Robust Feature Extension Algorithms

Ming-Wei Chang Dan Roth
Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL 61801
{mchang21, danr}@uiuc.edu

Since labeling data is very expensive, various machine learning algorithms are designed to use labeled and unlabeled data efficiently. An active area of research over the last few years has focused on attempts to exploit data from multiple tasks or domains in order to improve learning performance on a target task. Algorithms for "multi-task learning", "transfer learning" and "domain adaptation" belong to this category (Duamé, 2007; Raina et al., 2006; Blitzer et al., 2006; Evgeniou & Pontil, 2004). These algorithms are especially important in areas that are naturally divided into many *domains*. For example, a successful Part-of-Speech (POS) tagger trained on news articles can perform very badly on medical articles (Blitzer et al., 2006). Moreover, it is not possible to annotate a lot of labeled data for all domains. Therefore, the use of labeled data from different tasks is crucial.

In this paper, we are interested in problems where multiple tasks are defined, one of which is the *target task*. We refer to all other tasks as *source tasks*. We assume that the number of labeled instances for all source tasks is large enough to build good classifiers. However, the labeled data for target data is scarce and we wish to optimize the performance on the target task. In other words, we want to use the labeled data from the source tasks to improve the performance of the target task. Note that although the definition of the problem is similar to commonly defined "domain adaption", our definition of the problem is broader, since "domain adaption" usually refers to learning to perform on the *same* problem (e.x. POS tagging) on different datasets (news article domain, medical articles domain). We do not have this restriction in our definition.

In order to take advantage of labeled instances from source task, the algorithms need to build some relationship between source tasks and the target task. (Evgeniou & Pontil, 2004) proposed a "regularized" version of multitask learning using a modified support vector machine which assumes that the weight vector of every task has a "shared" linear component. Interestingly, the authors state that their approach is equivalent to the original support vector machine when an extended feature space is used. (Duamé, 2007) proposed an algorithm that projects the instances to a new feature space and shows good experimental results on domain adaptation tasks.

We first point out that these two proposals are essentially the same. While (Evgeniou & Pontil, 2004) solves an optimization problem, (Duamé, 2007) makes use of an analogous problem by extending the feature space without directly using support vector machines. We refer this approach, which extends the feature space, as **Feature Extension algorithm** (FE).

In this paper, we present an analysis of FE algorithms, that shows that these methods are good if the tasks are "close" to each other, under a well defined notion of "close". Specifically, our analysis shows that i) the FE algorithm can only work well if the tasks are close enough, and ii) when two tasks are too close, considering them as a single task can be better than applying the FE algorithm.

In practice, however, some of the source tasks may not satisfy the "closeness" condition, and using the FE algorithm will be sub-optimal. Based on our analysis, we develop a robust FE algorithm, that, on one hand, is designed to exploit "close" source tasks and, on the other, can tolerate source tasks that are both very close and different from the target tasks. We demonstrate experimentally that our algorithm outperforms the existing FE algorithms.

References

Blitzer, J., McDonald, R., & Pereira, F. (2006). Domain adaptation with structural correspondence learning. *Proc. of the Conference on Empirical Methods for Natural Language Processing (EMNLP).* Sydney, Australia.

Duamé, H. (2007). Frustratingly easy domain adaptation. *Proc. of the Annual Meeting of the ACL* (pp. 256–263). Prague, Czech Republic: Association for Computational Linguistics.

Evgeniou, T., & Pontil, M. (2004). Regularized multi–task learning. *Proceedings of the tenth ACM SIGKDD international conference.* (pp. 109–117). New York, NY, USA: ACM.

Raina, R., Ng, A. Y., & Koller, D. (2006). Constructing informative priors using transfer learning. *Proc. of the International Conference on Machine Learning (ICML)* (pp. 713–720). New York, NY, USA: ACM.

Topic: Transfer Learning, Domain adaptation
Preference: Oral