WaveletAE: A Wavelet-enhanced Autoencoder for Wind Turbine Blade Icing Detection

Binhang Yuan-SIAM(2019)

COTENTS

- Introduction
- WaveletAE Architechure
- Experimental Setup

COTENTS

SCADA System

SCADA is an acronym for *supervisory control and data acquisition*, a computer system for gathering and analyzing real time data. SCADA systems are used to monitor and control a plant or equipment in industries such as telecommunications, water and waste control, energy, oil and gas refining and transportation.

Wind speed

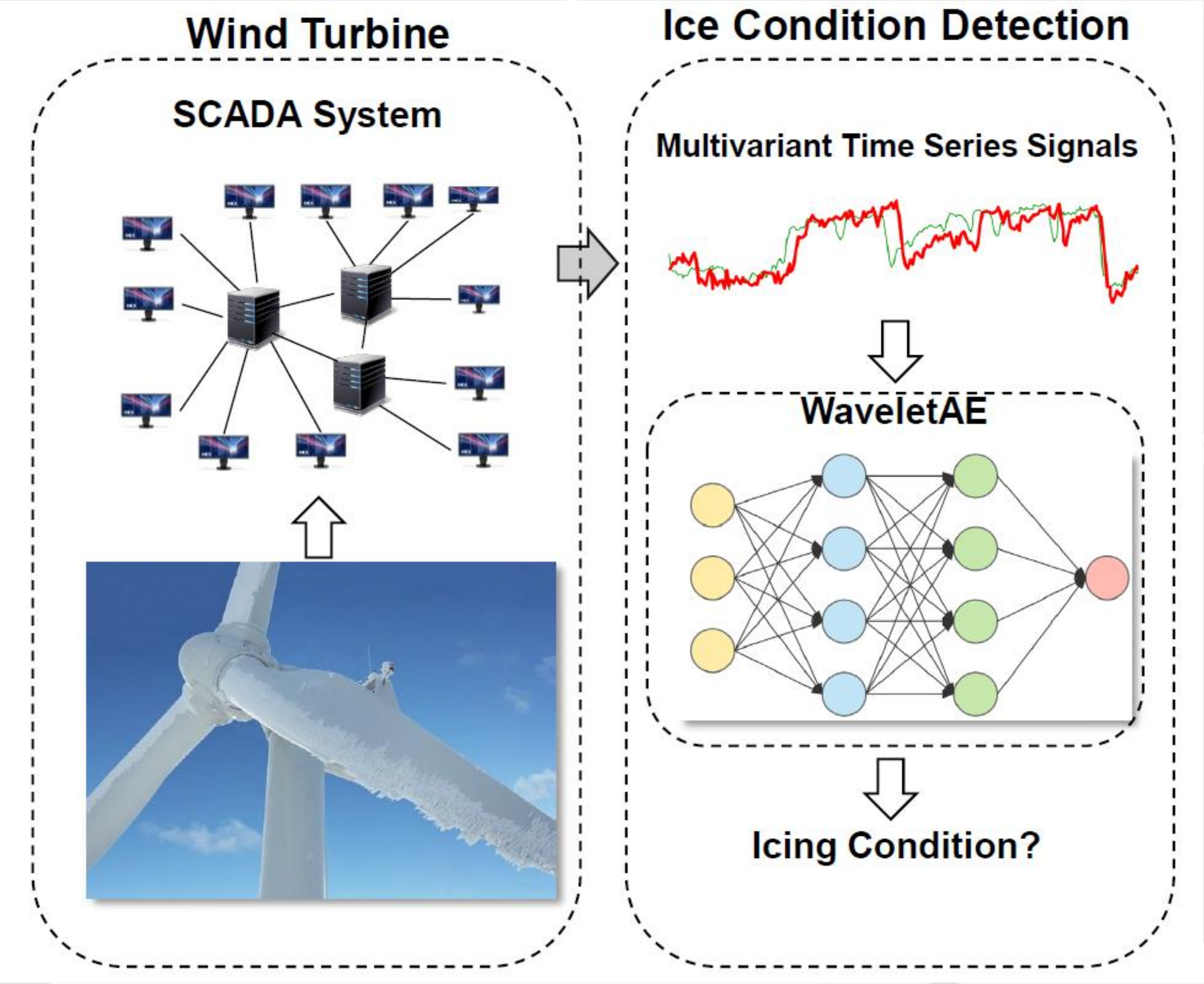
Internal temperature

Yaw positions

Pitch angles

Power outputs

Workflow


Wind Turbine
SCADA System
Ice Condition Detection
Multivariant Time Series Signals
WaveletAE
Icing Condition?

Discrete Wavelet Transform

Formally, given $x = [\ x_0 \quad x_1 \quad \dots \quad x_{N-1}\ ]^T$ that represents a length N signal, and the basis functions of the form $\varphi = [\ \varphi_0 \quad \varphi_1 \quad \dots \quad \varphi_{N-1}\ ]^T$ and $\psi = [\ \psi_0 \quad \psi_1 \quad \dots \quad \psi_{N-1}\ ]^T$, then the coecients for each translation (indexed by k) in each scale level (indexed by j0 or j) are projections of the signal onto each of the basis functions:

$$\boldsymbol{w}_\varphi[j_0,k] = (\ \boldsymbol{x}, \quad \varphi_{j_0,k}\ ) = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \boldsymbol{x}[m]\,\varphi_{j_0,k}[m]$$

$$\boldsymbol{w}_\psi[j,k] = (\ \boldsymbol{x}, \quad \psi_{j,k}\ ) = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \boldsymbol{x}[m]\,\psi_{j,k}[m]$$

where $\boldsymbol{w}_\varphi[j_0,k]$ is called approximation coecient, and $\boldsymbol{w}_\psi[j,k]$ is called detail coecient.

Discrete Wavelet Transform

The detail coecients at dffierent levels reveal variances of the signal on different scales, while the approximation coecient yields the smoothed average on that scale. One important property of the discrete wavelet transform is that detail coecients at each level are orthogonal, that says for any pair of detail coecients notin the same level, the inner product is 0:

$$\boldsymbol{w}_{\psi}\left[j, *\right] \cdot \boldsymbol{w}_{\psi}\left[j', *\right] = 0$$

Deep Autoencoder

A deep autoencoder is a multi-layer neural network, in which the desired output is the original input. Internally, the autoencoder includes a hidden layer h that describes a low-dimensional code to represent the input x. The network consists of two parts: an encoder function $h = f_E(x)$ parameterized by E that maps the input signal x to a code and a decoder function $x = f_D(h)$ parameterized by D that produces a reconstruction x of the input. Intuitively, learning a low dimensional representation forces the autoencoder to capture the most salient features of the training data.

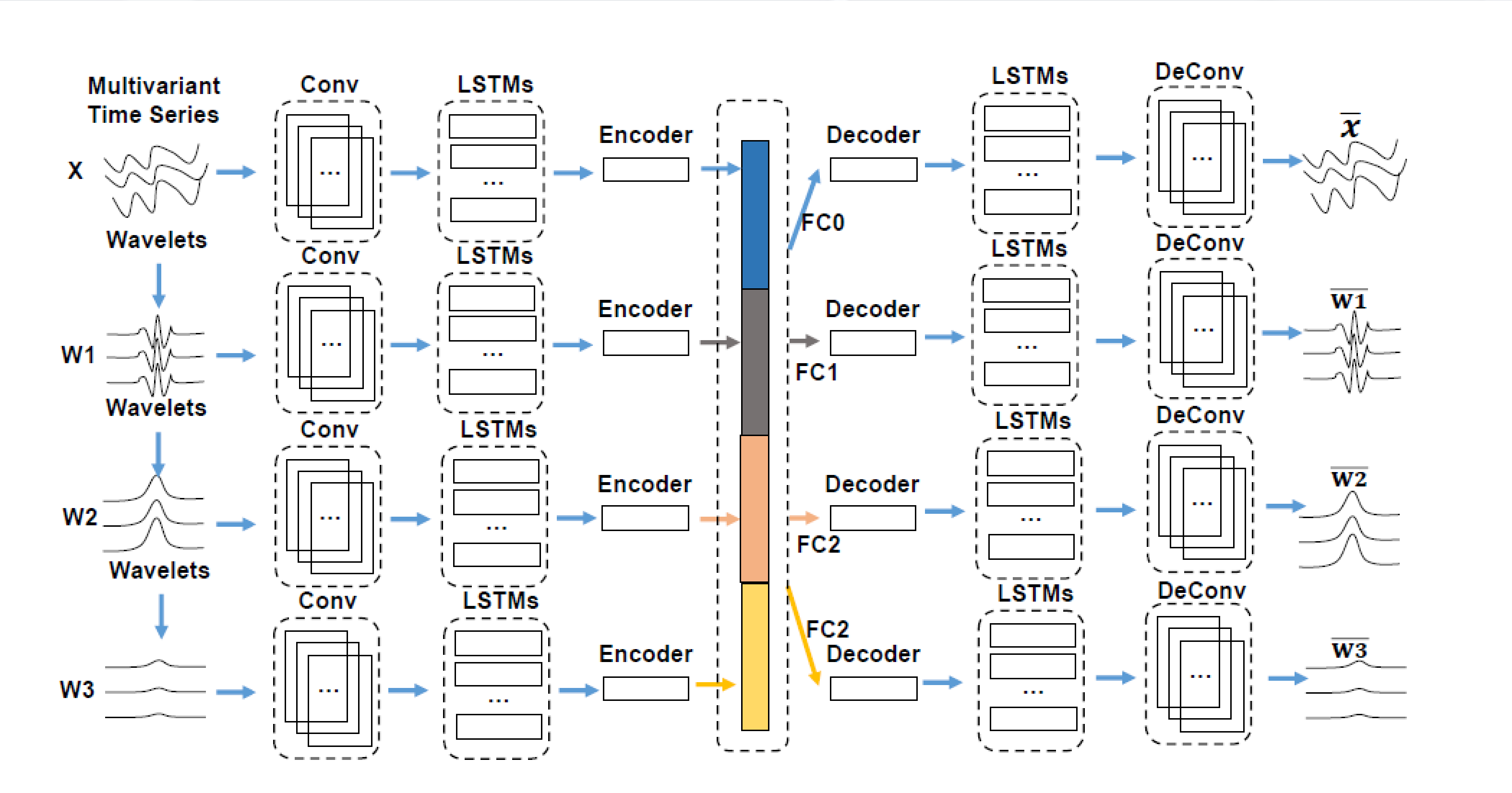
$$\min_{E,D} loss\left(\boldsymbol{x},\ f_D\left(f_E\left(\boldsymbol{x}\right)\right)\right)$$

Problem Formalization

Given the multivariate time series x R^{CxT} , where C is the number of channels, and T is the length of the signal, denoted equivalently by

- $\boldsymbol{x} = [\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(T)}]$ where $\boldsymbol{x}^{(t)} \in R^C$ represents the C-dimensional vector of variables at time t, and
- $\boldsymbol{x} = [\boldsymbol{x}_{(0)}, \boldsymbol{x}_{(1)}, ..., \boldsymbol{x}_{(C)}]^T$ $\boxed{2}$ where $\boldsymbol{x}_{(c)} \in R^T$ represents the T-dimensional vector of signal from channel c.

WaveletAE Architecture



Multilevel Discrete Wavelet Decomposition

According to the classic pyramid algorithm, we first apply the 1-dimensional discrete wavelet transform on each input signal channel $x_{(c)}$ to compute the wavelet detail coeffcients in each channel to a speciffic level L, where L can be viewed as a hyper-parameter of WaveletAE. The input signal is first augmented with the wavelet detail coeffcients to multiple scales. Formally,in scale level l, the wavelet details coeffcients are noted as $\boldsymbol{w}[l] = \left[\boldsymbol{w}_{(0)}[l], \boldsymbol{w}_{(1)}[l], ..., \boldsymbol{w}_{(C)}[l]\right]^T$

Convolutional Encoder

$$\boldsymbol{a_{out}} = f\left(W * \boldsymbol{a_{in}} + b\right)$$

Multiple Scale LSTM Encoder-Decoder

$$
\begin{aligned}
\boldsymbol{i}^{(t)} &= \sigma\left(W_{ii}\boldsymbol{a}^{(t)} + b_{ii} + W_{hi}\boldsymbol{h}^{(t-1)} + b_{hi}\right) \\
\boldsymbol{f}^{(t)} &= \sigma\left(W_{if}\boldsymbol{a}^{(t)} + b_{if} + W_{hf}\boldsymbol{h}^{(t-1)} + b_{hf}\right) \\
\boldsymbol{g}^{(t)} &= \tanh\left(W_{ig}\boldsymbol{a}^{(t)} + b_{ig} + W_{hg}\boldsymbol{h}^{(t-1)} + b_{hg}\right) \\
\boldsymbol{o}^{(t)} &= \sigma\left(W_{io}\boldsymbol{a}^{(t)} + b_{io} + W_{ho}\boldsymbol{h}^{(t-1)} + b_{ho}\right) \\
\boldsymbol{c}^{(t)} &= \boldsymbol{f}^{(t)} \cdot \boldsymbol{c}^{(t-1)} + \boldsymbol{i}^{(t)} \cdot \boldsymbol{g}^{(t)} \\
\boldsymbol{h}^{(t)} &= \boldsymbol{o}^{(t)} \cdot \tanh\left(\boldsymbol{c}^{(t)}\right)
\end{aligned}
$$

Deep Autoencoder

$$a_{out} = f\left(W \circledast a_{in} + b\right)$$

Reconstruct Loss

- the approximation coeffcients represent some smoothed averages of the input signal, such knowledge should be easily learned by the convolutional layers when processing the original signal.
- unlike the detail coeffcients,approximation coeffcients are not orthogonal to each other, the redundancy of the input can provide limited helpful information while enlarge the parameter space of the model.

$$loss = \|x - \bar{x}\|_2 + \sum_{l=1}^{L} \|w[l] - \bar{w}[l]\|_2$$

SCADA Dataset

- 60% of the signal as the training set;
- 20% of the signal as the validation set to compare the generalization performance between WaveletAE and the state-of-the-art approaches;
- 20% of the signal as the nal test data in the case study of simulated deployment.

Metrics

- Accuracy is the number of correct predictions made by the model over all the predictions.
- Precision is a measure that tells us what proportion of positions that we diagnose as an anomaly, actually are anomalies.
- Recall measures what proportion of samples that are anomalies is diagnosed by the model as an anomaly.
- F1 score is the Harmonic mean of precision and recall as a general evaluation of the model.

$$recall = \frac{true\ positives}{true\ positives\ +false\ negatives} \qquad precision = \frac{true\ positives}{true\ positives + false\ positives}$$

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

Semi-Supervised Anomaly Detection

Semi-supervised anomaly detection is required to construct models representing normal pattern from a given training data set which only include normal samples, and then output the likelihood of whether a test instanceis normal or abnormal.

$$\tau=\beta\cdot\text{mean}\{loss_{train}\}$$

Measurement	WaveletAE	LSTM [16]
Accuracy	**0.493**	0.428
Precision	**0.471**	0.437
Recall	**0.986**	0.928
F1 score	**0.637**	0.594

Supervised Anomaly Detection

In order to leverage the label information under supervised setting, we add a fully connected layer to map the global hidden states to the probability p_a of the abnormality, the training loss becomes a affine combination of reconstruction loss and the crossentropy binary classication loss $loss_c = -[y \log(p_a) + (1 - y) \log(1 - p_a)]$ dened as:

$$Loss = \alpha \cdot loss_{re} + (1-\alpha) \cdot loss_c$$

Measurement	WaveletAE	LSTM [16]	FCNN [29]
Accuracy	**0.873**	0.792	0.747
Precision	**0.857**	0.804	**0.857**
Recall	**0.863**	0.712	0.525
F1 score	**0.861**	0.755	0.651