Response to the reviewers

ICLR

Dataset transformations trade-offs to adapt machine learning methods across domains

Reviewer GQQh

Reviewer Comments P 0.1: *The authors find that different representations and models give different performance, which is not surprising. This is a known fact - in practice, domain specific models are generally trained tailored to the task at hand. These models encode implicit biases needed for the task. For vision tasks, convnets are trained and for NLP, transformers are used. Researchers are aware of this, and the conclusions of this paper add no value in my opinion. Furthermore, the experiments are performed only on MNIST with minimal architectural choices, and the results are not conclusive in any form. I recommend strongly rejecting the paper. The conclusions of the paper add no value, and the experiments are not performed well.*

Reply: We agree with the reviewer that different representations and models give different performance. However, we believe that it is worth exploring the advantages and disadvantages of using different models. As it is well known, a convolutional neural network (CNN) has higher accuracy than a fully connected network (FCN). At the same time, as we proved in the paper a CNN is more susceptible to adversarial attacks than a FCN.

Reviewer 57HV

Reviewer Comments P 0.2: *The draft fails to support the claims strongly. For instance, they do not consider data from CPS, instead they consider well known datasets from computer vision on which various DNN architectures have already been explored. Instead, authors should have worked with a different domain as claimed. Also, the datasets MNIST and Fashion MNIST are similar in some of the important aspects: both gray scale, same number of classes, and number of samples (complexity), etc. This does not support the draft's claim that it works with diverse dataset types. Experimental analysis presented in the paper is very weak. Only a couple simplest datasets are considered to train simple CNN and MLP classifier architectures. This kind of analysis is not very convincing to generalize from. It is not very clear why authors consider adversarial robustness to investigate the disadvantage of the dataset conversion. The take-away (according to the abstract) is that the data conversion is not always needed or beneficial. This is not very informative or novel, since it is understandable that the data formats need to be understood and domain knowledge also is required to avoid any structure or information loss while processing,*

Reply: In the paper, we select the model accuracy as a proxy for advantages and adversarial robustness as a proxy for disadvantages. The reason is to explore the tradeoffs between different models. For example, a convolutional neural network (CNN) has higher accuracy than a fully connected network (FCN). However, as we proved in the paper, a CNN is more susceptible to adversarial attacks than a FCN. To strengthen this claim, we added more results showing how the distances between samples in the dataset change for different models.

Reviewer JxT2

Reviewer Comments P 0.3: *In general, the authors clearly deliver their ideas and experimental procedure. While people are introducing ML methods to more and more real-world problems, the topic covered in this paper is certainly becoming more attractive. The problem is not well formulated. The idea of studying different input formates could be helpful when solving practical problems, but I would suggest the authors formulate this problem by starting from even simple NN models and assumptions on datasets. Is it possible to use an objective to describe this discrepancy caused by data transformations? Lack experiments. The authors are inspired by ML problems on CPSs datasets, however, the only experiment in this paper is on the MNIST and Fashion-MNSIT datasets, and the experiment on these part (MNIST, fashionMNIST) are already provided by public codebases [1]. Experiments on datasets from power systems, robotics, or autonomous driving are expected. There are multiple public datasets from CPSs mentioned above [2–4]. The definition of "domain" and "adapt". This paper defines the domain as a broad area of applications such as CV, audio recognition, and others, and the term "adapt" means changing the data formate. However, in many existing studies, the term domain is usually associated with the distribution of data [5]. Consider an autonomous driving case where we collect two datasets (image and vehicle trajectory) for each weather (sunny and raining) and thus have four datasets in total, how should we divide these four datasets into different domains? The usage of the Optimal Transport Data Distance. While this is a promising prior study on dataset geometric distances, one assumption is that the label-feature distributions are Gaussian. Does that still hold for the CPSs dataset that may contain continuous signals? I would like to hear more from the authors on this point. It is a good idea to use adversarial attacks to evaluate the transformations. It would be interesting to utilize the "Wasserstein adversarial attack" [6] and see if this method correlates with the optimal transport data distance.*

Reply: We appreciate your helpful feedback. To strengthen our claims, we included new simulations in Figs. 4 and 5. In specific, we show that while a convolutional neural network (CNN) has higher accuracy than a fully connected network (FCN), the CNN is more susceptible to adversarial attacks than a FCN.

References

[1] D. Alvarez-Melis, *Optimal Transport Dataset Distance (OTDD)*. [Online]. Available: https://github.com/microsoft/otdd.

[2] *PJM Hourly Energy Consumption Data*. [Online]. Available: https://www.kaggle.com/robikscube/hourly-energy-consumption.

[3] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[4] A. Mandlekar *et al.*, "Roboturk: A crowdsourcing platform for robotic skill learning through imitation," in *Conference on Robot Learning*, PMLR, 2018, pp. 879–893.

[5] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Optimal transport for domain adaptation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 9, pp. 1853–1865, 2016.

[6] E. Wong, F. Schmidt, and Z. Kolter, "Wasserstein adversarial examples via projected sinkhorn iterations," in *International Conference on Machine Learning*, PMLR, 2019, pp. 6808–6817.