Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Robust Classification under Covariate Shift with Application to Active Learning

Anqi Liu
Department of Computer Science
University of Illinois at Chicago
Chicago, Illinois 60607
aliu33@uic.edu

Introduction

In supervised machine learning, we usually assume the training data is independently and identically distributed (I.I.D.) with the testing data. We then use the trained model to generate predictions when receiving new data with the implicit assumption that they are I.I.D. with the training set, even though it could be untrue in many applications. Unfortunately, model performance can decrease significantly when the distribution generating the new data varies from the distribution that generated the training data. This situation can arise when labeled training data is missing, hard to get access to or just very expensive to uniformly collect. For instance, in a medical study it may be impossible to collect certain disease data in a certain area when the task is predicting that disease across many areas. Similarly, in pool-based active learning, the labeled data is naturally distributed differently from the pool since they are selectively sampled in each iteration from the pool. Therefore, methods that can deal with the discrepancy in training and testing distributions are highly desired.

Covariate shift (Shimodaira 2000) is a setting that assumes the source data distribution for training $P_{src}(x, y)$ is different with the target data distribution for testing $P_{trg}(x, y)$, but the difference only comes from $P_{src}(x) \neq P_{trg}(x)$, which means $P(y|x)$ is the same between training and testing. The assumption of an equal "labeling function" between source and target comes from a very natural idea that we are given a biased set of data to learn from but when given all the possible data, the "global" function that labels the data remains the same. This is a looser assumption than I.I.D. and there are a lot of situations like that. In medical study, whether a patient will suffer from certain disease given features, like a brain image, is independent with whether the training data is sampled from certain distribution. In pool-based active learning, labeled data is originally selected from the pool, where all the data share the same $P(y|x)$, which is also the true labeling function we are actively learning.

All (probabilistic) classifiers will suffer from covariate shift (Fan et al. 2005). This motivates our research. Generally, we try to answer this question: how can we deal with covariate shift and generate predictions that are robust and reliable? We propose to develop a general framework for classification under covariate shift that is robust, flexible and accurate. By "robust," we mean the classifier should provide reasonable performance even when confronted with large distribution discrepancy. By "flexible," we mean the classifier should handle different requirements of loss minimization. And by "accurate," it is essential that the prediction performance satisfies application demands as much as possible. The problems we are faced with include, but are not limited to:

- On the methodology: how to robustly minimize different loss functions subject to covariate shift, the ways to improve the performance or efficiency;
- On the application: how to use the framework for applications in different fields in AI, like active learning;
- On the analysis: how to prove the performance guarantees of the framework, like the ability to generalize under covariate shift.

Related Work

A previously preferred approach to correct covariate shift is to use importance weighting to estimate the prediction loss under the target distribution by reweighting the source samples according to the target-source density ratio, $P_{trg}(x)/P_{src}(x)$ (Shimodaira 2000; Zadrozny 2004). Machine learning research has primarily investigated covariate shift from this perspective, with various techniques for estimating the density ratio (Sugiyama et al. 2008; Huang et al. 2006). Despite asymptotic guarantees of minimizing target distribution loss, importance weighting is often extremely inaccurate for finite sample datasets, when distributions are very different. The reweighted loss will often be dominated by a small number of data points with large importance weights. Additionally, the specific data points with large importance weights vary greatly between random source samples, often leading to high variance model estimates. Theoretically, generalization bounds using importance weighting under covariate shift have only been established when the second moment of sampled importance weights is bounded (Cortes, Mansour, and Mohri 2010).

Existing approaches to active learning generally assumes the unknown datapoint labels follow the inductive biases of

the active learner. However, as previously stated, data produced from an active learner violates the I.I.D. data property broadly assumed by supervised machine learning techniques (Settles 2012). It poses serious pitfalls for active learning methods both in theory and in practice that have not yet been resolved. Current solutions using random labeled seed examples to start with or sampling from mixed strategies usually aim to make label solicitation more random, which undermines the advantages of active learning.

Current Progress

We proposed a novel approach to classification that embraces the uncertainty resulting from covariate shift (Liu and Ziebart 2014). Based on minimax robust estimation (Topsøe 1979; Grünwald and Dawid 2004), our approach is the worst-case distribution possible for a given loss function. We first focused on expected logarithmic loss minimization under covariate shift. The resulting robust bias-aware (RBA) classifier robustly minimizes the logarithmic loss of the target prediction task subject to known properties of data from the source distribution. The parameters of the classifier are optimized via convex optimization to match statistical properties measured from the source distribution.

Active learning, as an important application where natural covariate shift exists, is one of the main components in my research. The RBA classifier is applied to active learning by considering the problem as a covariate shift prediction task and adopting pessimism about all uncertain properties of the conditional label distribution (Liu, Reyzin, and Ziebart 2015). Theoretically, this aligns model uncertainty with prediction loss on remaining unlabeled data points, better justifying the use of the model's label estimates within active learning label solicitation strategies. Moreover, thanks to the cost sensitive method developed recently (Asif et al. 2015), the active learning framework using adversarial prediction is tractable in the 0-1 loss minimization under covariate shift. So it is also applied to active learning in a workshop paper (Liu et al. 2015), even though further effort to improve the performance is still required.

Future Plans

We plan to further develop this robust classification framework under covariate shift in the following three directions followed by problems we would want to solve:

- The methodology: We are now extending the method using kernel representation of features. And we would also like to develop the methodology of the framework in two aspects: a) The RBA classifier depends on a density ratio $P_{src}(x)/P_{trg}(x)$ to control how much the features from source data generalizes to the target and scale certainty of the prediction. Apart from using ad-hoc density estimation methods, we will study different ways to obtain density ratio and figure out the requirement for the density ratio. b) We would also adapt the method to slightly different problem settings. For example, when we have multiple sources data, how can we utilize connections between them to deal with covariate shift in that case?
- The application: We will continue exploring more sophisticated label strategies that will benefit us in both practice and theory in active learning. In addition, we will explore applications in other areas in broader context of AI to which our robust framework could contribute.
- The analysis: We would like to figure out the generalization error analysis under covariate shift setting because the expected target error no longer relates with the training error in the same way in the I.I.D. setting. We are also curious about relation between our model performance and the source-target discrepancy in theory.

References

Asif, K.; Xing, W.; Behpour, S.; and Ziebart, B. D. 2015. Adversarial cost-sensitive classification. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*.

Cortes, C.; Mansour, Y.; and Mohri, M. 2010. Learning bounds for importance weighting. In *Advances in Neural Information Processing Systems*, 442–450.

Fan, W.; Davidson, I.; Zadrozny, B.; and Yu, P. S. 2005. An improved categorization of classifier's sensitivity on sample selection bias. In *Proc. of the IEEE International Conference on Data Mining*, 605–608.

Grünwald, P. D., and Dawid, A. P. 2004. Game theory, maximum entropy, minimum discrepancy, and robust Bayesian decision theory. *Annals of Statistics* 32:1367–1433.

Huang, J.; Smola, A. J.; Gretton, A.; Borgwardt, K. M.; and Schlkopf, B. 2006. Correcting sample selection bias by unlabeled data. In *Advances in Neural Information Processing Systems*, 601–608.

Liu, A., and Ziebart, B. D. 2014. Robust classification under sample selection bias. In *Advances in Neural Information Processing Systems*.

Liu, A.; Asif, K.; Xing, W.; Behpour, S.; D. Ziebart, B.; and Reyzin, L. 2015. Addressing covariate shift in active learning with adversarial prediction. *ICML 2015 workshop of Active learning*.

Liu, A.; Reyzin, L.; and Ziebart, B. D. 2015. Shift-pessimistic active learning using robust bias-aware prediction. In *AAAI Conference on Artificial Intelligence*.

Settles, B. 2012. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 6(1):1–114.

Shimodaira, H. 2000. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference* 90(2):227–244.

Sugiyama, M.; Nakajima, S.; Kashima, H.; Buenau, P. V.; and Kawanabe, M. 2008. Direct importance estimation with model selection and its application to covariate shift adaptation. In *Advances in Neural Information Processing Systems*, 1433–1440.

Topsøe, F. 1979. Information theoretical optimization techniques. *Kybernetika* 15(1):8–27.

Zadrozny, B. 2004. Learning and evaluating classifiers under sample selection bias. In *Proceedings of the International Conference on Machine Learning*, 903–910. ACM.