Time Series Data Augmentation for Neural Networks by Time Warping with a Discriminative Teacher

Brian Kenji Iwana and Seiichi Uchida
Kyushu University

Summary

- We proposed a new data augmentation method for time series called **guided warping**. Guided warping warps the elements of one time series to the steps of another time series using Dynamic Time Warping (DTW).
- We demonstrated that the use of a discriminator can improve the performance of guided warping by selecting the **most discriminate teacher**.
- We evaluated the proposed method on all 88 datasets from the 2015 UCR Time Series Archive [1] using a temporal CNN and an LSTM.

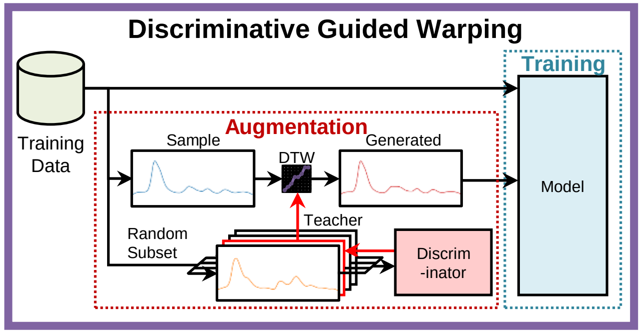
Background

- Time series classification datasets are often **small** and this isn't ideal for neural networks.
- Data augmentation is a common method to increase generalization.
- Common time series data augmentation methods include: random transformations, generative models, pattern mixing, and signal decomposition methods.

References

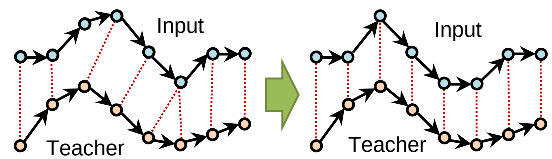
[1] Chen et al., the UCR Time Series Classification Archive, https://www.cs.ucr.edu/~eamonn/time_series_data/, 2015
[2] Um et al., Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks, ACM ICMI, 2017
[3] Le Guennec et al., Data augmentation for time series classification using convolutional neural networks, IWAATD, 2016
[4] Kamycki et al., Data augmentation with suboptimal warping for time-series classification, Sensors, 2019
[5] Forestier et al., Generating synthetic time series to augment sparse datasets, IEEE ICDM, 2017

Method



Guided Warping

- The **elements of the input sample** are warped to the **time steps of the teacher sample** using DTW.
- The generated pattern has the features of the input at the time steps of the teacher sample.



Discriminator

- To select the **most discriminative teacher**, a **nearest centroid classifier** is used, or:

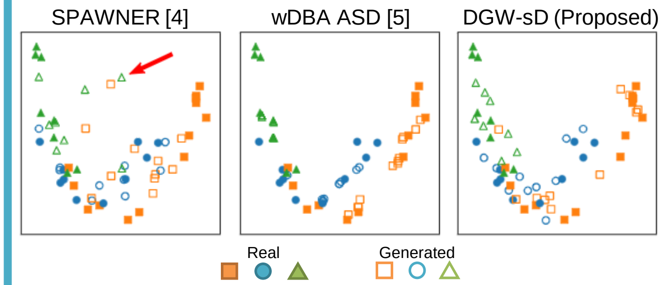
$$h(\mathbf{b}_m) = \frac{1}{\sum_{m'}[l_{m'} \neq l_m]} \sum_{m'} \mathcal{D}(\mathbf{b}_{m'}, \mathbf{b}_m)|[l_{m'} \neq l_m] - \frac{1}{\sum_{m'}[l_{m'} = l_m]} \sum_{m'} \mathcal{D}(\mathbf{b}_{m'}, \mathbf{b}_m)|[l_{m'} \neq l_m]$$

$$\mathbf{b}_{\text{disc}} = \underset{\{m=1,\dots,M\}}{\operatorname{argmax}}\ h(\mathbf{b}_m)$$

$\mathbf{b}_m$ is a time series in bootstrap set $\mathbf{B}$
$\mathcal{D}(\mathbf{b}_{m'}, \mathbf{b}_m)$ is the DTW distance between $\mathbf{b}_{m'}$ and $\mathbf{b}_m$
l_m is the label of pattern $\mathbf{b}_m$

Results

Visualization of the UCR CBF Dataset Using MDS



Results on 2015 UCR Time Series Archive Datasets

Method	CNN	LSTM
No Augmentation	76.44	57.24
Jittering [2]	77.32	58.35
Rotation [2]	74.84	54.78
Scaling [2]	77.06	57.98
Magnitude Warping [2]	78.30	58.04
Time Warping [2]	78.10	52.80
Window Slicing [3]	79.15	54.49
Window Warping [2]	79.58	57.49
SPAWNER [4]	78.84	**58.98**
wDBA [5]	77.42	56.01
RGW-D (Proposed)	79.39	57.42
RGW-sD (Proposed)	79.27	56.43
DGW-D (Proposed)	80.12	56.01
DGW-sD (Proposed)	**80.17**	56.99

RGW-D: Random Guided Warping with DTW
RGW-sD: Random Guided Warping with shapeDTW
DGW-D: Discriminative Guided Warping with DTW
DGW-sD: Discriminative Guided Warping with shapeDTW