Unifying Themes in Empirical and Explanation-Based Learning

Pat Langley (LANGLEY@ICS.UCI.EDU)
Department of Information & Computer Science, University of California, Irvine, CA 92717 USA

The Need for Unified Theories of Learning

A central activity of science is the search for unifying principles that account for apparently diverse phenomena within a single framework. However, recent work in machine learning has tended to emphasize the differences between learning methods. In this paper, I argue that two of the major paradigms – induction and explanation-based learning – are more similar than the literature suggests, and that we must focus on these similarities before we can build a unified theory of learning mechanisms.

Significant differences certainly exist between explanation-based and empirical methods, but the perceived chasm is far greater than the actual one. This perception has resulted partly from a literature that abounds with rhetorical statements claiming superiority of one method over another. Other causes for the perceived distinction include divergent notations and different measures of performance, which hide the underlying similarity of mechanisms and tasks. In this paper, I present examples of misleading rhetoric and conflicting metrics that the field must overcome before it can approach a unified theory of learning.

Learning from One Instance and Many Instances

One common claim is that empirical methods require many instances to learn, whereas EBL can learn from a single instance (e.g., Mitchell, Keller, & Kedar-Cabelli, 1986, pp. 47–48). This misleading statement probably results from comparisons between explanation-based methods (which are typically incremental) and nonincremental induction methods, such as Quinlan's (1986) ID3. However, if one examines *incremental* inductive methods, such as Fisher's (1987) COBWEB, the true situation becomes apparent. Any incremental approach to induction (even neural networks) can learn *something* from a single instance, though it may not learn *as much* as an EBL technique.

The above claim also suggests that EBL methods can learn *everything they need to know* from a single instance, but this is clearly false as well. Analytic techniques require one instance for each proof structure they compile. For example, Pazzani's (1988) OCCAM acquires four schemata for recognizing when economic sanctions will fail and three schemata for predicting when they will succeed; thus, it requires not one training instance for this domain, but seven. Although EBL techniques may learn more rapidly than empirical methods, this is a difference in learning *rate*, not a difference between one and many instances.

Learning With and Without Search

A second popular belief is that empirical methods require extensive search, whereas explanation-based methods can learn without search. Again, this statement is misleading on two fronts. First, it focuses on inductive methods like Mitchell's (1982) version-space algorithm, which use memory-intensive search techniques to consider competing hypotheses. However, many inductive methods rely on memory-limited methods such as greedy algorithms (Quinlan, 1986) and incremental hill climbing (Fisher, 1987). Although such methods operate within a space of hypotheses, they do not 'search' in the usual sense of this term.

On the other hand, if one views the explanation process as a component of learning (rather than as performance), then EBL itself can involve extensive search through the space of explanations. Work in this paradigm has not emphasized this search because, to date, most tests have involved relatively small domain theories. In addition, one goal of EBL is to improve efficiency, and Minton (1988) has shown that adding compiled rules to the knowledge base sometimes produces just the opposite effect. To deal with this issue, his PRODIGY system computes statistics for learned rules, deleting those that are not worth retaining. One can view this process as search through a space of compiled rules, just as empirical methods search a space of induced rules. Whether one labels either activity as 'search' is less important than the realization that both frameworks must deal with large rule spaces.

Learning With and Without Domain Knowledge

Yet another claim is that explanation-based methods take domain knowledge into account during learning, whereas empirical methods are knowledge free (e.g., Mitchell et al., 1986, p. 48). The first part of this statement is true enough, but the second half ignores the fact that any *incremental* induction system inevitably changes its knowledge level over time. After such a system has seen n instances, it will process instance $n + 1$ differently than if it had seen it first. For example, Fisher's (1987) COBWEB constructs a concept hierarchy that organizes instances it has encountered, and the structure of this memory influences not only the predictions it makes on new instances, but the learning that occurs. Thus, COBWEB takes advantage of domain knowledge to direct the learning process. The fact that it acquires this knowledge itself (rather than receiving it from the programmer) makes it no less knowledge intensive.

As another example, consider Wolff's (1982) SNPR algorithm, which is generally viewed as lying at the extreme end of the *tabula rasa* spectrum. This system accepts a sequence of letters as input, and carries out a hill-climbing search through the space of phrase-structure grammars, using two basic operators. The first notes frequently occurring sequences of symbols and defines new 'chunks', which correspond to words and phrases. The second learning operator notes when sets of symbols tend to occur in the same context (i.e., next to a common symbol); this defines new disjunctive classes, which correspond to parts of speech and alternative forms of phrases.

If one looks only at the relation between SNPR's inputs and outputs, it appears to be the prototypical 'knowledge free' induction system. However, the algorithm is semi-incremental, in that it processes only part of its input at a given time, using the knowledge it gains from earlier data in processing its later experience. Specifically, SNPR constructs a partial grammar to summarize the letter sequences it has observed, and it uses this grammar to rewrite new strings at a higher level of description (i.e., using nonterminal symbols in the grammar). One can view this activity as constructing *partial explanations* of the input, and one can view the later stages of grammar induction as a form of *knowledge-intensive* learning that involves extending an incomplete domain theory (the set of grammar rules). Although phrase-structure grammars are a constrained form of domain theory, they are very similar in structure to those used by many EBL systems.

Justified and Unjustified Learning

A fourth claim is that explanation-based methods are justified, whereas empirical learning is inherently unjustified (e.g., Mitchell et al., 1986, p. 48). The latter statement is clearly true, since empirical learning involves an inductive leap from instances to general rules. However, the justified nature of EBL is not so clear. Rules generated by analytic methods are guaranteed to be as *accurate* as the original domain theory, since the deductive closure does not change. However, they may not be as *efficient* as the original rule set. The common assumption that EBL will improve efficiency is based on the belief that training and test instances will follow similar distributions. Thus, analytic methods make an inductive leap with respect to efficiency that is no more justified than the leap made by empirical methods regarding accuracy.

In addition, one can extend the basic explanation-based learning framework to domains in which the inference rules, rather than being deductively valid, are plausible or probabilistic. In such domains, the process of compiling multi-step explanations may generate 'bad' inference rules that have very low predictive ability, since transitivity does not hold for probabilistic inference chains as it does for deductive chains. In such an extended framework, analytic learning methods are not even justified with respect to predictive accuracy. Given a reasonably accurate domain theory, such methods may still lead to more rapid learning, but they are not any more 'correct' than inductive methods.

Accuracy and Efficiency in Machine Learning

The term *learning* suggests some change in performance, and the empirical and explanation-based communities have been further divided by their concern with different performance measures. Most research on induction has focused on improving predictive accuracy, whereas most analytical work has (implicitly if not explicitly) focused on efficiency. However, both measures of performance have an important role to play in both approaches to learning.

For example, any performance system has limited memory size and processing time; thus, adding rules that reduce memory load or increase speed can let one finish complex tasks that were impossible before learning. This means that EBL can produce improvements in predictive accuracy, and suggests that researchers should measure it in future studies. Similarly, any induction system that deals with a complex domain will create many different concepts. If organized ineffectively, this acquired knowledge may drastically slow the performance system. This means that retrieval time is a central issue in empirical learning, and that induction researchers should examine this performance measure as well.

As work in both paradigms starts to bridge this gap, it may reveal previously unsuspected connections between induction and EBL. For instance, psychological studies suggest that humans recognize certain *basic-level* categories more rapidly than other concepts. Fisher's (1987) COBWEB/2 – an empirical learning system – models this effect with a mechanism that creates direct indices to some nodes in its concept hierarchy and that bypasses other concepts. In spirit, this operation is remarkably similar to the caching process by which many EBL methods store operationalized definitions of concepts to improve retrieval efficiency.

Like the examples in previous sections, this connection suggests that empirical and explanation-based methods have much more in common than the literature leads one to expect. If researchers in the two paradigms can rise above the rhetoric and assumptions that have kept them apart, they can move together toward a unified science of machine learning that incorporates insights from both frameworks.

References

Fisher, D. H. (1987). *Knowledge acquisition via incremental conceptual clustering*. Doctoral dissertation, Department of Computer Science, University of California, Irvine.

Minton, S. (1988). Quantitative results concerning the utility of explanation-based learning. *Proceedings of the Seventh National Conference on Artificial Intelligence* (pp. 564–569). St. Paul, MN: Morgan Kaufmann.

Mitchell, T. M. (1982). Generalization as search. *Artificial Intelligence*, *18*, 203–226.

Mitchell, T. M., Keller, R. M., & Kedar-Cabelli, S. T. (1986). Explanation-based generalization: A unifying view. *Machine Learning*, *1*, 47–80.

Pazzani, M. (1988). *Learning causal relationships: An integration of empirical and explanation-based techniques*. Doctoral dissertation, Department of Computer Science, University of California, Los Angeles.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*, 81–106.

Wolff, J. G. (1982). Language acquisition, data compression, and generalization. *Language and Communication*, *2*, 57–89.