Master Thesis Projects in Source-Free Active Domain Adaptation

April 2024

1 Introduction

Are you interested in Active Learning or Domain Adaptation? We are interested in supervising talented Bachelor/Master students and publishing high-quality publications.

2 Background

Domain Adaptation [3] is a setting that attempts to transfer a classification network trained on one data set/domain (called *source domain*) to a different, previously unseen domain (called *target domain*). Here, the goal is to use as few or as few labels as possible from the new domain as possible. One way of doing so is Active Domain Adaptation [12]. Here the algorithm is allowed to specify the unlabeled data points that would be most useful if they received labels. These are then sent to a so-called oracle (typically a human annotator) which provides a label for these instances. The algorithm then retrains the network with the additional labels. Depending on the budget, this process is repeated until said budget of data that is allowed to be labeled is reached. One aspect to take into account is that the goal is also to lose as little classification capability in the source domain as possible.
There are many different ways of selecting these instances. Typically the selection is done based on either uncertainty, diversity, domainness, or a mix of these properties. Even within these categories, there are many different selection strategies [4, 11, 20, 18].
Within the setting of domain adaptation, there is a special case called source-free domain adaptation [2]. Here, the original source domain data set that was used to train the original source network is no longer available. This requires specific strategies to handle the training process [19]. Here the usage of pseudo-labels [10, 6] or generative approaches [8] to supplement the data are common [14]. This setting finds use in Active Domain Adaptation as well, with multiple recent strategies being used to address it [9, 17, 7, 16].

3 Topics

3.1 Adjustment of ADA strategies into the ASFDA setting through the usage of pseudo-labels based on DiaNA (MA)

Description While not mainly meant to be an active source-free domain adaptation approach, another recently published approach called DiaNA [5] advertises its compatibility with SFDA. Its handling uses a pseudo-labeling approach.
The primary question of the thesis would be a broad investigation of how this specific approach works with other common ADA methods as well. This has already been done to some degree in the past, though not necessarily with the [14].

Thanks to a repository provided as part of another recent ADA approach [15], the relevant approaches are already implemented in Python. The repository was, however, not originally intended for source-free domain adaptation, so it is necessary to rework the code to account for the change in setting.
To investigate this, the goal would be to investigate the basic selection strategies (random, entropy, margin, least-confidence, and coreset), intermediate strategies (AADA [13], BADGE [1], k-Means), as well as current state-the-art methods (CLUE [11] and LAS [15]). The repository also includes the MHP strategy [17], which has already been used in the SFDA setting and may be useful for comparison.
Further implementations for SFDA methods [7, 9, 16] are also available, and some of them (depending on how well the repository codes work) should be used for comparison. DiaNA itself does not have any code available (yet). If the code becomes available (reasonably early) during the thesis, it can and should be included, but it does not need to be reimplemented as part of this thesis.
The goal here is to see if the approaches are compatible with the SFDA strategy used in DiaNA [5] and how the change in setting impacts the performance of the methods both compared to their prior behavior as well as other ASFDA approaches (including those created during the thesis). Given the aims of the thesis, the investigation should also look at the performance on the source dataset throughout the training process. Due to time and resource constraints, the investigation should be limited to solely the Office-31 datasets. The number of domain adaptation pairings can be restricted based on the state of the thesis, but the current plan is to use all domains available from that baseline.
In case of the approaches not being compatible, an analysis of the reasons why would be required. If the approaches are compatible, depending on the state of the thesis, further investigations into incorporating the unique properties of the selection strategies into the pseudo-labeling process could be performed.

Relevant Resources : `https://github.com/tsun/LADA/tree/master`, `https://github.com/divyakraman/SALAD_SourcefreeActiveLabelAgnosticDomainAdaptation`, `https://github.com/TL-UESTC/ELPT`, `https://github.com/fanwang826/BIAS`

Required Reading : [5, 14, 15, 11, 17, 7, 9, 16]

Requirements

- Study in the field of computer science
- Good understanding of machine learning
- Good programming skills in Python
- Independent working
- Beneficial: understanding of active learning/active domain adaptation

Additional Information

- Start: in April
- Length of the work: 26 weeks
- Location: AoE

Contact information If you are interested, please send your CV and transcripts to jahn@dbs.ifi.lmu.de

References

[1] Jordan T Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. Deep batch active learning by diverse, uncertain gradient lower bounds. *arXiv preprint arXiv:1906.03671*, 2019.

[2] Boris Chidlovskii, Stephane Clinchant, and Gabriela Csurka. Domain adaptation in the absence of source domain data. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 451–460, 2016.

[3] Hal Daume III and Daniel Marcu. Domain adaptation for statistical classifiers. *Journal of artificial Intelligence research*, 26:101–126, 2006.

[4] Bo Fu, Zhangjie Cao, Jianmin Wang, and Mingsheng Long. Transferable query selection for active domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7272–7281, 2021.

[5] Duojun Huang, Jichang Li, Weikai Chen, Junshi Huang, Zhenhua Chai, and Guanbin Li. Divide and adapt: Active domain adaptation via customized learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7651–7660, 2023.

[6] Youngeun Kim, Donghyeon Cho, Kyeongtak Han, Priyadarshini Panda, and Sungeun Hong. Domain adaptation without source data. *IEEE Transactions on Artificial Intelligence*, 2(6):508–518, 2021.

[7] Divya Kothandaraman, Sumit Shekhar, Abhilasha Sancheti, Manoj Ghuhan, Tripti Shukla, and Dinesh Manocha. Salad: Source-free active label-agnostic domain adaptation for classification, segmentation and detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 382–391, 2023.

[8] Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. Model adaptation: Unsupervised domain adaptation without source data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9641–9650, 2020.

[9] Xinyao Li, Zhekai Du, Jingjing Li, Lei Zhu, and Ke Lu. Source-free active domain adaptation via energy-based locality preserving transfer. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 5802–5810, 2022.

[10] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International conference on machine learning*, pages 6028–6039. PMLR, 2020.

[11] Viraj Prabhu, Arjun Chandrasekaran, Kate Saenko, and Judy Hoffman. Active domain adaptation via clustering uncertainty-weighted embeddings. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8505–8514, 2021.

[12] Piyush Rai, Avishek Saha, Hal Daumé III, and Suresh Venkatasubramanian. Domain adaptation meets active learning. In *Proceedings of the NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing*, pages 27–32, 2010.

[13] Jong-Chyi Su, Yi-Hsuan Tsai, Kihyuk Sohn, Buyu Liu, Subhransu Maji, and Manmohan Chandraker. Active adversarial domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 739–748, 2020.

[14] Xin Su, Yiyun Zhao, and Steven Bethard. A comparison of strategies for source-free domain adaptation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8352–8367, 2022.

[15] Tao Sun, Cheng Lu, and Haibin Ling. Local context-aware active domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18634–18643, 2023.

[16] Fan Wang, Zhongyi Han, and Yilong Yin. Bias: Bridging inactive and active samples for active source free domain adaptation. *Knowledge-Based Systems*, 284:111151, 2024.

[17] Fan Wang, Zhongyi Han, Zhiyan Zhang, Rundong He, and Yilong Yin. Mhpl: Minimum happy points learning for active source free domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20008–20018, 2023.

[18] Binhui Xie, Longhui Yuan, Shuang Li, Chi Harold Liu, Xinjing Cheng, and Guoren Wang. Active learning for domain adaptation: An energy-based approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8708–8716, 2022.

[19] Zhiqi Yu, Jingjing Li, Zhekai Du, Lei Zhu, and Heng Tao Shen. A comprehensive survey on source-free domain adaptation. *arXiv preprint arXiv:2302.11803*, 2023.

[20] Fan Zhou, Changjian Shui, Shichun Yang, Bincheng Huang, Boyu Wang, and Brahim Chaib-draa. Discriminative active learning for domain adaptation. *Knowledge-Based Systems*, 222:106986, 2021.