The Changing Science of Machine Learning

Pat Langley (langley@asu.edu)
Computer Science and Engineering, Arizona State University, P.O. Box 87-8809, Tempe, AZ 85287 USA

Machine Learning saw its first issues appear in 1986. Naturally, the publication did not spring fully grown from its founders' brows; we laid the first plans in 1983, during the Second International Workshop on Machine Learning at Allerton House, Illinois. My partners in crime were Ryszard Michalski, Jaime Carbonell, and Tom Mitchell, who had organized the first two workshops and who had edited the book based on the initial event (Michalski, Carbonell, & Mitchell, 1983).

The reasons for launching a new journal were familiar ones. The first workshop in 1980 at Carnegie Mellon University had identified a community of researchers with common interests in computational approaches to learning and arrived at a name for its activities. Moreover, the parent fields of artificial intelligence and cognitive science were showing little interest at the time in learning-related issues, preferring to focus on the role of knowledge in intelligence, regardless of its origin. As a result, we encountered some difficulty publishing and, more generally, felt we were not getting the attention we deserved. Finally, there was the common urge of young, energetic researchers to create something of their own to which they could attach their names. At least that was part of my own motivation; I cannot speak for my co-founders.

Our group agreed that the new field needed a journal, and we also decided that I would be the first Executive Editor, with the other three serving as Editors. In 1984, I moved from Carnegie-Mellon University to the University of California, Irvine, where I received department support for the enterprise and set up an editorial office. We began to explore publishers, eventually settled on Kluwer, and started to solicit submissions. We made many decisions in those early days, from selecting an initial editorial board to determining the publication's size and format[1] that I cannot detail here.

More importantly, I began to think about what it would mean for an article to make a genuine contribution to machine learning. Between 1984 and 1990, when I turned over the executive editorship to Jaime Carbonell, I formed a number of opinions about these issues, some of which readers will find in editorials (e.g., Langley, 1986, 1987, 1989) contained in the journal's early volumes. One requirement was that

[1] I recommended that the journal have a green cover, but I did *not* propose the pea green shade that Kluwer ultimately selected.

authors should specify their learning task and their learning method clearly enough to allow replication, with pseudocode being a useful means to that end. I also worked with authors, sometimes more than they desired, to make their articles transparent to readers.

In the remainder of this essay, I will focus on two key changes I encouraged during my editorship that I believe were beneficial to the field and on some unexpected side effects that I feel have been far less desirable. This will involve contrasting the state of machine learning in the mid-1980s with that in later periods, including the present. I will assume that many readers have little knowledge of the field's origins or how it reached its current state, and I hope they will find my observations about its history informative and thought provoking.

First, early research on machine learning adopted an informal approach to evaluation. Papers typically reported runs of learning methods on a small set of training cases or problems, the outputs of these runs, and arguments for why the latter were reasonable or desirable results. This was appropriate in the field's initial days, when it was still demonstrating that machine learning was possible in any form. However, experience with reviewing submissions to the journal led a few of us to believe it was time for more formal approaches to evaluation. My discussions on this topic with Dennis Kibler, who was also at UCI, were especially productive.

The central idea, already proposed by Simon (1983), was that the purpose of learning is to improve performance on some class of tasks. In this view, it made little sense to describe a learning method in the absence of some performance system that utilized the acquired knowledge. Different performance measures would be appropriate for distinct tasks (e.g., accuracy for classification or efficiency for problem solving), but it seemed that such metrics were always possible and that many claims about learning could be linked to them.

Of course, one could approach this idea in different ways. Some formal treatments were already available, but they typically focused on worst-case analyses. Another option was to compare a system's error rates or reaction times to those obtained in psychological studies, but most researchers were interested in demonstrating functionality rather than in fitting such data. Nevertheless, the empirical methods used in psychology suggested ways to design and run controlled experiments with learning systems. Kibler and Langley (1988) laid out a framework for such an experimental science of machine learning, including examples from the emerging literature in this area. Many authors adopted this perspective and, within a few years, the vast majority of *Machine Learning* articles reported experimental results about performance improvement on well-defined tasks.

The experimental movement was aided by another development. David Aha, then a PhD student at UCI, began to collect data sets for use in empirical studies of machine learning. This grew into the UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/), which he made available to the community by FTP in 1987. This was rapidly adopted by many researchers because it was easy to use and because it let them compare their results to previous findings on the same tasks. The Repository focused on supervised concept learning for classification, but this was already a topic of widespread interest and it proved attractive to many in the field.

The early movement in machine learning was also characterized by an emphasis on 'symbolic' representations of learned knowledge, such as production rules, decision trees, and logical formulae. In fact, the original call for papers to *Machine Learning*, which I authored, explicitly welcomed submissions on these topics but discouraged papers on neural networks and other 'nonsymbolic' approaches. This bias was understandable in that the machine learning movement was an outgrowth of symbolic AI and cognitive science, with most of these researchers being concerned with automatically constructing expert systems or modeling human acquisition of knowledge structures.

However, the new emphasis on performance improvement suggested that this definition of the field was too narrow. In principle, it seemed that machine learning should include the study of any methods that improved performance with experience. For this reason, by the late 1980s the journal had begun to consider and publish papers on other approaches to learning, some of them incorporating ideas from the pattern-recognition community, which had been exploring the topic for over two decades. Thus, early volumes of *Machine Learning* included articles on learning with probabilistic and instance-based representations. This increase in breadth was a healthy development for the field.

Yet the two changes I have described also led the machine learning community, gradually and unexpectedly, in some directions that I believe have produced harmful effects over the long term. One such effect was an increased emphasis on classification and regression tasks, as opposed to more complex tasks like reasoning, problem solving, and language understanding that had played important roles earlier. This shift occurred partly because it was far easier to carry out experiments in classification and regression domains, and the situation was compounded by the UCI Repository's focus on such tasks, which made cross-domain studies in this area straightforward and commonplace.

Another factor was that methods borrowed from the pattern-recognition community lent themselves to these simpler tasks. Ironically, the increased breadth in representational paradigms led to a narrowing of the problems addressed by the field. Like the first edited book (Michalski et al., 1983), the early volumes of *Machine Learning* included a variety of papers on problem solving, reasoning, and language, but by the mid-1990s they had nearly disappeared from the literature, having been displaced by work on less audacious tasks.

A related trend involved a shift away from the field's original concern with learning in the context of intelligent systems. Many early programs embedded learning methods within sophisticated systems that carried out multi-step reasoning, heuristic problem solving, language understanding, or other complex cognitive activities (e.g., DeJong & Mooney, 1986; Laird, Rosenbloom, & Newell, 1986). The new emphasis on classification led to an increased focus on component learning algorithms that were studied in isolation from larger-scale intelligent systems. The latter invariably included some performance element, but it was often quite simple[2] and typically borrowed from earlier work.

The first three volumes of *Machine Learning* saw the beginnings of this change, but in the 1990s it became one of the field's dominant themes. In 1986, when we launched the journal, machine learning was still viewed as a branch of artificial intelligence. By 2000, many researchers committed to machine learning treated it as a separate field with few links to its parent discipline. There are now active PhD-level researchers who have never taken a course in artificial intelligence and who see no reason why they should, as their interests lie in completely different areas. I think this change has been a great loss to both fields.

A third shift concerned the representation of learning methods' inputs and outputs. Early work had emphasized relational encodings of experience and learned knowledge, even for relatively simple tasks like classification, but especially for reasoning, problem solving, and language processing. These were well supported by the logical and rule-based formalisms that were common in initial research efforts. However, the late 1980s and early 1990s saw an increasing number of papers that assumed attribute-value or propositional representations. This change was influenced by the growing interest in pattern-recognition techniques, but it was further reinforced by the UCI Repository, which emphasized attribute-value representations of training cases. The commercial success of many data-mining applications, which almost invariably used attribute-value schemes, was another contributing factor.

The early 1990s saw a countermovement, inductive logic programming, that emphasized learning with logical and other relational formalisms. This paradigm has continued and articles still appear regu-

[2] Work on reinforcement learning, although concerned with agents that operate over time, also incorporated this bias by assuming simple reactive control schemes.

larly in the journal, but it has remained a distinct community with limited impact on the broader machine learning discipline. Recent research on 'statistical relational learning', which combines ideas from inductive logic programming with statistical approaches, has been more influential, but the majority of work in the discipline remains focused on attribute-value notations. Given the importance of relations in human learning, I believe the field has become more impoverished as a result.

A final change concerned the role of knowledge in machine learning. Because of its connections with knowledge-based systems in AI and the study of expertise in cognitive science, early work in the field assumed that the product of learning was knowledge stated in some explicit form. Another major theme of research during the 1980s was the use of background knowledge as input to guide and constrain the learning process. The explanation-based learning movement (e.g., DeJong & Mooney, 1986) was the most visible example, but work on theory revision (e.g., Richards & Mooney, 1995) also adopted this position. These paradigms were concerned with learning from relatively few training cases, thus obtaining rates of learning comparable to those in humans.

During the 1990s, a number of factors led to decreased interest in the role of knowledge. One was the growing use of statistical and pattern-recognition approaches, which improved performance but which did not produce knowledge in any generally recognized form. The increasing reliance on experimental evaluation that revolved around performance metrics meant there was no evolutionary pressure to study knowledge-generating mechanisms. At the same time, the advent of large data sets convinced many that learning rate was not an important issue, reducing interest in using background knowledge to improve it. In some circles, learning from very large data sets became almost a fetish, marginalizing work on learning rapidly from few experiences.

The discipline of machine learning has seen other changes from its early days. In recent years, there has been increased emphasis on mathematical formalization and a bias against papers that do not include such treatments. Also, although the field retains a strong empirical component, few articles report the type of controlled experiments that Kibler and I proposed originally, with most studies involving 'bake offs', that is, mindless comparisons among the performance of algorithms that reveal little about the sources of power or the effects of domain characteristics. Both trends are outgrowths of those already discussed, but I do not have the space to examine their causes here.

I should clarify that I intend none of my comments as critiques of any particular research efforts or even paradigms for approaching machine learning. The incorporation of ideas from pattern recognition and statistics has enriched the field, and the introduction of experiments that measure performance improvement has placed it on a firm foundation. But my observations do imply a critique of the discipline as a whole. Machine learning was originally concerned with developing intelligent systems that exhibited rich behavior on complex tasks, while many modern researchers seem content to tackle problems that do not require either intelligence or systems. Machine learning focused initially on using and acquiring knowledge cast as rich relational structures, while many researchers now appear to care only about statistics.

I do not believe that we should abandon any of the computational advances that have occurred in the 25 years since *Machine Learning* published its first papers. Each has been a valuable contribution to our understanding of learning. However, I think it is equally important that we not abandon the many insights revealed during the field's early period, which remain as valid today as when they initially came to light. The challenge for machine learning is to recover the discipline's original breadth of vision and its audacity to develop learning mechanisms that cover the full range of abilities observed in humans – who remain our only example of truly intelligent systems.

References

DeJong, G., & Mooney, R. (1986). Explanation-based learning: An alternative view. *Machine Learning*, *1*, 145–176.

Kibler, D., & Langley, P. (1988). Machine learning as an experimental science. *Proceedings of the Third European Working Session on Learning* (pp. 81–92). Glasgow: Pittman.

Laird, J. E., Rosenbloom, P. S., & Newell, A. (1986). Chunking in SOAR: The anatomy of a general learning mechanism. *Machine Learning*, *1*, 11–46.

Langley, P. (1986). Human and machine learning. *Machine Learning*, *1*, 243–248.

Langley, P. (1987). Research papers in machine learning. *Machine Learning*, *2*, 195–198.

Langley, P. (1989). Toward a unified science of machine learning. *Machine Learning*, *3*, 253–259.

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (1983). *Machine learning: An artificial intelligence approach*. San Mateo, CA: Morgan Kaufmann.

Richards, B. L., & Mooney, R. J. (1995). Automated refinement of first-order horn-clause domain theories. *Machine Learning*, *19*, 95–131.

Simon, H. A. (1983). Why should machines learn? In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach*. San Mateo, CA: Morgan Kaufmann.