Diana F. Gordon

Navy Center for Applied Research in Artificial Intelligence
Naval Research Laboratory, Code 5510
Washington, D.C. 20375
gordon@aic.nrl.navy.mil

Abstract

Selecting a good bias prior to concept learning can be difficult. Therefore, dynamic bias adjustment is becoming increasingly popular. Current dynamic bias adjustment systems, however, are limited in their ability to identify erroneous assumptions about the relationship between the bias and the target concept. Without proper diagnosis, it is difficult to identify and then remedy faulty assumptions. We have developed an approach that makes these assumptions explicit, actively tests them with queries to an oracle, and adjusts the bias based on the test results.

1 Introduction

Bias is a fundamental aspect of any supervised concept learner. Numerous papers have noted this importance (e.g., Mitchell 1980; Haussler 1988). The type of bias that we discuss here is the choice of a hypothesis language. The hypothesis language defines the space of hypotheses. A *strong* bias defines a small hypothesis space; a *weak* bias defines a large hypothesis space; a *correct* bias defines a space that includes the target concept. A strong correct bias, e.g., one with fewer features, is generally desirable because it reduces the number of hypothesis choices and thereby promotes rapid convergence to the target concept.

The bias can be adjusted (shifted) dynamically during incremental concept learning by strengthening the bias when possible and weakening it to regain correctness. Recently, interest has grown in systems that dynamically shift the bias (e.g., Utgoff 1986; Rendell 1990; Spears & Gordon 1991). These systems, however, are limited in their ability to identify erroneous assumptions about the relationship between the bias and the target concept. Proper diagnosis aids in the recovery from faulty assumptions. We have developed an approach to bias adjustment that addresses this need for proper diagnosis. Our method consists of a bias tester and adjuster that can be added to an incremental concept learner to improve the learner's performance.

Unlike previous approaches to bias testing, our approach uses formal definitions of assumptions about the bias, called *biasing assumptions*, to guide an analysis of *why* the bias is inappropriate (e.g., too weak, or incorrect) for learning the target concept. An example of a biasing assumption is the irrelevance of a feature for learning the target concept. The bias tester performs this analysis (called a *biasing assumption test*) by actively testing the bias with queries to an oracle. Each query is a request to an instance generator for a new instance. For example, the irrelevance of a feature might be tested by querying an oracle for the class (positive/negative) of instances having different values of that feature. The bias adjuster then records the analysis results and adjusts the bias accordingly. If a biasing assumption holds, the adjuster strengthens the bias, e.g., by removing the irrelevant feature from the hypothesis language. Otherwise, the adjuster weakens the bias or allows the bias to stay the same if no adjustments are needed.

Our approach has three primary advantages. First, the bias tests are composed of queries. Queries can accelerate learning significantly (see Gordon 1990; 1992). Second, our approach is designed to be incorporated into an existing concept learner. Third, our approach diagnoses the bias to find and record specific erroneous biasing assumptions. This enables the bias to be *minimally weakened* as well as corrected. Minimal weakening is most advantageous when a stronger bias is desirable. In that case, bias strengthening along with minimal bias weakening can enable very rapid acquisition of the target concept (see Gordon 1990; 1992).

In our framework, the bias is the set of features and their values in the hypothesis language. These values appear in *value trees* (e.g., see Figure 1), which are input by a user or knowledge engineer who is somewhat familiar with the domain. Value trees are