

Editorial: Human and Machine Learning

The goals of machine learning

One can identify a number of different themes within the machine learning community, each corresponding to central goals of its parent field, artificial intelligence. For instance, many AI researchers are concerned with implementing knowledge-intensive systems, but these often take man-years to construct. Machine learning may provide methods for automating this process, promising considerable savings in time and effort. Similarly, many AI researchers view artificial intelligence as a scientific discipline rather than an engineering one, and hope to formulate general principles of intelligent behavior that hold across a variety of domains. Since machine learning focuses on the acquisition of domain-specific knowledge rather than the knowledge itself, it holds considerable potential for such general principles.

Still other AI researchers are concerned with explaining human behavior, viewing AI techniques as an ideal tool for stating complex theories of the human information processing system. Since learning is a central phenomenon in human cognition, these researchers evaluate machine learning methods in terms of their ability to explain human learning. Although this is a minority opinion within machine learning, below we present some reasons why more members of the field should take this approach seriously.

Science as search

One of the central insights of AI is that intelligence involves search, and that effective search is constrained by domain-specific knowledge. This framework can be applied to problem solving, language understanding, and learning from experience. One can even apply this search metaphor to machine learning as a field of scientific study. In this framework, machine learning researchers are exploring a vast space of possible learning methods, searching for techniques with useful characteristics and looking for relations between these methods.

Expanding on this analogy, any domain-specific constraints that we can place on this search should ease our task considerably. In particular, our knowledge of

human learning provides an important set of constraints on our models of learning. Some of this knowledge is limited to specific areas, but other aspects (such as the incremental nature of human learning) hold across domains. We will not argue that all machine learning research should attempt to model human behavior, but we feel the field as a whole would benefit by paying more attention to human learning phenomena. We will see that, in many cases, there already exist empirical laws of human learning that are waiting for process explanations. In turn, these results can guide our search through the space of possible learning mechanisms.

Below we consider some examples of successful machine learning models of human behavior. At least two senses of ‘success’ are relevant to our discussion. The most obvious is that an AI system may account for well-established phenomena in human learning, making it a useful computational theory. The second sense is more subtle — the methods incorporated in the program may lead to other interesting AI systems, whether or not they model human learning. We will see examples of both types of success.

The elementary perceiver and memorizer

One of the earliest AI systems was Feigenbaum’s (1963) EPAM, a model of human behavior on verbal learning and memory tasks. Experimental psychologists had long studied human memory using nonsense syllables, and by the 1950s they had converged on a number of robust empirical generalizations. These laws summarized behavior in both paired associate tasks and serial learning tasks, and included regularities such as the serial position effect and selective forgetting.

EPAM modeled these phenomena, using a *discrimination network* to represent associations and using the processes of *discrimination* and *familiarization* to modify these networks. One might argue that experiments with nonsense syllables tell us little about the human information processing system, since our memories handle such abstract input only rarely. Despite this criticism, EPAM still stands out as the first computational model that accounted for a large body of psychological data. As such, it serves as a useful prototype for future attempts to model human learning and memory.

However, EPAM’s effect on machine learning goes beyond its success as an initial psychological model. The system’s use of discrimination nets influenced Hunt, Marin, and Stone’s (1966) design of their CLS system, and this in turn led to Quinlan’s (1986) well-known ID3 system for learning from examples. Feigenbaum’s early work on EPAM has also impacted recent work on incremental models of concept formation and conceptual clustering. For instance, both Lebowitz (1982) and Kolodner (1983) represent memory for concepts as a sophisticated discrimination network, and employ extensions of EPAM’s methods to acquire these memory structures.

Production system models of learning

Research on production system models of human cognition provides another success story. Newell (1967) originally proposed production systems as a model of the human cognitive architecture, and listed as one attraction their ability to account for the incremental nature of human learning. In fact, one of the earliest implemented production system programs was Waterman's (1970) poker player, which learned from experience. After this initial work, most research on production systems focused on performance until the late 1970s, when learning again emerged as a central topic in AI and cognitive psychology.

Anderson and his colleagues used the ACT production system architecture to model a variety of learning phenomena. In particular, Anderson (1982) has described the process of *knowledge compilation*, which consists of transforming declarative knowledge (propositions) into procedural form (production rules) and gradually combining these rules into larger chunks. Knowledge compilation accounts for general speedup effects observed in human learning. Anderson's theory also explains why sufficient practice eliminates the effects of set size in short-term memory experiments, and how *Einstellung* arises from problem solving experience.

Other researchers (Anzai & Simon, 1979; Langley, 1982; Ohlsson, 1983) have also used production systems to model human learning. However, the 'adaptive' production system approach has also migrated outside the realm of cognitive simulation. In fact, nearly all research on the task of learning search heuristics (Mitchell, Utgoff, & Banerji, 1983; Langley, 1985; Laird, Rosenbloom, & Newell, 1986; Porter & Kibler, 1986) has been carried out within the production system paradigm. Although few of these researchers were concerned with human behavior, their work clearly benefited from earlier attempts to use production systems as psychological models of learning.

Open areas in computational models of learning

Although machine learning research has contributed significantly to our understanding of human learning (and vice versa), much more remains to be done. Let us consider some areas that seem ripe for computational theories of the learning process. In each case, a number of solid empirical results exist, but few process models have been proposed that address these results.

Perhaps the most obvious area is first language acquisition. Experimental work and diary studies have converged on a common picture of the major stages that children traverse in learning their first language. Children begin their linguistic careers saying one word at a time, then move into a multi-word stage in which function words like *the* and *ing* are absent. They gradually master the nuances of these function words and gradually recover from overgeneralizations such as *foots* and *runned*. Only later do children acquire more complex forms like negation and

passive constructions.

Although language acquisition has long been a major topic within machine learning, only a few researchers have attempted to model first language learning in humans (Selfridge, 1981; Langley, 1982; Hill, 1983). Undoubtedly, one problem is that language development is a complex field in its own right. However, this field would benefit greatly from computational models of the language learning process, and machine learning would benefit in turn from careful attempts to account for the child language data.

Unlike language acquisition, the domain of human motor learning has received almost no attention within machine learning. However, this is another area in which significant empirical results exist with no computational theory to explain them. For instance, motor learning appears to involve a gradual shift from slower, feedback-based 'closed loop' processing to faster 'open loop' behavior in which little or no feedback is used. Also, there is considerable transfer between motor skills involving the same basic structure but differing in speed and force, but little transfer when the structure is altered. Methods from machine learning hold real potential for explaining these phenomena, providing a solid theoretical framework for future work on motor behavior. In turn, attempts in this direction might force us to develop new learning mechanisms, influencing machine learning research in other domains.

A third area involves the problem of concept formation. Although concept learning has been a mainstay of the machine learning community, most research in the area has ignored a number of well-established psychological phenomena. For instance, few human concepts can be defined in terms of necessary and sufficient conditions, and the evidence suggests that some form of prototype-based or probabilistic scheme is necessary. Also, *basic level* categories (Rosch & Mervis, 1975) appear to have a special status in human memory, being retrieved more quickly and being acquired earlier than other concepts. Finally, many human concepts seem to be defined functionally, in terms of the procedures in which they take part. Although some machine learning research has started to address these issues, we need more work on concept formation that is constrained by our knowledge of human concept learning. Such research would yield a better understanding of human concepts and their acquisition, but it should also lead to improved methods for nonhuman concept learning.

Summary

Although science can be characterized in terms of search, some search methods let one explore multiple paths in parallel. We have argued that more machine learning researchers should focus their efforts on modeling human behavior, but we have not argued that the field should limit itself to this approach. For those interested in general principles, the study of nonhuman learning methods is also necessary for useful results. In terms of applications, some of machine learning's greatest

achievements have involved nonincremental methods that are clearly poor models of human learning. Planes are terrible imitations of birds (and fly less efficiently), but there are still excellent reasons for using aircraft.

However, we do believe that too little research has focused on results from the literature on human learning, and that greater attention in this direction would benefit the field as a whole. Science is a complex and bewildering process, and the scientist should employ all available knowledge to direct his steps in useful directions. This strategy seems especially important in young fields like machine learning, in which conflicting views and methods abound. We encourage the reader to join us in applying machine learning techniques to explain the mysteries of human behavior, and in using knowledge of human behavior to constrain our computational theories of learning.

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

References

- Anzai, Y., & Simon, H.A. (1979). The theory of learning by doing. *Psychological Review*, 86, 124–140.
- Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369–406.
- Feigenbaum, E.A. (1963). The simulation of verbal learning behavior. In E.A. Feigenbaum & J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill.
- Hill, J.A.C. (1983). *A computational model of language acquisition in the two-year-old*. Doctoral dissertation, Department of Computer Science, University of Massachusetts, Amherst.
- Hunt, E.B., Marin, J., & Stone, P. (1966). *Experiments in induction*. New York: Academic Press.
- Kolodner, J.L. (1983). Maintaining organization in a dynamic long-term memory. *Cognitive Science*, 7, 243–280.
- Laird, J.E., Rosenbloom, P.S., & Newell, A. (1986). SOAR: The anatomy of a general learning mechanism. *Machine Learning*, 1, 11–46.
- Langley, P. (1982). Language acquisition through error recovery. *Cognition and Brain Theory*, 5, 211–255.
- Langley, P. (1985). Learning to search: From weak methods to domain-specific heuristics. *Cognitive Science*, 9, 217–260.
- Lebowitz, M. (1982). Correcting erroneous generalizations. *Cognition and Brain Theory*, 5, 367–381.
- Mitchell, T.M., Utgoff, P., & Banerji, R.B. (1983). Learning problem solving heuristics by experimentation. In R.S. Michalski, J.G. Carbonell, & T.M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach*. Palo Alto, CA: Tioga.
- Newell, A. (1967). *Studies in problem solving: Subject 3 on the crypt–arithmetic task Donald + Gerald = Robert* (Technical Report). Center for the Study of Information Processing, Carnegie Institute of Technology, Pittsburgh, PA.
- Ohlsson, S. (1983). A constrained mechanism for procedural learning. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence* (pp. 426–428). Karlsruhe, FRG: Morgan-Kaufmann.
- Porter, B.W., & Kibler, D.F. (1986). Experimental goal regression: A method for learning problem solving heuristics. *Machine Learning*, 1, 249–286.
- Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning*, 1, 81–106.
- Rosch, E., & Mervis, C.B. (1975). Family resemblance studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.
- Selfridge, M. (1981). A computer model of child language acquisition. *Proceedings of the Seventh*

- International Joint Conference on Artificial Intelligence* (pp. 92–96). Vancouver, BC, Canada: Morgan-Kaufmann.
- Waterman, D.A. (1970). Generalization learning techniques for automating the learning of heuristics. *Artificial Intelligence, 1*, 121–170.