

Machine Learning: A Maturing Field

With this volume I complete my four-year term as executive editor of *Machine Learning*, and Tom Dietterich, who has been co-executive editor with me recently, takes over the helm—or starts serving his sentence, depending upon one's point of view. Let me take this opportunity to make a few reflections about the state of the field; past, present and future, based on personal observations.

A decade ago machine learning was regrouping from the rather uneventful 1970s. The first machine learning workshop was held in 1980 at Carnegie Mellon University with some two dozen participants and photocopied preprints. Shortly thereafter we started preparing the first machine learning book, and I was in charge of finding a publication venue. However the title "Machine Learning" raised skeptical eyebrows in publishers. By "machine learning" did we not really mean learning *about* machines rather than learning *by* machines? Couldn't we think of something more scientific-sounding to call the book? And anyway hadn't Minsky and Papert debunked this learning nonsense? Since it proved difficult to explain the difference between linear perceptrons and symbolic learning to those publishers (who shall go unnamed, to protect the guilty), we approached Nils Nilsson and his Tioga Press. Nils embraced the project with foresight and enthusiasm, and the rest, as they say in the tired cliché, is history.

The uneventful '70s, and the struggling turn of the decade, were followed by the booming '80s. Machine learning became respectable: some would argue (this writer included) it even became scientifically sound. In the process, it also became much more popular. Machine learning conferences in the late '80s drew ten times as many participants as that first CMU workshop: IJCAI and AAAI have very well represented machine learning tracks; this journal was established (with several publishers bidding against each other) and expanded from four to six and then to eight issues per year; books on the subject are no longer a rarity; and even undergraduate artificial intelligence curricula routinely include machine learning.

What do the steady-as-you-go '90s have in store for machine learning? This question is best addressed if one extrapolates trends in the field of a somewhat more subtle nature than raw popularity. First, machine learning is not a monolithic field. Some areas such as genetic algorithms have had a life of their own from the very start, well before the machine learning boom. Other areas have evolved in the '80s in a somewhat independent manner: connectionist learning (more deeply rooted in neural networks than in symbolic machine learning), and theoretical ML (as represented in the COLT workshops) are clear exemplars. Still other areas are splitting from the mainstream. Decision-tree induction, especially the ID3 variety, is being successfully incorporated into a number of commercial products and systems, with much of the applied work not reported in the standard ML media. Even "main-line" ML is split into the larger inductive learning camp and the smaller analytic (EBL) learning paradigm.

Second, the criteria for success in machine learning vary across the subdisciplines, and this is not just the traditional theoretical vs. experimental distinction of proofs vs. empirical validation. For instance, the current trend in analytical learning (EBL, analogy, abstraction, etc.) is to incorporate the learning mechanism as part of a larger performance engine (e.g., a problem solver or planner) and measure improved performance of the entire system over time and with multiple problem sets and different variants of the learning method(s). Inductive learning, on the other hand, is not normally reported in the context of a large performance engine (except for work such as Langley's ICARUS). Instead, the standard Irvine data sets are used to determine percent accuracy of concept classification, without regard to performance on a larger external task. Still others argue that ML should concentrate more on exploratory research—it is too early to be constrained by precise measurements. And, there is always the Cognitive Science approach where psychological fidelity (or at least psychological plausibility) are more important evaluative criteria than algorithmic or system performance.

These separatist trends are counterbalanced by several factors. Some researchers combine different ML approaches into integrated systems (such as Mitchell's THEO containing neural nets, decision trees and EBL), whereas other researchers compare the performance of different learning methods on common tasks (Shavlik, for instance). The ML conferences and the journal seek to represent the entire field, with articles, special issues, guest presentations and encouragements to submit work that spans different areas of machine learning. To some degree this process is successful as researchers return from conferences, or read articles that open their eyes to other exciting approaches, with an invigorating effect on their own work.

What then does the future hold in store for ML? In brief, machine learning has become a "normal" scientific discipline, with all the trappings and trimmings: scientific methods, competing paradigms, evaluative criteria, a recognized archival publication, regular conferences, a place in the educational curriculum, and so on. For better or worse, it is headed for continued evolutionary progress in the standard mode. It is no longer a discipline seeking recognition. As such, it has lost a bit of its uncertainty and its excitement. Perhaps we should do more to put the fun back into machine learning—but by all means keep all of the progress and scientific methods in place. The \$64 question, of course, is just how to do that. Here is one way: let us craft some exciting challenges for machine learning, similar to that for man-powered flight, or the Fredkin prize for computer chess. However, instead of having one grand challenge for ML, I propose several different challenges that could be within reach of present-day researchers from different ML paradigms, borrowing liberally from discussions with my colleagues:

1. **MACHINE DISCOVERY:** The first machine learning program that proposes a new law explaining a phenomenon leading to a published refereed journal article in the discipline where the discovery was made. Publication must be primarily on the merits of the newly discovered law (or theory), not on the novelty of it having been discovered by machine.
2. **ML IN CHESS:** The first machine learning program that improves its playing strength by at least 500 USCF (or FIDE) rating points, starting from a class B or higher playing level. Both before and after ratings must be officially recognized.

3. **ML IN PLANNING.** The first machine learning program that improves planning performance over 100-fold in at least two disparate real-world domains (e.g., in medicine, transportation, manufacturing, space) as measured in a large suite of test problems drawn from each domain, all different from any training problems.
4. **ML IN FINANCE.** The first machine learning program that learns in some aspect of investment (e.g., the equities market), producing an investment strategy that earned the investors over one million dollars in a one-year period when the strategy was followed to the letter without human intervention. Running the program in a hypothetical investment program or on historical data does not count for the one million dollar test. The prize, if any, would go to the researcher(s) who developed the ML system, not to the investors who would be more than adequately rewarded by its use.
5. **ML IN LANGUAGE:** The first machine learning system that would outperform the best hand-built natural language processing program at a well-defined, well-studied, important NLP task. For instance, the IBM effort to develop machine translation by statistical learning techniques would qualify, if it succeeded at outperforming SYSTRAN—or the then best machine translation system—in a useful suite of translation tasks. Acoustical speech recognition does not qualify because statistical and neural net techniques already outperform hand-built speech recognition methods. The challenge is to spur new ML research, more than to recognize past achievements.
6. **MEDICINE:** The first machine learning program to produce a diagnostic and/or treatment system that outperforms all pre-existing hand-built systems and that is deployed and under regular use in at least two institutions to help treat (or diagnose) at least two different and serious maladies, and that is accepted generally by the medical community in the areas of specialization as valid treatment. “Fringe” medical practices—those not taught in major medical schools—are excluded.

These challenges should be revised at least once a decade, to raise the proverbial bar a notch or two by replacing those challenges already accomplished with new and more exciting ones. Finally, it would be nice to find philanthropic (or “scientropic”) benefactors to sponsor a prize fund for one or more of these challenges that strike their fancy.

Jaime Carbonell
Carnegie Mellon University