

Budgeted Learning of Probabilistic Classifiers

Russell Greiner
Department of Computing Science
and Alberta Ingenuity Centre for Machine Learning
University of Alberta
Edmonton, Alberta, Canada T6G 2E8
greiner@cs.ualberta.ca

ABSTRACT

Researchers often use clinical trials to collect the data needed to evaluate some hypothesis, or produce a classifier. During this process, they have to pay the cost of performing each test. Many studies will run a comprehensive battery of tests on each subject, for as many subjects as their budget will allow – i.e., “round robin” (RR). We consider a more general model, where the researcher can sequentially decide which single test to perform on which specific individual; again subject to spending only the available funds. Our goal here is to use these funds most effectively, to collect the data that allows us to learn the most accurate classifier.

We first explore the simplified “*coins version*” of this task. After observing that this is NP-hard, we consider a range of heuristic algorithms, both standard and novel, and observe that our “biased robin” approach is both efficient and much more effective than most other approaches, including the standard RR approach. We then apply these ideas to learning a *naive-bayes classifier*, and see similar behavior. Finally, we consider the most realistic model, where *both* the researcher gathering data to build the classifier, and the user (e.g., physician) applying this classifier to an instance (patient) must pay for the features used --- e.g., the researcher has \$10,000 to acquire the feature values needed to produce an optimal \$30/patient classifier. Again, we see that our novel approaches are almost always much more effective than the standard RR model.

This is joint work with Aloak Kapoor, Dan Lizotte and Omid Madani. See the Budgeted Learning Webpage at <http://www.cs.ualberta.ca/~greiner/BudgetedLearning>.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

UBDM '06, August 20, 2006, Philadelphia, Pennsylvania, USA.

Copyright 2006 ACM 1-59593-440-5/06/0008...\$5.00.