Reproducibility Report for ACM SIGMOD 2022 Paper: "dCAM: Dimension-wise Class Activation Map for Explaining Multivariate Data Series Classification"

PARIS CARBONE, ADAM HASSELBERG, KTH Royal Institute of Sweden, Sweden

The paper is proposing a new convolutional architecture for discriminant feature discovery and classification explanation in multivariate time series analysis. The main experiment workload which involves heavy ML training could not be fully completed during the evaluation period due to time and resource constraints. The training script stalls in CPU-based execution and require access to a HPC cluster comparable to what the authors used. Using a HPC cluster we are able to reproduce and validate most of the authors results.

1 INTRODUCTION

This reproducibility report concerns the paper dCAM: Dimension-wise Class Activation Map for Explaining Multivariate Data Series Classification [1], that is joint work between Université Paris Cité and EDF R&D (France). The paper features extensive model comparisons between the proposed model (dCNN/dResNet/dInceptionTime) and its baselines. The training execution could not be reproduced in full due to time and resource constraints. However, the authors results were reproduced to a convincing degree, making the overall contributions credible.

2 SUBMISSION

The reproducibility submission [2] contained source code and scripts necessary to run the result. However, the datasets had to be downloaded manually from Google Drive.

The project was based on pytorch and a set of common python libraries which could be trivially installed. All experiment scripts were included in the directory which further allowed parameterization for either "CPU" or "CUDA" -based execution, and support-scripts to facilitate deployment of batch-jobs to HPC clusters. Finally, a notebook file was included for plotting all results as well as additional descriptions and instructions to reproduce the experiments.

3 HARDWARE AND SOFTWARE ENVIRONMENT

Table 1 summarizes the hardware specifications of the original paper as well as those used in the reproduction of the results.

Table 1. Hardware & Software environment

	Paper	Repro Review
HPC Cluster	Jean Zay (idris.fr)	Tetralith (nsc.liu.se)
CPU	Intel Cascade Lake 6248	Intel Xeon Gold 6130
cores	2x20	32
GHz	2.5	2.1
RAM	384 GB	96 GB
GPU	Nvidia Tesla V100 SXM2 32GB	Nvidia Tesla T4 16GB

4 REPRODUCIBILITY EVALUATION

4.1 Process

Setup: The preparation phase was good but not ideal. Google Drive files cannot be downloaded e.g., using `wget` from our remote server, thus, all data files had to be manually downloaded and transferred to the project direct directory.
Experiments: The experiments could not be reproduced using only cpu-mode, the execution stalled with no information about why. After communicating with the authors we were advised to use a cluster with GPU accelerators. Running the experiments on the Tetralith cluster required minor reconfiguration for some of the batch jobs: extending the predefined maximum allowed run time, and, reducing the maximum batch size. Due to time and resource constraints we were unable to complete all training batch jobs that were part of the experiments. In total 293 out of 332 training tasks were reproduced successfully.
Evaluation: The reproduced results could easily be gathered and loaded using the provided python notebook and we were able to produce side-by-side plots of our own results and the authors pre-compiled results. We find that after reproducing the authors experiments we achieved similar results with minor differences. Figure 1 shows an example of the minor differences we found in reproducing the experiments.

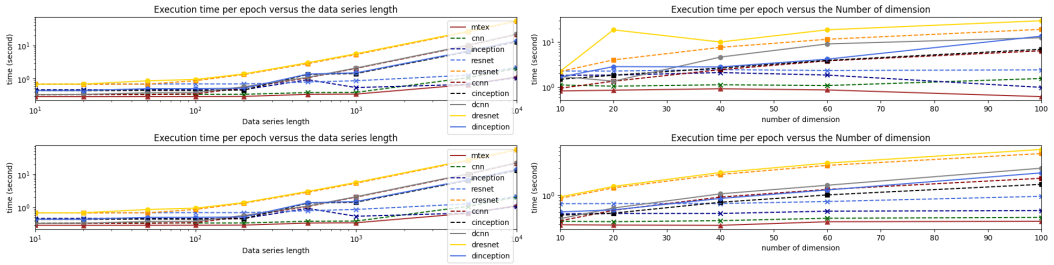

Fig. 1. Reproduced results (top) and original results (bottom).

4.2 Results

The results were consistent with the paper for the given experiment output and all figures and tables could be validated. Figures 8-12 could be adequately validated. A minor remark for Table 2 to be made is that within 10 runs overall accuracies seemed slightly lower (e.g., 5-10%), however, the trends were identical to those presented in the paper.

5 SUMMARY

Overall most paper claims have been partially reproduced, we were unable to reproduce 39 out of the 332 required training tasks due to time and resource constraints. We find that the results could be sufficiently reproduced to certify a reproducibility badge.

REFERENCES

[1] Boniol, Paul, Mohammed Meftah, Emmanuel Remy, and Themis Palpanas. "dcam: Dimension-wise class activation map for explaining multivariate data series classification." In Proceedings of the 2022 International Conference on Management of Data, pp. 1175-1189. 2022.
[2] Reproducibility Repository (last access 1 Feb 2023) https://github.com/boniolp/dCAM/tree/main/src