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

Siteng Huang[1], Donglin Wang[1*], Xuehan Wu[2], Ao Tang[3]

[1] Machine Intelligence Laboratory, Westlake University; [2] Huawei Technologies Co., Ltd.; [3] WeCar Technology Co., Ltd.
{huangsiteng, wangdonglin}@westlake.edu.cn

Introduction

Challenge

The difficulty of the task lies in that traditional time series forecasting methods fail to capture complicated non-linear dependencies between time steps and between multiple time series. Recently, recurrent neural network and attention mechanism have been used to model periodic temporal patterns across multiple time steps. However, these models fit not well for time series with dynamic-period patterns or nonperiodic patterns.

Method

We propose a **D**ual **S**elf-**A**ttention **Net**work (DSANet) for highly efficient multivariate time series forecasting, especially for dynamic-period or nonperiodic series. DSANet completely dispenses with recurrence and utilizes two parallel convolutional components, a self-attention module and a traditional autoregressive linear model.

Contributions

- We adopt global and local temporal convolutions to capture complex mixtures of global and local temporal patterns.
- We incorporate self-attention modules to extract dependencies between different time series. To the best of our knowledge, this is the first work to apply self-attention mechanism in time series forecasting.
- Experiments on real-world multivariate time series data show that the proposed model is effective and outperforms baselines.

Model

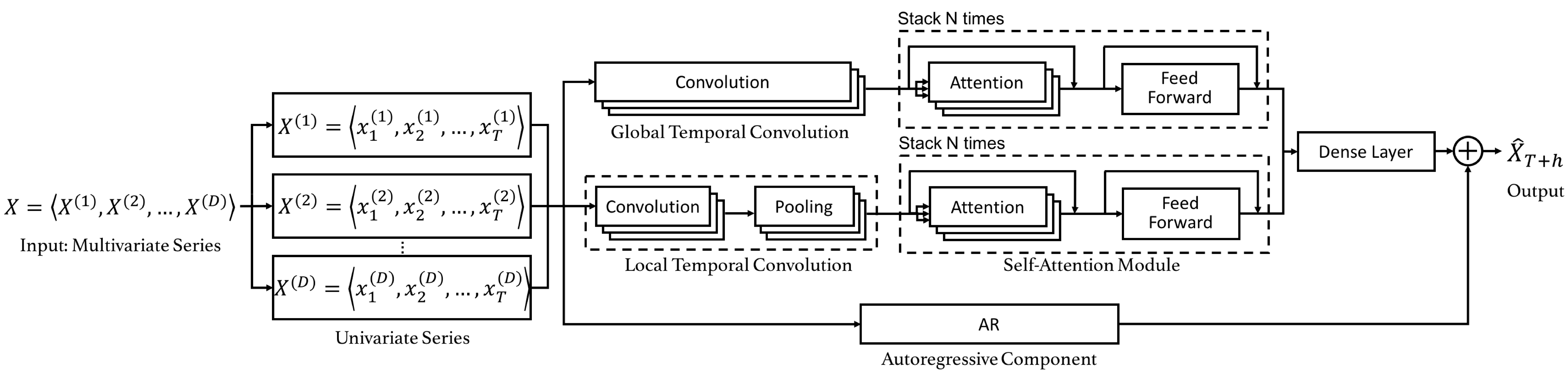

Figure 1: Dual Self-Attention Network (DSANet).

Global Temporal Convolution

DSANet applies 1D convolution over all time steps to extract global temporal patterns for univariate time series.

Local Temporal Convolution

Considering that time steps with a shorter relative distance have a larger impact on each other, DSANet uses another 1D convolution with shorter length of filters to model local temporal patterns.

Self-Attention Module

A self-attention module is applied to capture the dependencies between different series.

Autoregressive Component

To address the drawback that the scale of neural network output is not sensitive to that of input, the final prediction of DSANet is a mixture of the non-linear component and a classical autoregressive model.

Experiment

We conduct experiments on a large multivariate time series data set, which contains the daily revenue of geographically close gas stations. Data visualization analysis is performed to ensure that the data set does not contain distinct repetitive patterns.

Main Results

We use root relative squared error (RRSE), mean absolute error (MAE) as evaluation metrics, for which a lower value is better.

		Methods								
window-horizon	Metrics	VAR	LRidge	LSVR	GP	GRU	LSTNet-S	LSTNet-A	TPA	DSANet
32-3	RRSE	0.9401	0.8114	0.8934	1.0564	0.8297	0.8222	0.8120	0.8441	**0.7817**
	MAE	0.4914	0.4302	0.4687	0.5676	0.4311	0.4214	0.4220	0.4348	**0.4074**
32-6	RRSE	0.9170	0.8094	0.9144	1.0677	0.8524	0.8544	0.8718	0.8482	**0.7713**
	MAE	0.4743	0.4374	0.4778	0.5616	0.4380	0.4371	0.4465	0.4336	**0.4102**
32-12	RRSE	0.9335	0.9132	0.9600	1.0878	0.8938	0.8753	0.9033	0.8887	**0.8297**
	MAE	0.4746	0.4619	0.4956	0.5580	0.4536	0.4524	0.4529	0.4487	**0.4367**
32-24	RRSE	1.0188	0.9789	1.0178	1.1280	0.9457	0.9941	0.9814	0.9310	**0.9277**
	MAE	0.4988	0.4811	0.5174	0.5611	0.4807	0.4916	0.4921	0.4499	**0.4422**

Table 1: RRSE and MAE scores for our proposed model and baselines.

Ablation Study

To justify the efficiency of our architecture design, a careful ablation study is conducted. Specifically, we remove each of the global temporal convolution branch, the local temporal convolution branch, and the AR component one at a time in our DSANet model, and each new model is named DSAwoGlobal, DSAwoLocal, and DSAwoAR.

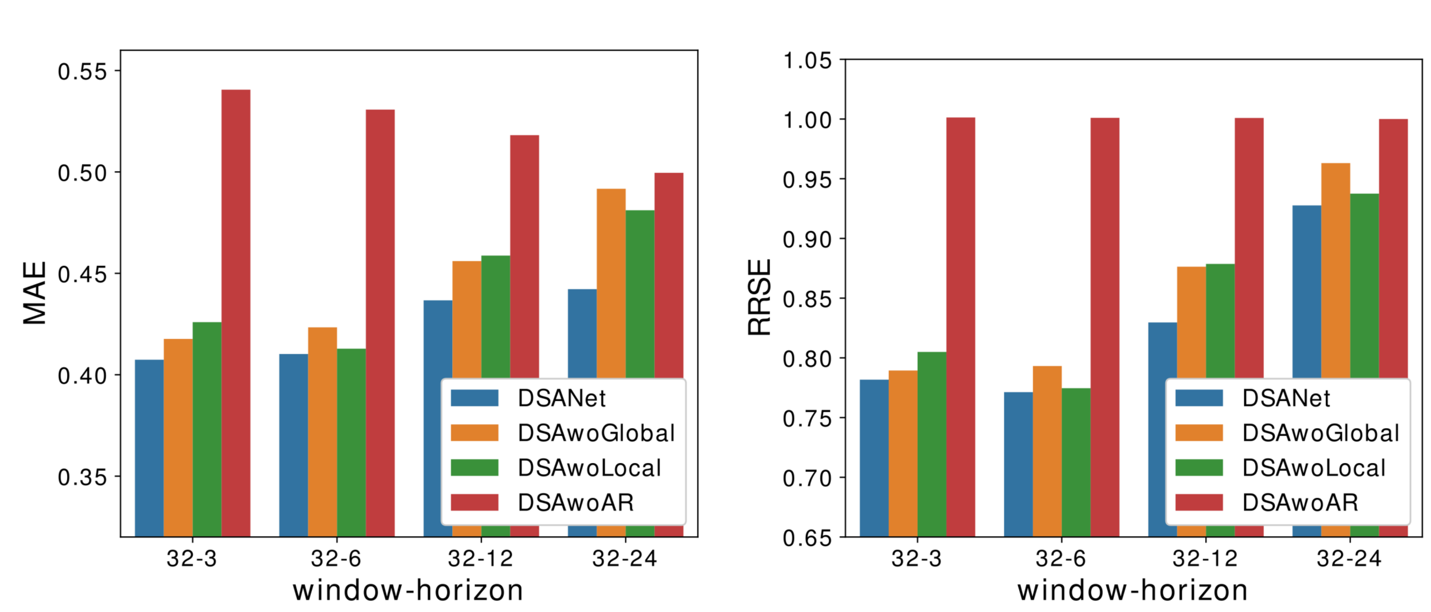

Figure 2: Ablation test results of DSANet.

Conclusion

We present a novel deep learning model, dual self-attention network (DSANet), for the task of multivariate time series forecasting, especially for those data with dynamic-period or nonperiodic patterns. Experiments on a large real-world dataset demonstrate the accuracy and robustness of the proposed method.

Misc

References

1. Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. 2018. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In *Proc. of SIGIR 2018*.
2. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proc. of NIPS 2017*. 5998–6008.

Links

- code: github.com/bighuang624/DSANet
- website: https://kyonhuang.top/publication/dual-self-attention-network

西湖大學
WESTLAKE UNIVERSITY

CIKM2019 BEIJING CHINA
November 03-07, 2019