

Discussion of  
"Exploiting the Absence of Irrelevant  
Information"  
by Rao, Greiner and Hancock

This paper has two main innovations. The first is to provide a new model for blocked attributes. The second is to redefine the notion of an irrelevant variable.

As stated in the paper, many standard models for blocked attributes assume that attributes are blocked according to a random process. Other non-random models of blocked attributes also exist. Consider the cases where it is too expensive or time-consuming to find the value of an attribute, or an attribute itself is unknown. In these cases, the same attributes are blocked in every example. The learner may see many examples which look identical (because they differ only on blocked attributes) but are labeled differently. The target concept then looks like a probabilistic function, labeling an example as positive with probability  $p$ , and negative with probability  $1 - p$ . Work on learning probabilistic concepts (e.g. [1]) was motivated in part by this model of blocked attributes.

The authors of this paper consider yet another model, in the context of decision-tree learning. In their model, instances are generated attribute by attribute traversing the decision tree until a leaf is reached; all other attributes are then blocked. The authors motivate their model well; there are certainly cases where it is appropriate.

In the context of functions, the standard definition of an irrelevant variable is that it never affects the output of the function. In this paper, the definition of relevance is relative, and dependent on prior knowledge. For example, once I know you are six years old, I know you can't vote in the United States. Whether you are a U.S. Citizen is irrelevant.

The paper gives a simple algorithm for PAC learning decision trees in the new model (as well as incremental algorithms which I will not discuss). Although blocked data sometimes makes learning harder, in this model, the blocked data provides added information. First, given a particular example, the learner obtains functional information; no matter what values are given to the blocked attributes, the output of the decision tree will be the same. Second, the learner obtains representational information about the decision tree itself; the unblocked variables correspond to the variables along a root-leaf path in the decision tree.

It is actually the representational information on which the algorithm relies, and not the functional information. One might complain therefore that the model is artificially powerful. In the case of decision trees, though, the choice of which attributes to block (in the learning model) is natural, and the transmission of representational information by the blocked attributes can be viewed as a fortunate side-effect.

For other classes, it is less clear that a natural choice exists. The authors note that in the case of DNF, the difficulty of the learning problem can vary tremendously depending on how the blocked attributes are chosen. If only functional information is provided, instances are prime implicants of the target formula or its complement. This model has the advantage of being applicable to all classes of boolean functions. However, it may have no practical motivation. Nevertheless, as a contrast to the results in this paper, it would be interesting to determine the difficulty of learning decision trees in this weaker model.

In some cases, the functional information alone would be extremely helpful. Consider, for example, the problem of learning functions of  $\log n$  out of  $n$  variables. It is still not known whether this class is poly-time PAC predictable, although certain types of algorithms are provably ineffective [2]. For this class, the  $\log n$  variables determining the function are the only ones ever unblocked. The number of possible different examples (prime implicants) is  $O(n)$ , and learning is trivial.

L. Hellerstein

[1] Schapire, R. E. *The Design and Analysis of Efficient Learning Algorithms*. MIT Press. Cambridge, MA, 1992.

[2] M. Kearns. Efficient noise-tolerant learning from statistical queries. In *Proceedings of the 25th Annual ACM Symposium on the Theory of Computing*, 1993.