EDITORIAL

Machine Learning and Concept Formation

The task of concept formation

The vast majority of research in machine learning has focused on the acquisition of *concepts*. Yet despite this emphasis, little of the work has any direct relevance to modeling human concept learning. For a variety of reasons, machine learning researchers have tended to make assumptions that violate our knowledge about the representation, use, and acquisition of human concepts.

For instance, most AI researchers have attempted to describe concepts in logical terms, as the conjunction or disjunction of features. However, we know that many human concepts cannot be described in such a logical formalism. Some instances of a concept are more typical than others; e.g., robins are more typical birds than penguins. This would seem to require a more flexible representation, combined with some form of partial matching.

Another common assumption is that concept learning is *supervised*; i.e., a benevolent teacher is present to direct the learning process. However, we know that much human concept learning is *unsupervised*. Children form useful concepts long before they acquire enough language to benefit from a tutor's advice, and they seem to accomplish this through direct interaction with their environment.

Researchers have also tended to focus on learning one or a few concepts at a single level of abstraction. However, we know that humans organize their concepts into large hierarchies that describe categories at varying levels of specificity. Thus, models of concept learning should address the formation of conceptual hierarchies as well as the formation of individual categories.

Finally, much of the work on concept acquisition has used nonincremental learning methods. Yet we know that humans interact with their environment over time, encountering new instances and incorporating them into memory one after another. This strongly suggests that humans form their conceptual hierarchies in an incremental fashion, and that this process is closely related to the classification of new instances. Thus, models of human concept learning should employ incremental methods that are fully integrated with a performance element.

Taken together, these observations suggest a new task for machine learning – the task of *concept formation*. We will use this term to refer to the acquisition of conceptual hierarchies in which each concept has a flexible, non-logical definition and in which learning occurs incrementally and without supervision. We believe these features characterize much of the process of human concept learning.

Models of concept formation

Exceptions to the above trends certainly exist. For instance, much of the work on learning from examples has taken an incremental approach. Similarly, research on conceptual clustering has focused on the construction of concept hierarchies from unsupervised data. And recent work on learning from examples has explored more flexible concept representations and partial matching (Michalski, 1987; Quinlan, 1987). But very few researchers have broken with all the assumptions listed above and focused on the full task of concept formation.

However, the papers by Lebowitz and Fisher in the current issue do exactly that. Neither author puts forth his system as a psychological model *per se*, but both reject the four assumptions listed above and replace them with more plausible ones. Thus, Lebowitz's UNIMEM and Fisher's COB-WEB allow partial matching to occur, and the latter system replaces logical concept definitions with probabilistic descriptions. Both systems learn without aid from a tutor and both construct entire concept hierarchies rather than single-level categories. Finally, both COBWEB and UNIMEM acquire knowledge incrementally, modifying their concepts and hierarchies in the very act of classifying new instances. In other words, they fully integrate the learning and performance aspects of classification.

Fisher's and Lebowitz's systems also share a number of other features. Like Feigenbaum's (1963) early EPAM model of human learning, both UNIMEM and COBWEB construct their concept hierarchies in a top-down fashion, though each includes other means for modifying the hierarchy as well.[1] In addition, both incorporate numeric information into their concept representations, though the interpretation of these numbers is quite different.

Moreover, both Fisher and Lebowitz explicitly reject the systematic search paradigm that has dominated research on empirical learning methods. In contrast to Mitchell's (1982) breadth-first search or Michalski's (1983) beam search, the authors in this issue allow only one 'hypothesis' in memory at a time. Their systems can be viewed as searching the space

---

[1] We should briefly mention Hanson and Bauer's (1986) WITT, a concept formation system that also rejects the assumptions listed earlier. Unlike UNIMEM and COBWEB, this system constructs hierarchies in an agglomerative manner, from the bottom up.

of concept hierarchies, but the search method is very weak and best characterized as hill climbing. This strategy is not guaranteed to achieve the optimum solution, but the heuristic approaches used by COBWEB and UNIMEM give good results in practice and they have very low computational requirements. We expect that other machine learning researchers will soon follow their lead on this dimension.

Evaluating models of concept formation

The two papers in this issue also share a common theme in their attempt to carefully evaluate concept learning behavior. Lebowitz examines the effect of varying UNIMEM's parameters on the quality of the resulting concept hierarchies. Using the domain of universities, he shows how two parameters – the percentage of features needed to create concepts and the percentage needed to retain them – affect the quality of the resulting hierarchies, using a number of dependent measures. These empirical studies help clarify the inner workings of UNIMEM.

Such a parametric analysis would be inappropriate for COBWEB, since that system has no parameters. Instead, Fisher emphasizes that the knowledge acquired during concept formation can be used for *prediction*, and he uses this fact to evaluate COBWEB's behavior. Using the domain of soybean diseases, he measures his system's ability to make accurate predictions as a function of two variables – the number of instances encountered and the dependence between features. As one would expect, predictive ability improves as more instances have been seen and as one increases the correlations between features.

Both Lebowitz and Fisher test their systems on real-world tasks, but despite their attractions, such natural domains have drawbacks as well. For instance, one has no control over the characteristics of the domain, nor can one easily determine whether the optimal solution has been found. Thus Fisher turns to artificial tasks, studying the effect of domain regularity on whether COBWEB finds the optimal clustering and on its time to convergence. The use of artificial domains lets him systematically vary task characteristics in a way that is impossible for real-world domains.

Finally, both authors show a concern for the time required to incorporate new instances into memory. Lebowitz shows empirically that UNIMEM's time for incorporation increases only gradually with the number of instances encountered. Fisher shows analytically that the update cost for COBWEB increases only logarithmically with the number of instances. Thus, the papers arrive at similar results using two different approaches.

In summary, the two papers in this issue are innovative along two important dimensions. First, they question some assumptions commonly held in the machine learning community and then present running systems that are based on alternative assumptions. Second, they support their novel approaches with extensive empirical studies that clarify the nature of the learning algorithms and the conditions under which they are useful. Fisher and Lebowitz each make significant additions to our understanding of concept formation, but their contribution extends beyond this class of problems. We encourage other researchers to imitate their examples by proposing novel methods and by supporting them with careful experiments and analyses. Therein lies the path to a true science of machine learning.

Pat Langley
University of California, Irvine
Langley@CIP.UCI.EDU

References

Feigenbaum, E. A. (1963). The simulation of verbal learning behavior. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill.

Hanson, S. J., & Bauer, M. (1986). Conceptual clustering, semantic organization and polymorphy. *Proceedings of the International Meeting on Advances in Learning* (pp. 53–77). Les Arc, France.

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

Michalski, R. S. (1983). A theory and methodology of learning from examples. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach*. Los Altos, CA: Morgan Kaufmann.

Michalski, R. S. (1987). How to learn imprecise concepts: A method for employing a two-tiered knowledge representation in learning. In *Proceedings of the Fourth International Workshop on Machine Learning* (pp. 50–58). Irvine, CA: Morgan Kaufmann.

Quinlan, J. R. (1987). Decision trees as probabilistic classifiers. In *Proceedings of the Fourth International Workshop on Machine Learning* (pp. 31–37). Irvine, CA: Morgan Kaufmann.