Smart Transfer Learning methods for time series classification

Pratikshit Singh
ps71@illinois.edu

Simran Sandhu
sandhu5@illinois.edu

Siddharth Patra
patra3@illinois.edu

Abstract

The objective of this research project is to investigate the domain of transfer learning in Time Series Classification (TSC) problems. The focus is on exploring methods to analyze the best source dataset for pre-training CNN models for transfer learning to improve accuracy for the targeted time series dataset. To achieve this, the project aims to leverage various similarity measures between time series datasets. The experiments will be performed using the UCR archive, the largest publicly available TSC benchmark containing 85 datasets. The project will evaluate the performance of similarity measures such as DTW and a one dimensional Fully Convolutional Neural Network (FCN) will be used as the network architecture. The project will compare the results of the similarity measures and select the best measure for the final combination of datasets to be used for transfer learning capabilities for the CNN model. The final phase of the project will include cross-validation testing to evaluate the performance of the proposed method and explain the results.

1 Introduction

Through this research project we aim to explore the transfer learning problem domain, particularly domain adaptation for Time Series classification(TSC) problems [1]. Transferring deep CNN for TSC seems to deliver high accuracy results by fine tuning a pre-trained model. The model prediction's improvements or degradations depend on the choice of source dataset chosen for the transfer learning approach. Thus, we aim to find methods to analyze the best source of time series dataset for pre-training CNN models that can be transferred to improve accuracy for the targeted time series dataset. We plan to explore and leverage various similarity measures between time series datasets to improve model predictions.

2 Motivation

Generalization with transfer learning has shown significant accuracy improvements in classification tasks[2]. Deep CNN image classification transfer models like ResNets [3] and others trained on large dataset - ImageNet dataset [4], containing 100 classes, gave substantial accuracy improvements. Transfer learning is easier in computer vision tasks because images have a high level of similarity in terms of the underlying structures, patterns, and features between different visual classes when a pre-trained model have seen a distribution from 1000 classes.

However, time series data is highly dependent on the underlying trends, patterns, and dependencies over time. These patterns can vary significantly across different time series classification tasks, making it more challenging to transfer the knowledge learned from one task to another. Hence, it is more important to choose an optimum source dataset which can be used for pre-training deep CNN model. Thus, transfer learning and similarity measures with time series has many unexplored fields and motivated us to research in this area.

3 Dataset

We perform experiments using the UCR archive [5] which is the largest publicly available TSC benchmark containing 85 datasets. For each dataset in the archive, pre-training a model and then fine-tuned it on the other datasets would result in 7140 different deep neural networks. This is computationally expensive, so in the project we will perform analysis, using similarity measures et al., before training models to find some interesting combination of source and target datasets.

4 Methodology

This section talks about in detail the experiments that we plan to perform as well as their performance evaluation. Any post-analysis that we do, will be decided based on the results of the performance metrics.

4.1 Network Architecture

Network architecture which we plan to use is a one dimensional Fully Convolutional Neural Network (FCN). The performance of similarity measures for transfer learning methods is not dependent on the model chosen, however we choose this since it is a proven state-of-art model for UCR archive time series classification [6].

4.2 Similarity Measures

We do a literature survey in related work such as [7] and find some similarity measures that may give a performance benefit with transfer learning approaches. We would then perform a critical post-analysis in order to account for any discrepancies observed from intuitive behaviour. We've identified below potential candidates for measuring the similarity between datasets.

- Dynamic Time Warping(DTW)[8] with different distance measures
- Proxy A-Distance (PAD)[9]
- Pearson Correlation Coefficient
- (DBA)[10]
- Time lagged cross correlation

4.3 Evaluation

We plan to compare the similarity measure's performance by comparing their effects on the model's loss. We plan to perform analysis on the different techniques and reason about the perofrmance differences.

5 Anticipated Challenges

Limited computational resources - due to large number of possible models by combination of source and target(7140 unique models) we would not be exploring all the possible pre-training(source dataset) and fine turning(target dataset) combination on CNN models.

Implementation of selected similarity measure between all 85 datasets in UCR archive, problems may occur due to different lengths of time series.

6 Plan to address or overcome the challenges

We plan to find similarly and dissimilarity measures between datasets before actually generating pre-trained transfer CNN models and fine tuning them to get accuracy analysis. We'll only preceed with model implementation for datasets that give interestingly high or lower similarity, or datasets which have large variance in various similarity measures.

7 Desired outcomes and goals

We will use smart transfer learning to generate deep CNN models to avoid negative transfer learning. Exploring various similarity measure and then transferring pre-trained models from source to target will give us correlation between similarity and accuracy improvement by transfer learning for TSC.

Another goal is to justify some anomalies observed in accuracy results in paper[1] using DTW algorithm as an similarity measure. We aim to explore more similarity measure that justify such differences.

8 Future Work

UCR archive has recently extended to include 43 new datasets. As future work, this project can be extended to generate results on all 128 datasets, and generate analysis on interesting observations.

In another future work, we can aim to reduce the deep neural network's overfitting phenomena by generating synthetic data using a Weighted DTW Barycenter Averaging method [10], since the latter distance gave encouraging results in guiding a complex deep learning tool such as transfer learning.

9 Schedule

As initial work of the project we will get the UCR archive datasets, followed by preparation and analysis of the datasets, identifying various classes and outliers, removal of outliers, and setting up first similarity measure algorithms. Further using similarity results between datasets we'll select the datasets we want to use for our domain adaptation task. Only these datasets will be used for creating transfer models.

Around mid term checkpoint we will design and implement an optimal CNN model that we will use for transfer learning. Following this, we'll be implementing more similarity measures[add more similarity measures references]. We will showcase the different similarity results for different combination of datasets from the subset of UCR dataset. And show analysis on accuracy improvements by transfer learning between datasets vs similarity between datasets.

As final phase of project we plan to compare the results of the similarity measures and generalize/choose the best measure to create the final combination of dataset that we will use for transfer learning capabilities for our CNN model.Finally we'll pre-train CNN models using these source dataset and fine-tune on target dataset. We'll perform cross-validation testing to evaluate the performance of whole machinery and explain the results. We also plan to justify the anomalies observed for most similar neighbour in section 2 while not performing te best in accuracy improvement on transfer learning.

References

[1] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Transfer learning for time series classification. In *2018 IEEE international conference on big data (Big Data)*, pages 1367–1376. IEEE, 2018.

[2] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data mining and knowledge discovery*, 33(4):917–963, 2019.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[5] Dau, Hoang Anh and Keogh, Eamonn and Kamgar, Kaveh and Yeh, Chin-Chia Michael and Zhu, Yan and Gharghabi, Shaghayegh and Ratanamahatana, Chotirat Ann and Yanping and Hu, Bing and Begum, Nurjahan and Bagnall, Anthony and Mueen, Abdullah and Batista,

Gustavo, and Hexagon-ML. The ucr time series classification archive, October 2018. `https://www.cs.ucr.edu/~eamonn/time_series_data_2018/`.

[6] Zhiguang Wang, Weizhong Yan, and Tim Oates. Time series classification from scratch with deep neural networks: A strong baseline, 2016.

[7] Joan Serra and Josep Ll Arcos. An empirical evaluation of similarity measures for time series classification. *Knowledge-Based Systems*, 67:305–314, 2014.

[8] *Dynamic Time Warping*, pages 69–84. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[9] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems*, volume 19. MIT Press, 2006.

[10] François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3):678–693, March 2011.