Refining Inductive Bias in Unsupervised Learning via Constraints

Kiri Wagstaff
Department of Computer Science
Cornell University
Ithaca, NY 14853
wkiri@cs.cornell.edu

Algorithmic bias is necessary for learning because it allows a learner to generalize rationally. A bias is composed of all assumptions the learner makes outside of the given data set. There exist some approaches to automatically selecting the best algorithm (and therefore bias) for a problem or automatically shifting bias as learning proceeds. In general, these methods are concerned with supervised learning tasks. However, reducing reliance on supervisory tags or annotations enables the application of learning techniques to many real-world data sets for which no such information exists. We therefore propose the investigation of methods for refining the bias in unsupervised learning algorithms, with the goal of increasing accuracy and improving efficiency. In particular, we will investigate the incorporation of background knowledge in the form of constraints that allow an unsupervised algorithm to automatically avoid unpromising areas of the hypothesis space.

Background Knowledge as Constraints. There is a natural connection between the bias in an algorithm and background knowledge. Often, the bias hardcoded into an algorithm was chosen due to background knowledge about the class of tasks to be targeted. This bias encodes certain assumptions about what sort of hypotheses are valid solutions for *any* problem it is applied to. However, for a specific task it is often the case that more precise information is available that can be used to augment the bias in useful ways. In such cases, it is desirable to leverage this background knowledge to refine the algorithmic bias in the proper direction.

In particular, we are interested in improvements that can be obtained with the addition of problem-specific constraints. Constraints are derived from background knowledge and specify relationships between instances that may not be expressible in the traditional feature-value representation used for machine learning data sets.

Current and Proposed Work. To date, we have investigated the incorporation of instance-level hard constraints into one clustering algorithm (a partitioning variation of COBWEB (Fisher 1987)). We found that incorporating constraints results in improved clustering accuracy (Wagstaff & Cardie in press). The types of constraints investigated were specific to algorithms that create flat partitions of the input data. We plan to investigate the relative merits of different kinds of constraints (e.g. hard vs. soft, feature-level vs. instance-level, probabilistic vs. deterministic) as applied to a variety of algorithms (partitioning vs. hierarchical, those that use a distance measure vs. those that do not, etc.).

In addition, a number of interesting questions were raised in the course of our previous work. First, does the distribution of instances (how many are from each class) affect the efficacy of constraints? Second, in our experiments, we observed that the category utility (CU) of the "correct" (fully-constrained) partition was *lower* than that obtained without using constraints. Does this indicate that CU is a poor choice of objective function in clustering? What does this signify about the correlation of the class label and other attributes, and ultimately about the relative "difficulty" of the data set? Lastly, how can we generate constraints to be used by these techniques?

Evaluation of Constraint Techniques. In order to assess the techniques developed, we plan to evaluate them on a variety of real-world and artificial data sets. Of particular importance is a determination of the relationship between the amount of information contained in the constraints and the magnitude of any accuracy improvements observed.

For some domains, constraints on which instances can or cannot reside in the same cluster are known or are automatically computable from background knowledge. In the problem of noun phrase coreference, for example, instance-level constraints can be computed from background linguistic knowledge. Other good candidates for evaluating constraint techniques are domains where class labels are known for a small subset, but not all, of the instances. Artificial data sets will be useful for exploring what effect the class distribution of instances has on clustering accuracy. In addition, we expect to use them to investigate the observed effect on category utility when constraints are used.

Acknowledgements: We would like to thank Claire Cardie for research advice and Westley Weimer for proofreading and suggestions. This work is supported by an NSF Graduate Fellowship.

References

Fisher, D. 1987. Knowledge Acquisition Via Incremental Conceptual Clustering. *Machine Learning* 2:139–172.

Wagstaff, K., and Cardie, C. in press. Clustering with Instance-Level Constraints.