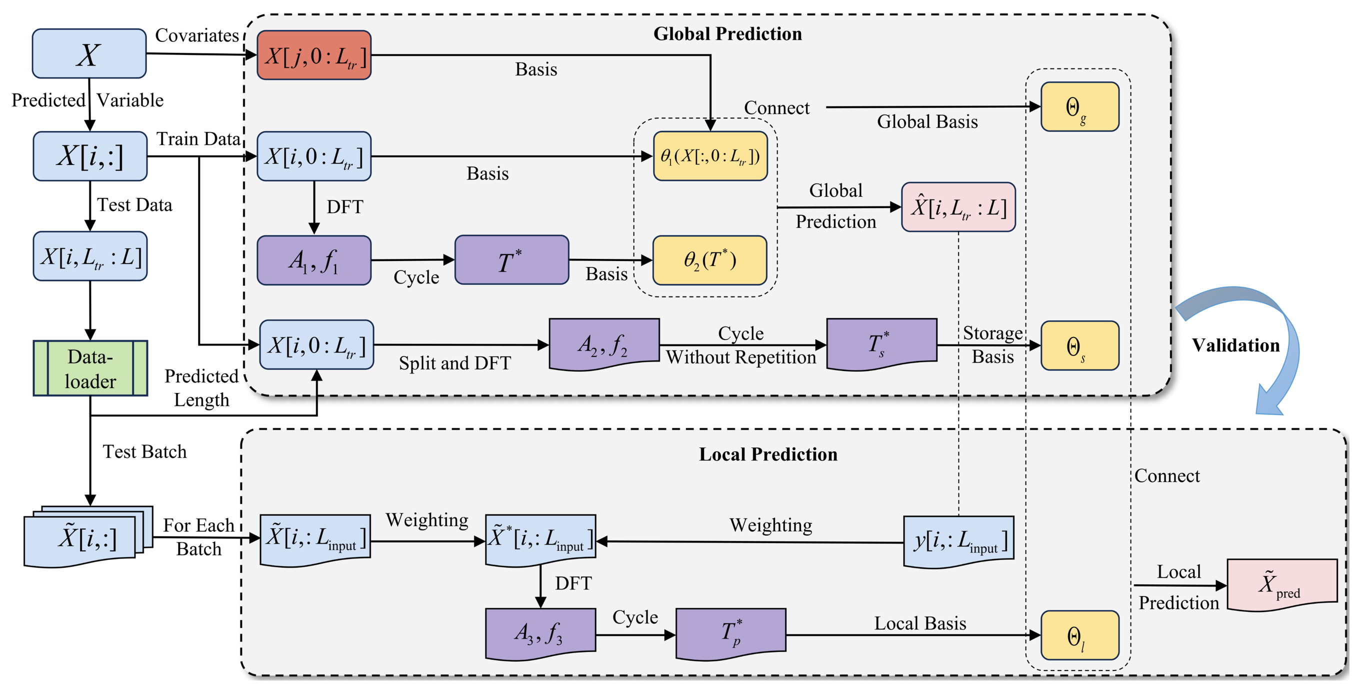
$$\arg\min_{\Xi} \int_0^T \left(\| X - \Theta \cdot \Xi \|^2 \right) \mathrm{d}t + \lambda \| \Xi \|_1$$

The functions of the three basis functions are as follows:

- ➢ **Global Basis:** This basis reflects the changes in the global time series, aiding our understanding of the overall trends.
- ➢ **Storage Basis:** This basis captures the local variations in the training set, providing ample information for the prediction.
- ➢ **Local Basis:** This basis assists in capturing the input data's basis for local rolling prediction scenarios in the test set, providing local information for localized predictions.

The detailed procedures of global prediction and local rolling prediction are illustrated in Figure 1.

Figure 1



The GLIP model architecture

Results

- **Local Rolling Prediction**

As shown in Table 1, GLIP achieved excellent results in local rolling prediction across the four benchmark datasets. Overall, compared to neural network methods, GLIP achieved a total improvement of 23.64%. We observed that the improvement tends to be greater when the O/I is larger. This indicates that GLIP has the ability to predict long-term data based on short-term data in local rolling prediction, addressing the challenge of long-term prediction that traditional methods cannot handle. Figure 2 (d)-(f) showcase the visualization results of the three datasets when performing local rolling prediction with the longest output.

- **Global Prediction**

Global prediction is a distinctive advantage of GLIP, which sets it apart from neural network-based methods that necessitate an ample number of training samples and are therefore incapable of accomplishing global prediction. This implies that neural networks may not consistently capture the genuine trends of time series from a macroscopic perspective. In certain scenarios, GLIP can provide a reference for global prediction, as depicted in Figure 2 (a)-(c).

- **Environment and Efficiency**

Each of the aforementioned experiments can be completed in a matter of seconds to minutes on a personal computer running on a CPU environment. This greatly reduces our computational requirements. According to the principle of Occam's razor, GLIP undoubtedly stands as a superior prediction model.

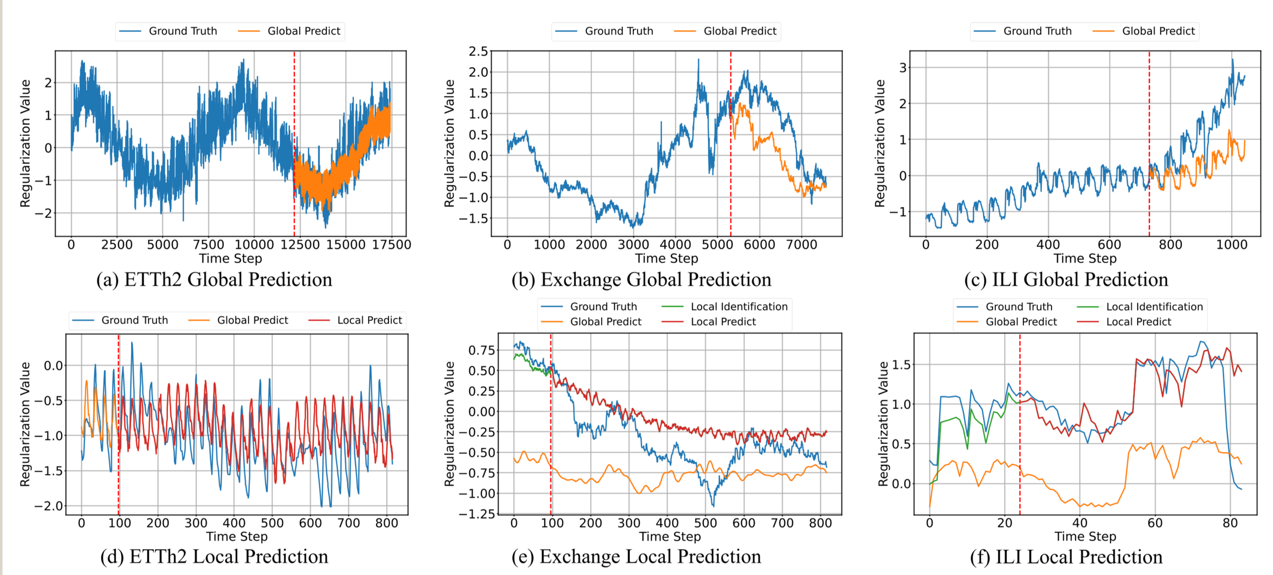
Table 1

Models		IMP.	GLIP		DLinear		FEDformer		Autoformer		Informer		Reformer		Pyraformer		LogTrans	
Metric		MSE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh	96	24.91%	**0.217**	**0.332**	0.289	0.353	0.346	0.388	0.358	0.397	3.755	1.525	0.845	0.693	0.645	0.597	2.116	1.197
	192	34.46%	**0.251**	**0.364**	0.383	0.418	0.429	0.439	0.456	0.452	5.602	1.931	0.958	0.741	0.788	0.683	4.315	1.635
	336	38.39%	**0.276**	**0.383**	0.448	0.465	0.496	0.487	0.482	0.486	4.721	1.835	1.044	0.787	0.907	0.747	1.124	1.604
	720	32.18%	**0.314**	**0.408**	0.605	0.551	0.463	0.474	0.515	0.511	3.647	1.625	1.458	0.987	0.963	0.783	3.188	1.540
ETTm	96	11.38%	**0.148**	**0.260**	0.167	**0.260**	0.203	0.287	0.255	0.339	0.365	0.453	0.658	0.619	0.435	0.507	0.768	0.642
	192	17.86%	**0.184**	**0.302**	0.224	0.303	0.269	0.328	0.281	0.340	0.533	0.563	1.078	0.827	0.730	0.673	0.989	0.757
	336	21.35%	**0.221**	**0.331**	0.281	0.342	0.325	0.366	0.339	0.372	1.363	0.887	1.549	0.972	1.201	0.845	1.334	0.872
	720	34.00%	**0.262**	**0.360**	0.397	0.421	0.421	0.415	0.433	0.432	3.379	1.338	2.631	1.242	3.625	1.451	3.048	1.328
Exchange	96	11.11%	**0.072**	**0.202**	0.081	0.203	0.148	0.278	0.197	0.323	0.847	0.752	1.065	0.829	0.376	1.105	0.968	0.812
	192	20.38%	**0.125**	**0.275**	0.157	0.293	0.271	0.380	0.300	0.369	1.204	0.895	1.188	0.906	1.748	1.151	1.040	0.851
	336	31.48%	**0.209**	**0.348**	0.305	0.414	0.460	0.500	0.509	0.524	1.672	1.036	1.357	0.976	1.874	1.172	1.659	1.081
	720	31.10%	**0.443**	**0.516**	0.643	0.601	1.195	0.841	1.447	0.941	2.478	1.310	1.510	1.016	1.943	1.206	1.941	1.127
ILI	24	18.65%	**1.802**	**0.908**	2.215	1.081	3.228	1.260	3.483	1.287	5.764	1.677	4.366	1.382	1.420	2.012	4.480	1.444
	36	9.52%	**1.766**	**0.943**	1.963	0.963	2.679	1.080	3.103	1.148	4.755	1.467	4.446	1.389	7.394	2.031	4.799	1.467
	48	18.73%	**1.731**	**0.952**	2.130	1.024	2.622	1.078	2.669	1.085	4.763	1.469	4.572	1.436	7.551	2.057	4.800	1.468
	60	22.80%	**1.828**	**0.992**	2.368	1.096	2.857	1.157	2.770	1.125	5.264	1.564	4.743	1.487	7.662	2.100	5.278	1.560

Local rolling prediction errors in terms of MSE and MAE.

Furthermore, considering the structure of GLIP, its efficiency in learning the local rolling prediction model is essentially independent of the prediction length. Once the model structure is determined, it can directly perform predictions without being affected by the increase in prediction length.

Figure 2



The visualization results depict the prediction of a specific variable in the ETT, Exchange, and ILI datasets. Figures~(a), (b), and (c) represent global predictions, while Figures~(d), (e), and (f) illustrate a slice of local rolling predictions. The labels in the figure hold the following implications: "Ground Truth" signifies the veritable time series data; "Global Prediction" denotes the visual representation of global prediction, aligning with the output outcomes from the global identification module; "Local Identification" represents the result of local identification after the weighted combination of global predictions and local input data; "Local Prediction" encompasses the outcomes of local rolling prediction; and the red vertical line demarcates the division between the training set and test set.

- **Ablation Experiment**

Effect of Sparse Identification: We refer to the absence of the sparse identification component as Case 1.

Effect of Local Identification Curve: The absence of local identification curves is denoted as Case 2.

Effect of Storage Basis: Case 3 is designated to represent the scenario where trend storage bases are absent.

Table 2

L_{output}	Metric	GLIP	Case 1	Case 2	Case 3
96	MSE	**0.072**	1.912	0.087	0.091
	MAE	**0.202**	1.151	0.212	0.220
192	MSE	**0.125**	1.186	0.206	0.232
	MAE	**0.275**	1.135	0.328	0.318
336	MSE	**0.209**	1.182	0.485	0.792
	MAE	**0.348**	1.122	0.496	0.453
720	MSE	**0.443**	1.759	0.602	2.450
	MAE	**0.516**	1.014	0.599	0.757

Results for ablation experiments on the Exchange dataset.

Conclusion

We have provided a detailed account of the GLIP model to global prediction and local rolling prediction. This breakthrough overcomes the limitations of traditional methods with a heavy computational burden and opens up a new direction for long-term prediction.
In future work, we plan to enhance both the scale of the data and the scope of the predictions. Additionally, we aim to integrate the identification methods with neural network approaches.