Induction: Weak but Essential

Thomas G. Dietterich

Department of Computer Science

Oregon State University

Corvallis, OR 97331

(503) 754-4466

Schank, Collins, and Hunter are certainly correct when they state that inductive learning methods do not provide an adequate theory of human learning. However, few researchers would make such a claim. Induction is a weak method, but it is still appropriate for situations in which weak methods are appropriate—namely, situations in which there is no knowledge available. Furthermore, induction can be applied as a method of last resort, after all available knowledge has been exploited.

What *is* induction? It is a computational method for searching a space of alternative category definitions by matching candidate definitions against given members and non-members of the category. Candidate definitions that match members (positive examples) are preferred to candidate definitions that don't match members or that match non-members (negative examples). This definition highlights two things.

First, the key operation of induction is *matching.* A consequence of this is that inductive methods are critically dependent on the representation of training examples and of category definitions. No other knowledge, aside from that implicit in the representation, is brought to bear. It is for this reason that Newell (1969) has grouped induction with the other weak methods such as generate-and-test, hill climbing, and means-ends analysis. The great virtue of weak methods is that they place very few demands on the task environment. When the stronger explanation-based methods described by Schank, et al., break down, induction can still be applied.

The second point to note about the definition of induction is that it doesn't mention necessary and sufficient conditions. When a category definition fails to match a positive example or matches a negative example, this does not necessarily eliminate the candidate definition—it merely produces a preference against it.

An AI system that illustrates both of these points is the Meta-DENDRAL system (Buchanan & Mitchell, 1978). Meta-DENDRAL employs an explanation-based approach to learn category definitions that do not involve necessary and sufficient conditions. Yet induction methods play a major role in the system.

The task of Meta-DENDRAL is to form a causal model of the operation of a chemical instrument called a mass spectrometer. In a mass spectrometer, small quantities of an unknown chemical sample are bombarded by high energy electrons. This bombardment causes the molecules in the sample to break into electrically-charged fragments. The fragments are then subjected to a magnetic field, which separates them according to the ratio of their mass to their charge. The result is a mass spectrum—a histogram showing the number of observed fragments plotted against their mass-to-charge ratio. Different molecular structures produce different mass spectra—somewhat like finger prints—and a skilled chemist can often identify the structure of a molecule by analyzing its spectrum. However, it is difficult to predict in advance how the mass spectrum (i.e., the finger print) of a molecule will appear.

Meta-DENDRAL discovers rules for predicting the mass spectra of classes of molecules. It is given training examples, each consisting of a molecule with known chemical structure and known mass spectrum. Essentially, Meta-DENDRAL's task is to discover rules, called *cleavage rules*, that tell when and where the molecule will break during the electron bombardment. A typical cleavage rule looks like

$$C - C - N - C \Rightarrow C - C * N - C,$$

which says that if the molecule contains a chain of two carbons bonded to a nitrogen bonded to a carbon then that chain will break at the carbon-nitrogen bond.[1]

To learn these rules, Meta-DENDRAL proceeds as follows. First, it analyzes the spectra of the given molecules and attempts to find an explanation for each fragment in the spectrum. To construct these explanations, it applies a *half-order theory* of mass spectrometers that basically says "all bonds can break." For each fragment observed in the spectrum, it asks "What bonds in the molecule could have broken to yield this fragment?" If Meta-DENDRAL had a completely accurate theory of mass spectrometry, each fragment would have a unique explanation and no learning would be necessary. However, the half-order theory is so weak that each fragment typically has several alternative explanations.

Consequently, Meta-DENDRAL must find some way of sifting through them to identify the most plausible ones. It does this by induction. The induction task is to discover a definition for the category "bonds that break during bombardment." Each alternative explanation can be viewed as a positive example that says "These specific bonds broke in this molecule." Meta-DENDRAL starts the induction process by considering the most general rule

$$X - X \Rightarrow X * X,$$

which is the same as the half-order theory (i.e., all bonds break; X denotes any atom). It then adds features (conditions) to this rule by matching it against the training examples. The matching

[1] There are other requirements on the atoms that are not shown here including bonding to other atoms, membership in rings, etc.

process usually divides this rule into sub-rules to handle positive examples that cannot be matched by a single cleavage rule. The process is similar to a top-down clustering process. It terminates when roughly half of the training examples are covered by some cleavage rule and when the rules are sufficiently specific (according to some domain-specific, but relatively syntactic, criteria).

There are three points to observe in this example. First, although Meta-DENDRAL employs explanations in the learning process (along the lines suggested by Schank, et al.), these explanations are faulty and cannot solve the entire learning problem. If they could, there would be no need for learning at all—the system would already know the correct theory of mass spectrometry. Second, when the knowledge employed to construct explanations was exhausted, induction was still applicable, because it is such a weak method. Third, the category definitions constructed by Meta-DENDRAL are neither necessary nor sufficient to predict exactly where molecules will break under bombardment. There is too much noise in the data, and the explanation construction process (using the half-order theory) is too weak, to provide high quality training instances.

In conclusion, there is a role for induction in a theory of learning. While Meta-DENDRAL is not a model of human learning, the lessons it teaches are relevant (see VanLehn 1983). Explanation-based learning, if it is not to trivialize the learning process, must be error-prone and incomplete. To overcome these errors, induction may need to be performed *on the explanations themselves.* The same kinds of syntactic criteria (e.g., prefer syntactically short category definitions) will need to be applied to explanations. Occam's Razor, after all, tells us to "Prefer the simplest explanation." When our knowledge gives out, induction is the only method we have left to apply.

Buchanan, B. G. & Mitchell, T. M. (1978) Model-based learning of production rules in *Pattern-Directed Inference Systems,* ed. D. A. Waterman, & F. Hayes-Roth. pp 297–312 Academic Press.

Newell, A. (1969) Heuristic programming: Ill-structured problems. in *Progress in Operations Re-*

search, ed J. Arnofsky pp 363–414 Wiley.

VanLehn, K. (1983) Human procedural skill acquisition: theory, model, and psychological validation in *Proceedings of the National Conference on Artificial Intelligence,* pp 420–423 Morgan-Kaufmann.