Surrogate Loss Learning for DTW

Abstract

In Time Series analysis Dynamic Time Warping (DTW) is a technique used to quantify similarity between two time sequences which might be varying in length and speed. Research has manifested that DTW, which is a metric can also be used as a loss function for training deep learning networks. The basic implementation of DTW makes use of 'min' operation which is non-differentiable. In order to use DTW as loss function, authors proposed **soft-DTW** which is differentiable as it makes use of soft-min instead of min operation. The pitfall of this approach is that its closely dependent on hyper-parameter γ and focuses only on shape error between two sequences. The efficiency of this function is hindered by its high dependency on parameter γ and is a major stumbling block for it.

Considering these factors, surrogate loss functions turns out to be best alternative for non-differential or computationally complex loss functions. As suggested by the authors in **Learning Surrogate Losses**, we would try to develop a surrogate loss function for **DTW** as meta-level neural network that will approximate the desired loss. This network will be trained initially with train data-set such that **DTW** is differentiable during training and then the training of prediction model will be carried out. We will evaluate the model on the basis of error and epoch-time. In perspective of DTW we'd like to make use of GANs for surrogate loss training[3].

References

[1] Cuturi, Marco, and Mathieu BlondelCuturi, Marco, and Mathieu Blondel. Soft-DTW: a differentiable loss function for time-series, International Conference on Machine Learning. PMLR, 2017.

[2] Grabocka, Josif, Randolf Scholz, and Lars Schmidt-Thieme. Learning surrogate losses, arXiv:1905.10108 (2019).

[3] Lamb Alex, et al. Professor Forcing: A new algorithm for training neural networks, arXiv:1610.09038 (2016).