An Implementation of "Adversarial Discriminative Domain Adaptation"

Ali Jafari, Behrouz Ghamkhar

March 30, 2023

1 Introduction

While methods based on deep learning have achieved state-of-the-art results on numerous tasks, most of these methods are based on the assumption that the data used to train the model is drawn from the same distribution that the data used to test the model is drawn from. This assumption can be violated in the real world, where AI models should be able to adapt themselves to new environments with little human supervision. In order to address this problem, methods based on *single-source unsupervised domain adaptation* utilize labeled data from a *source domain* to achieve satisfactory performance on unlabeled data from a different, but related *target domain* [1].

In this document, we discuss our implementation of "Adversarial Discriminative Domain Adaptation" [2]. In addition to documenting our implementation of the paper, we provide the results of our replications of the experiments performed in [2]. Furthermore, we comment on the result of each replicated experiment, and offer possible reasons for the success or failure of every experiment.

1.1 Overview of the Method

In [2], the authors propose a framework for single-source unsupervised domain adaptation that is trained in three stages. Firstly, a *source encoder* and a *classifier* are jointly trained on data from the source domain. Secondly, a *target encoder* is trained adversarially against a *discriminator*. This is performed in a way that given data from the target domain, the target encoder can generate representations that are similar to the representations generated by the source encoder, given data from the source domain. In this sense, the second stage of training in the framework proposed by [2] can be understood like the training of a Generative Adversarial Network [3]. In this analogy, the target encoder is the generator, and the 'real data'

References

[1] G. Wilson and D. J. Cook, "A survey of unsupervised deep domain adaptation," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, no. 5, pp. 1–46, 2020.

[2] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7167–7176, 2017.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.