

5 Conclusion

In this paper we proposed an algorithm that requests labels for training points actively selected from an unlabeled pool, and also requests labels for randomly sampled points from the pool for model selection. The algorithm outputs a single model, from a set of candidate models, that has high accuracy on test data, while requesting fewer total labels than several baselines. Moreover, it performs almost as well as an active learning oracle that knows the optimal regularization parameters in advance. An interesting direction for future research is to investigate how to mitigate the downsides of importance weighting with small samples so that the need to sample an unbiased validation set for model selection can be reduced or eliminated.

References

- Baram, Y.; El-Yaniv, R.; and Luz, K. 2003. Online choice of active learning algorithms. In *ICML*, 19–26.
- Beygelzimer, A.; Hsu, D.; Langford, J.; and Tong, Z. 2010. Agnostic active learning without constraints. In Lafferty, J.; Williams, C. K. I.; Shawe-Taylor, J.; Zemel, R.; and Culotta, A., eds., *Advances in Neural Information Processing Systems 23*. 199–207.
- Cortes, C.; Mohri, M.; Riley, M.; and Rostamizadeh, A. 2008. Sample selection bias correction theory. In *Proceedings of the 19th international conference on Algorithmic Learning Theory*, ALT '08, 38–53. Berlin, Heidelberg: Springer-Verlag.
- Cortes, C.; Mansour, Y.; and Mohri, M. 2010. Learning bounds for importance weighting. In Lafferty, J.; Williams, C. K. I.; Shawe-Taylor, J.; Zemel, R.; and Culotta, A., eds., *Advances in Neural Information Processing Systems 23*. 442–450.
- Dasgupta, S.; Kalai, A. T.; and Monteleoni, C. 2009. Analysis of perceptron-based active learning. *J. Mach. Learn. Res.* 10:281–299.
- Dudík, M.; Langford, J.; and Li, L. 2011. Doubly robust policy evaluation and learning. In *ICML*, 1097–1104.
- Greiner, R.; Grove, A. J.; and Roth, D. 2002. Learning cost-sensitive active classifiers. *Artif. Intell.* 139(2):137–174.
- Huang, J.; Smola, A. J.; Gretton, A.; Borgwardt, K. M.; and Schölkopf, B. 2006. Correcting sample selection bias by unlabeled data. In Schölkopf, B.; Platt, J.; and Hoffman, T., eds., *NIPS*, 601–608. MIT Press.
- Kapoor, A., and Horvitz, E. 2009. Breaking boundaries between induction time and diagnosis time active information acquisition. In *NIPS*, 898–906.
- Lewis, D. D., and Gale, W. A. 1994. A sequential algorithm for training text classifiers. In *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '94, 3–12. New York, NY, USA: Springer-Verlag New York, Inc.
- Madani, O.; Lizotte, D. J.; and Greiner, R. 2004. Active model selection. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, UAI '04, 357–365. Arlington, Virginia, United States: AUAI Press.
- Mazzoni, D.; Wagstaff, K.; and Burl, M. C. 2006. Active learning with irrelevant examples. In *ECML*, 695–702.
- Miller, D., and Yan, L. 1999. Critic-driven ensemble classification. *Trans. Sig. Proc.* 47(10):2833–2844.
- Platt, J. C. 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *ADVANCES IN LARGE MARGIN CLASSIFIERS*, 61–74. MIT Press.
- Roy, N., and McCallum, A. 2001. Toward optimal active learning through sampling estimation of error reduction. In *In Proc. 18th International Conf. on Machine Learning*, 441–448. Morgan Kaufmann.
- Sawade, C.; Landwehr, N.; Bickel, S.; and Scheffer, T. 2010. Active risk estimation. In *ICML*, 951–958.
- Seung, H. S.; Opper, M.; and Sompolinsky, H. 1992. Query by committee. In *Proceedings of the fifth annual workshop on Computational learning theory*, COLT '92, 287–294. New York, NY, USA: ACM.
- Sugiyama, M., and Rubens, N. 2008. Active learning with model selection in linear regression.
- Sugiyama, M.; Krauledat, M.; and Müller, K.-R. 2007. Covariate shift adaptation by importance weighted cross validation. *J. Mach. Learn. Res.* 8:985–1005.
- Sutton, R., and Barto, A. 1998. *Reinforcement Learning: An Introduction*. A Bradford book. Bradford Book.
- Tong, S., and Koller, D. 2000. Support vector machine active learning with applications to text classification. In *Journal of Machine Learning Research*, 999–1006.
- Vapnik, V., and Chapelle, O. 2000. Bounds on error expectation for support vector machines. *Neural Comput.* 12(9):2013–2036.
- Yan, Y.; Rosales, R.; Fung, G.; Farooq, F.; Rao, B.; and Dy, J. G. 2012. Active learning from multiple knowledge sources. In *AISTATS*, 1350–1357.