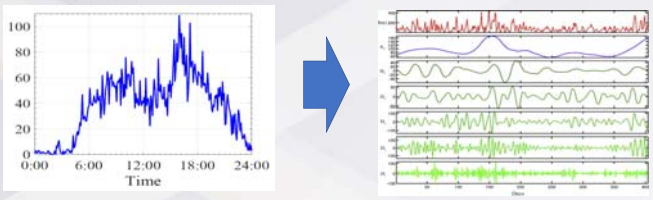
Multilevel Wavelet Decomposition Network for Interpretable Time Series Analysis

Jingyuan Wang, Ze Wang, Jianfeng Li, Junjie Wu Beihang University, Beijing, China

Motivation

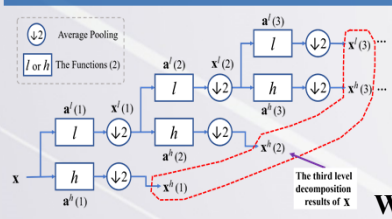


- Frequency is an import feature of time series data.
- Wavelet is a powerful tool to analyze frequency.
- Deep learning achieved great success in time series applications.

Our idea:

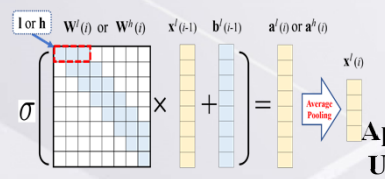
Integrating wavelet and deep neural network in a whole framework

Multilevel Wavelet Decomposition Network



$$a_n^l(i+1) = \sum_{k=1}^{K} x_{n+k-1}^l(i) \cdot l_k,$$
$$a_n^h(i+1) = \sum_{k=1}^{K} x_{n+k-1}^l(i) \cdot h_k,$$

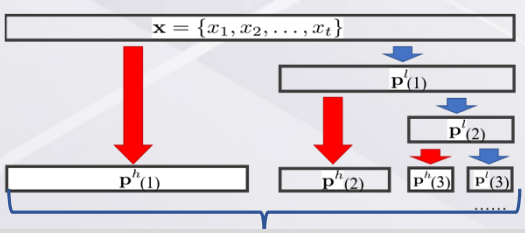
Wavelet Decomposition



$$a^l(i) = \sigma\left(W^l(i)x^l(i-1) + b^l(i)\right),$$
$$a^h(i) = \sigma\left(W^h(i)x^l(i-1) + b^h(i)\right).$$

Approximating Wavelet Using Neural Networks

In the similar form

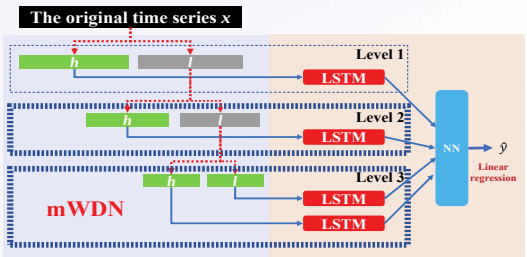


Multilevel Wavelet Decomposition Network

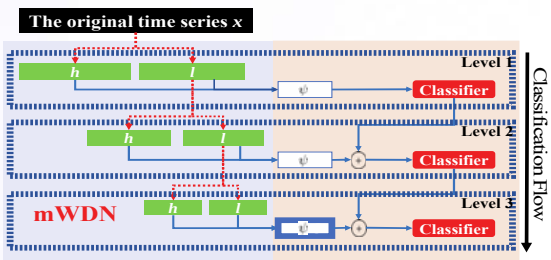
Approximated Wavelet Frequency Component

Applications

Based on the mWDN model, we further extend the model to two variants for time series forecasting and time series classification.



Multi-frequency LSTM for forecasting



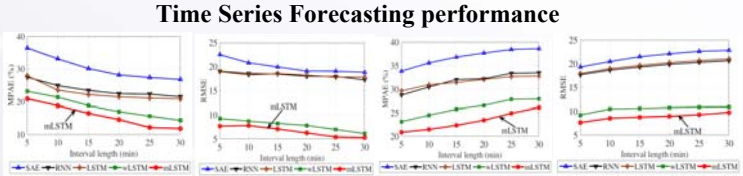
Residual Classification Flow for classification

Performance

Time Series Classification over the UCR data set

Err Rate	RNN	LSTM	MLP	FCN	ResNet	MLP-RCF	FCN-RCF	ResNet-RCF	Wavelet-RCF
Winning times	2	2	0	9	6	2	**19**	7	7
AVG arithmetic ranking	7.425	6.825	7.2	4.025	4.55	5.15	**2.175**	3.375	*3.075*
AVG geometric ranking	6.860	6.131	7.043	3.101	3.818	4.675	**1.789**	2.868	*2.688*
MPCE	0.039	0.043	0.041	0.023	0.025	0.028	**0.017**	0.021	*0.019*

Time Series Forecasting performance



The mWDN based deep neural networks achieved the best classification and forecasting performance compared with other benchmarks.

Interpretability

We **quantify the importance** of each middle layer to the final output of the mWDN based models by estimating the **partial derivatives** of final prediction to intermediate sub-sequence generated by Multilevel Wavelet Decomposition Network.

$$S(x_i) = \left|\frac{\partial M(x_i)}{\partial x_i}\right| = \left|\lim_{\varepsilon \to 0} \frac{M(x_i) - M(x_i - \varepsilon)}{\varepsilon}\right|$$

Sensitivity of inputs and each frequency