A Self-training Approach to Cost Sensitive Uncertainty Sampling⋆

Alexander Liu, Goo Jun, and Joydeep Ghosh

Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, Texas 78712
U.S.A.
{ayliu,gjun}@mail.utexas.edu, ghosh@ece.utexas.edu

Abstract. Uncertainty sampling is an effective method for performing active learning that is computationally efficient compared to other active learning methods such as loss-reduction methods. However, unlike loss-reduction methods, uncertainty sampling cannot minimize total misclassification costs when errors incur different costs. This paper introduces a method for performing cost-sensitive uncertainty sampling that makes use of self-training. We show that, even when misclassification costs are equal, this self-training approach results in faster reduction of loss as a function of number of points labeled and more reliable posterior probability estimates as compared to standard uncertainty sampling. We also show why other more naive methods of modifying uncertainty sampling to minimize total misclassification costs will not always work well.

Reference

1. Liu, A., Jun, G., Ghosh, J.: A Self-training Approach to Cost Sensitive Uncertainty Sampling. Machine Learning (2009) DOI: 10.1007/s10994-009-5131-9

⋆ This is an extended abstract of an article published in the machine learning journal [1].