西湖大學
WESTLAKE UNIVERSITY

CIKM2019 BEIJING CHINA
November 03-07. 2019

DSANet: Dual Self-Attention Network for Multivariate Time Series Forecasting

Siteng Huang, Donglin Wang, Xuehan Wu, Ao Tang

Presenter: Siteng Huang

Machine Intelligence Laboratory, Department of Engineering,
Westlake University, Hangzhou, China

November, 2019

Outline

1. Introduction and Previous Works
2. Proposed Model
3. Experiments
4. Conclusion

Introduction

- The purpose of time series forecasting is to predict the future value based on historical data.
- The difficulty lies in that traditional methods fail to capture complicated non-linear dependencies between time steps and between multiple time series.

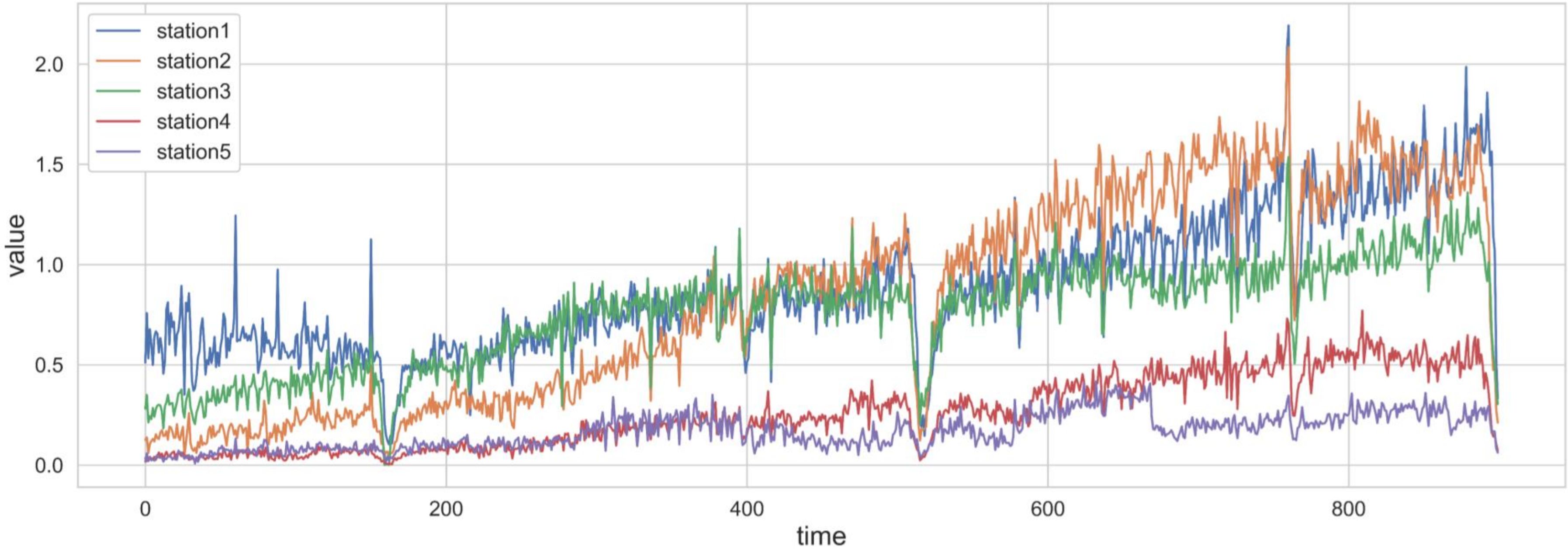

Figure 1: An example of chaotic multivariable time series.

Previous Works

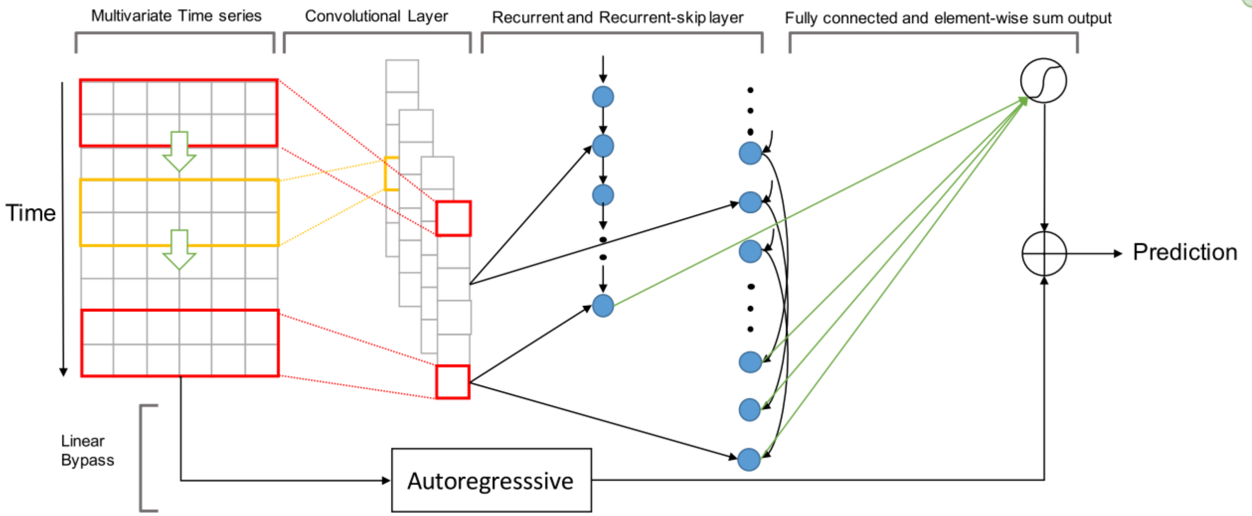

Figure 2: Long- and Short-term Time-series Network (LSTNet).

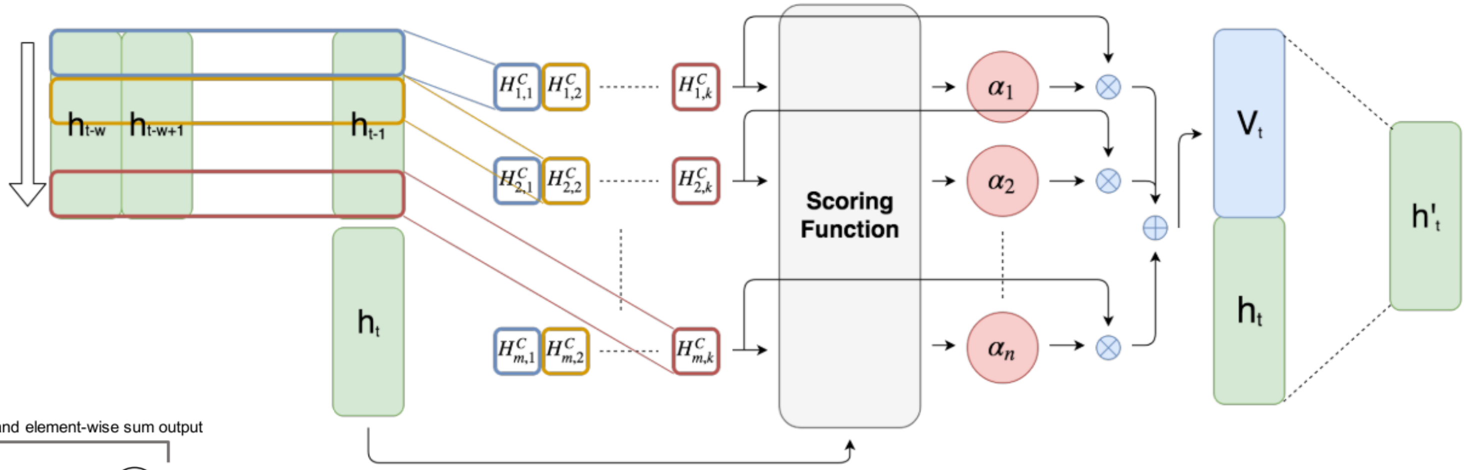

Figure 3: Temporal Pattern Attention.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, Hanxiao Liu. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. SIGIR 2018: 95-104
Shun-Yao Shih, Fan-Keng Sun, Hung-yi Lee. Temporal pattern attention for multivariate time series forecasting. Machine Learning 108(8-9): 1421-1441 (2019)

Proposed Model

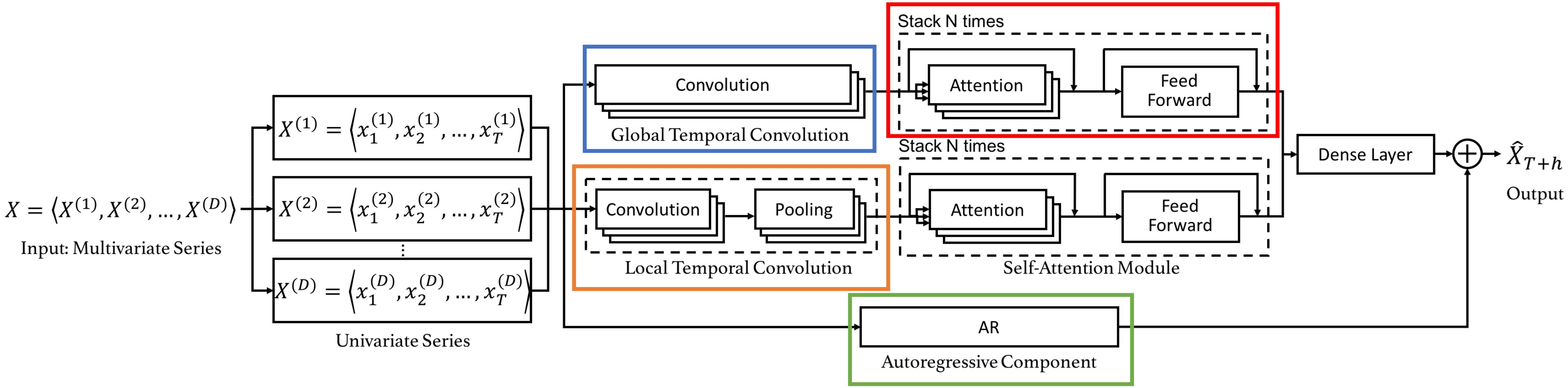

Figure 4: Dual Self-Attention Network (DSANet).

Experimental Settings

- Dataset: A large multivariate time series dataset, which contains the daily revenue of geographically close gas stations.
- Baselines: VAR, LRidge, LSVR, GRU, LSTNet-S, LSTNet-A, TPA
- Problem Parameters:
  - *window*
    - The length of the input time series
    - Value range: {32, 64, 128}
  - *horizon*
    - The desirable horizon ahead of the current time stamp
    - Value range: {3, 6, 12, 24}

西湖大學 WESTLAKE UNIVERSITY

Experimental Settings

- Evaluation Metrics:
  - Root relative squared error (RRSE)

$$\mathrm{RRSE} = \frac{\sqrt{\sum_{(i,t)}(Y_{i,t} - \hat{Y}_{i,t})^2}}{\sqrt{\sum_{(i,t)}(Y_{i,t} - \mathrm{mean}(Y))^2}}$$

  - Mean absolute error (MAE)

$$\mathrm{MAE} = \mathrm{mean}(\sum_{(i,t)} |Y_{i,t} - \hat{Y}_{i,t}|)$$

  - Empirical correlation coefficient (CORR)

$$\mathrm{CORR} = \mathrm{mean}(\sum_{i} \frac{\sum_t (Y_{i,t} - \mathrm{mean}(Y_i))(\hat{Y}_{i,t} - \mathrm{mean}(\hat{Y}_i))}{\sqrt{\sum_t (Y_{i,t} - \mathrm{mean}(Y_i))^2 \sum_t (\hat{Y}_{i,t} - \mathrm{mean}(\hat{Y}_i))^2}})$$

Experimental Results

		Methods							
window-horizon	Metrics	VAR	LRidge	LSVR	GRU	LSTNet-S	LSTNet-A	TPA	DSANet
32-3	RRSE	0.9401	0.8114	0.8934	0.8297	0.8222	0.8120	0.8441	**0.7817**
	MAE	0.4914	0.4302	0.4687	0.4311	0.4214	0.4220	0.4348	**0.4074**
	CORR	0.3600	0.6294	0.4575	0.5850	0.5953	0.6193	0.5592	**0.6416**
32-6	RRSE	0.9170	0.8094	0.9144	0.8524	0.8544	0.8718	0.8482	**0.7713**
	MAE	0.4743	0.4374	0.4778	0.4380	0.4371	0.4465	0.4336	**0.4102**
	CORR	0.3973	0.6477	0.4283	0.5587	0.5822	0.5678	0.5577	**0.6602**
32-12	RRSE	0.9335	0.9132	0.9600	0.8938	0.8753	0.9033	0.8887	**0.8297**
	MAE	0.4746	0.4619	0.4956	0.4536	0.4524	0.4529	0.4487	**0.4367**
	CORR	0.3824	0.5887	0.3657	0.5102	0.5878	0.5069	0.5011	**0.6057**
32-24	RRSE	1.0188	0.9789	1.0178	0.9457	0.9941	0.9814	0.9310	**0.9277**
	MAE	0.4988	0.4811	0.5174	0.4807	0.4916	0.4921	0.4499	**0.4422**
	CORR	0.2709	0.5141	0.3237	0.4708	0.4513	0.4535	0.3951	**0.5464**

Table 1: RRSE, MAE and CORR scores for our proposed DSANet and baselines when *window*=32.

Experimental Results

		Methods							
window-horizon	Metrics	VAR	LRidge	LSVR	GRU	LSTNet-S	LSTNet-A	TPA	DSANet
64-3	RRSE	0.9732	0.8948	0.9121	0.9348	0.8710	0.8703	0.9302	**0.8446**
	MAE	0.4712	0.4490	0.4455	0.4626	0.4226	0.4314	0.4537	**0.4214**
	CORR	0.3339	0.5034	0.4470	0.4066	0.5209	0.5041	0.3899	**0.5444**
64-6	RRSE	0.9926	0.8887	0.9081	0.9208	0.8715	0.8738	0.9160	**0.8211**
	MAE	0.4839	0.4574	0.4565	0.4553	0.4382	0.4427	0.4531	**0.4189**
	CORR	0.3008	0.5290	0.4608	0.4176	0.5241	0.5309	0.4298	**0.5554**
64-12	RRSE	1.0578	1.0193	0.9690	0.9661	0.9511	0.9659	0.9622	**0.8953**
	MAE	0.4898	0.4812	0.4766	0.4555	0.4520	0.4705	0.4519	**0.4459**
	CORR	0.2462	0.4311	0.3869	0.3146	0.3844	0.4255	0.3094	**0.4480**
64-24	RRSE	1.1505	1.0707	1.0330	0.9885	1.0099	1.0065	0.9984	**0.9329**
	MAE	0.5403	0.5122	0.5087	0.4716	0.4886	0.4845	0.4763	**0.4497**
	CORR	0.1422	0.3657	0.3067	0.2595	0.2426	0.3282	0.2688	**0.4204**

Table 2: RRSE, MAE and CORR scores for our proposed DSANet and baselines when *window*=64.

西湖大學 WESTLAKE UNIVERSITY

Experimental Results

		Methods							
window-horizon	Metrics	VAR	LRidge	LSVR	GRU	LSTNet-S	LSTNet-A	TPA	DSANet
128-3	RRSE	1.0707	0.9369	0.9573	1.0035	0.9232	0.9262	1.0039	**0.8730**
	MAE	0.4982	0.4335	0.4750	0.4732	0.4399	0.4412	0.4768	**0.4175**
	CORR	0.1718	0.3181	0.3444	0.0722	0.3652	0.3678	0.2176	**0.4227**
128-6	RRSE	1.1419	0.9098	0.9694	0.9971	0.9490	0.9137	0.9993	**0.8438**
	MAE	0.5358	0.4295	0.4892	0.4685	0.4530	0.4378	0.4658	**0.4095**
	CORR	0.0766	0.3675	0.3566	0.1534	0.3230	0.4137	0.1899	**0.4147**
128-12	RRSE	1.1893	0.9465	1.0366	1.0014	0.9281	0.9341	0.9994	**0.8802**
	MAE	0.5758	0.4453	0.5319	0.4667	0.4507	0.4514	0.4744	**0.4314**
	CORR	0.0669	0.3297	0.3071	0.1180	0.3739	0.4043	0.1433	**0.4288**
128-24	RRSE	1.2814	0.9929	1.1782	1.0239	0.9583	0.9471	1.0059	**0.8987**
	MAE	0.6263	0.4613	0.6268	0.4708	0.4466	0.4418	0.4633	**0.4144**
	CORR	-0.0329	0.2478	0.2090	-0.0114	0.2480	0.3328	-0.0219	**0.4060**

Table 3: RRSE, MAE and CORR scores for our proposed DSANet and baselines when *window*=128.

Ablation Study

- DSAwoGlobal: Remove the global temporal convolution branch;
- DSAwoLocal: Remove the local temporal convolution branch;
- DSAwoAR: Remove the autoregressive component.

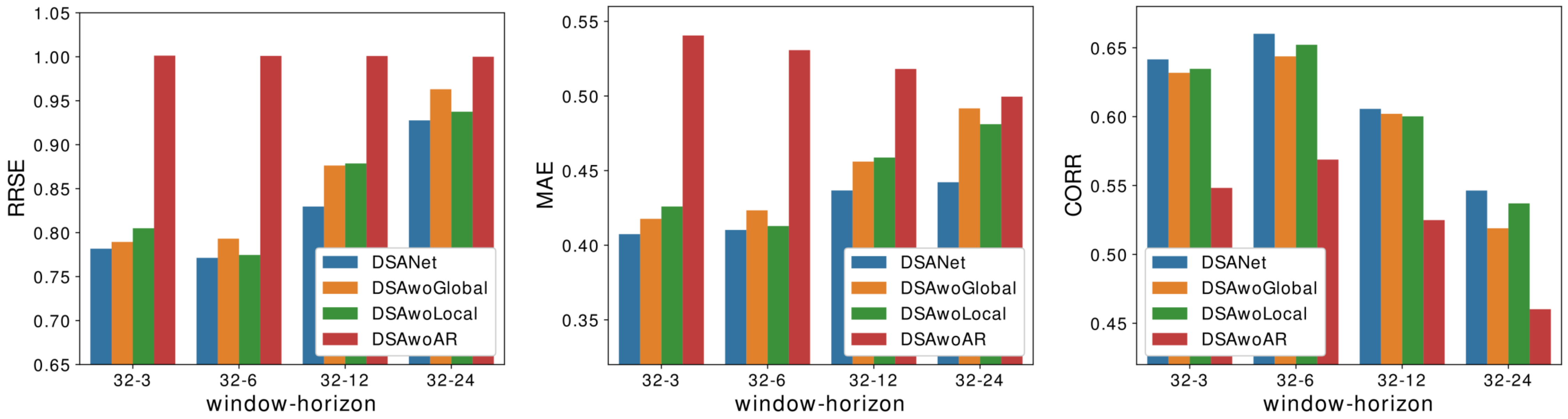

Figure 5: Ablation test results of DSANet.

西湖大學 WESTLAKE UNIVERSITY

Conclusion

- Multivariate time series with dynamic-period or nonperiodic patterns is chaotic and hard to forecast.
- Dual convolutions help to capture mixtures of global and local temporal patterns.
- Self-attention mechanism helps to capture the dependencies between different series.
- Our model shows promising results and outperforms baselines.
- All components have contributed to the effectiveness and robustness of the whole model.

Thanks For Attention

Question?