The Discipline of Machine Learning

Tom M. Mitchell

July 2006
CMU-ML-06-108

Machine Learning Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Over the past 50 years the study of Machine Learning has grown from the efforts of a handful of computer engineers exploring whether computers could learn to play games, and a field of Statistics that largely ignored computational considerations, to a broad discipline that has produced fundamental statistical-computational theories of learning processes, has designed learning algorithms that are routinely used in commercial systems for speech recognition, computer vision, and a variety of other tasks, and has spun off an industry in data mining to discover hidden regularities in the growing volumes of online data. This document provides a brief and personal view of the discipline that has emerged as Machine Learning, the fundamental questions it addresses, its relationship to other sciences and society, and where it might be headed.

Keywords: machine learning

1 Defining Questions

A scientific field is best defined by the central question it studies. The field of Machine Learning seeks to answer the question

> "How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?"

This question covers a broad range of learning tasks, such as how to design autonomous mobile robots that learn to navigate from their own experience, how to data mine historical medical records to learn which future patients will respond best to which treatments, and how to build search engines that automatically customize to their user's interests. To be more precise, we say that a machine *learns* with respect to a particular task T, performance metric P, and type of experience E, if the system reliably improves its performance P at task T, following experience E. Depending on how we specify T, P, and E, the learning task might also be called by names such as data mining, autonomous discovery, database updating, programming by example, etc.

Machine Learning is a natural outgrowth of the intersection of Computer Science and Statistics. We might say the defining question of Computer Science is "How can we build machines that solve problems, and which problems are inherently tractable/intractable?" The question that largely defines Statistics is "What can be inferred from data plus a set of modeling assumptions, with what reliability?" The defining question for Machine Learning builds on both, but it is a distinct question. Whereas Computer Science has focused primarily on how to manually program computers, Machine Learning focuses on the question of how to get computers to program themselves (from experience plus some initial structure). Whereas Statistics has focused primarily on what conclusions can be inferred from data, Machine Learning incorporates additional questions about what computational architectures and algorithms can be used to most effectively capture, store, index, retrieve and merge these data, how multiple learning subtasks can be orchestrated in a larger system, and questions of computational tractability.

A third field whose defining question is closely related to Machine Learning is the study of human and animal learning in Psychology, Neuroscience, and related fields. The questions of how computers can learn and how animals learn most probably have highly intertwined answers. To date, however, the insights Machine Learning has gained from studies of Human Learning are much weaker than those it has gained from Statistics and Computer Science, due primarily to the weak state of our understanding of Human Learning. Nevertheless, the synergy between studies of machine and human learning is growing, with machine learning algorithms such as temporal difference learning now being suggested as explanations for neural signals observed in learning animals. Over the coming years it is reasonable to expect the synergy between studies of Human Learning and Machine Learning to grow substantially, as they are close neighbors in the landscape of core scientific questions.

Other fields, from biology to ecomonics to control theory also have a core interest in the question of how systems can automatically adapt or optimize to their environment, and machine learning will likely have an increasing exchange of ideas with these fields over the coming years. For example, economics is interested in questions such as how distributed collections of self-interested individuals may form a system (market) that learns prices leading to pareto-optimal allocations for the greatest common good. And control theory, especially adaptive control theory, is interested in questions such as how a servo-control system can improve its control strategy through experience. Interestingly, the mathematical models for adaptation in these other fields are somewhat different from those commonly used in machine learning, suggesting significant potential for cross-fertilization of models and theories.

The following sections discuss the state of the art of Machine Learning, a sample of successful applications, and a sample of open research questions.

2 State of Machine Learning

Here we describe some of the progress in machine learning, as well as open research questions.

2.1 Application Successes

One measure of progress in Machine Learning is its significant real-world applications, such as those listed below. Although we now take many of these applications for granted, it is worth noting that as late as 1985 there were almost no commercial applications of machine learning.

- *Speech recognition.* Currently available commercial systems for speech recognition all use machine learning in one fashion or another to train the system to recognize speech. The reason is simple: the speech recognition accuracy is greater if one trains the system, than if one attempts to program it by hand. In fact, many commercial speech recognition systems involve two distinct learning phases: one before the software is shipped (training the general system in a speaker-independent fashion), and a second phase after the user purchases the software (to achieve greater accuracy by training in a speaker-dependent fashion).

- *Computer vision.* Many current vision systems, from face recognition systems, to systems that automatically classify microscope images of cells, are developed using machine learning, again because the resulting systems are more accurate than hand-crafted programs. One massive-scale application of computer vision trained using machine learning is its use by the US Post Office to automatically sort letters containing handwritten addresses. Over 85% of handwritten mail in the US is sorted automatically, using handwriting analysis software trained to very high accuracy using machine learning over a very large data set.

- *Bio-surveillance.* A variety of government efforts to detect and track disease outbreaks now use machine learning. For example, the RODS project involves real-time collection of admissions reports to emergency rooms across western Pennsylvania, and the use of machine learning software to learn the profile of typical admissions so that it can detect anomalous patterns of symptoms and their geographical distribution. Current work involves adding in a rich set of additional data, such as retail purchases of over-the-counter medicines to increase the information flow into the system, further increasing the need for automated learning methods given this even more complex data set.

- *Robot control.* Machine learning methods have been successfully used in a number of robot systems. For example, several researchers have demonstrated the use of machine learning to acquire control strategies for stable helicopter flight and helicopter aerobatics. The recent Darpa-sponsored competition involving a robot driving autonomously for over 100 miles in the desert was won by a robot that used machine learning to refine its ability to detect distant objects (training itself from self-collected data consisting of terrain seen initially in the distance, and seen later up close).

- *Accelerating empirical sciences.* Many data-intensive sciences now make use of machine learning methods to aid in the scientific discovery process. Machine learning is being used to learn models of gene expression in the cell from high-throughput data, to discover unusual astronomical objects

from massive data collected by the Sloan sky survey, and to characterize the complex patterns of brain activation that indicate different cognitive states of people in fMRI scanners. Machine learning methods are reshaping the practice of many data-intensive empirical sciences, and many of these sciences now hold workshops on machine learning as part of their field's conferences.

2.2 Place of Machine Learning within Computer Science

Given this sample of applications, what can we infer in general about the future role of machine learning in the field of computer applications? One way to think about this is to imagine the space of all software applications, and to recognize the above applications suggest a niche within this space where machine learning has a special role to play. In particular, machine learning methods are already the best methods available for developing particular types of software, in applications where:

- The application is too complex for people to manually design the algorithm. For example, software for sensor-base perception tasks, such as speech recognition and computer vision, fall into this category. All of us can easily label which photographs contain a picture of our mother, but none of us can write down an algorithm to perform this task. Here machine learning is the software development method of choice simply because it is relatively easy to collect labeled training data, and relatively ineffective to try writing down a successful algorithm.

- The application requires that the software customize to its operational environment after it is fielded. One example of this is speech recognition systems that customize to the user who purchases the software. Machine learning here provides the mechanism for adaptation. Software applications that customize to users are growing rapidly - e.g., bookstores that customize to your purchasing preferences, or email readers that customize to your particular definition of spam. This machine learning niche within the software world is growing rapidly.

Viewed this way, machine learning methods play a key role in the world of computer science, within an important and growing niche. While there will remain software applications where machine learning may never be useful (e.g., to write matrix multiplication programs), the niche where it will be used is growing rapidly as applications grow in complexity, as the demand grows for self-customizing software, as computers gain access to more data, and as we develop increasingly effective machine learning algorithms.

Beyond its obvious role as a method for software development, machine learning is also likely to help reshape our view of Computer Science more generally. By shifting the question from "how to program computers" to "how to allow them to program themselves," machine learning emphasizes the design of self-monitoring systems that self-diagnose and self-repair, and on approaches that model their users, and the take advantage of the steady stream of data flowing through the program rather than simply processing it. Similarly, Machine Learning will help reshape the field of Statistics, by bringing a computational perspective to the fore, and raising issues such as never-ending learning. Of course both Computer Science and Statistics will also help shape Machine Learning as they progress and provide new ideas to change the way we view learning.

2.3 Some Current Research Questions

As the above applications suggest, substantial progress has already been made in the development of machine learning algorithms and their underlying theory. For example, we now have a variety of algorithms for supervised learning of classification and regression functions; that is, for learning some initially unknown

function $f : X \rightarrow Y$ given a set of labeled training examples $\{\langle x_i, y_i \rangle\}$ of inputs x_i and outputs $y_i = f(x_i)$. For example, in training an image recognition program x_i may be a single image, and y_i the label of the object in the image. Algorithms from Support Vector Machines, to Bayesian classifiers, to Genetic Algorithms may be used to estimate the function f from the data. We also have a useful body of theory that helps characterize how accurately one should expect to learn the function f, depending on the number of labeled training examples available, assumptions about the nature of the data (e.g., whether the examples are drawn independently), and properties of the learning algorithm such as the complexity of the set of hypotheses it considers. Of course there are many other types of learning problems and associated algorithms and theories, including unsupervised clustering (e.g., cluster genes based on their time series expression patterns), anomaly detection (e.g., find unusual patterns of emergency room admissions), reinforcement learning (e.g., learn to pick good chess moves, where the only training data is the final win/lose outcome of the game after making many moves), data modeling (e.g., find a small set of factors that can be combined to reconstruct a sequence of high-dimensional brain images), etc.

The field is moving forward in many directions, exploring a variety of types of learning tasks, and developing a variety of underlying theory. Here is a sample of current research questions:

- *Can unlabeled data be helpful for supervised learning?* Supervised learning involves estimating some function $f : X \rightarrow Y$ given a set of labeled training examples $\{\langle x_i, y_i \rangle\}$. We could dramatically reduce the cost of supervised learning if we could make use of unlabeled data as well (e.g., images that are unlabeled). Are there situations where unlabeled data can be guaranteed to improve the expected learning accuracy? Interesting, the answer is yes, for several special cases of learning problems that satisfy additional assumptions. These include practical problems such as learning to classify web pages or spam. Exploration of new algorithms and new subclasses of problems where unlabeled data is provably useful is an active area of current research.

- *How can we transfer what is learned for one task to improve learning in other related tasks?.* Note the above formulation of supervised learning involves learning a single function f. In many practical problems we might like to learn a family of related functions (e.g., a diagnosis function for patients in New York hospitals, and one for patients in Tokyo hospitals). Although we expect the diagnosis function to be somewhat different in the two cases, we also expect some commonalities. Methods such as hierarchical Bayesian approaches provide one way to tackle this problem, by assuming the learning parameters of the NY function and the Tokyo function share similar prior probabilities, but allowing the data from each hospital to override these priors as appropriate. The situation becomes more subtle when the transfer between functions is more complex – e.g.., a robot learning both a next-state function and a function to chose control actions should be able to learn better by taking advantage of the logical relationship between these two types of learned information.

- *What is the relationship between different learning algorithms, and which should be used when?.* Many different learning algorithms have been proposed and evaluated experimentally in different application domains. One theme of research is to develop a theoretical understanding of the relationships among these algorithms, and of when it is appropriate to use each. For example, two algorithms for supervised learning, Logistic Regression and the Naive Bayes classifier, behave differently on many data sets, but can be proved to be equivalent when applied to certain types of data sets (i.e., when the modeling assumptions of Naive Bayes are satisfied, and as the number of training examples approaches infinity). This understanding suggests, for example, that Naive Bayes should be preferred if data is sparse but one is confident of the modeling assumptions. More generally, the theoretical

characterization of learning algorithms, their convergence properties, and their relative strengths and weaknesses remains a major research topic.

- *For learners that actively collect their own training data, what is the best strategy?* Imagine a mobile robot charged with the task of learning to find its master's slippers anywhere in the house, and imagine that it is allowed to practice during the day, by viewing the slippers from different viewpoints of its choice, and moving the slippers to different locations with different lighting conditions. What is the most efficient training strategy for actively collecting new data as its learning proceeds? A second example of this problem involves drug testing where one wishes to learn the drug effectiveness while minimizing the exposure of patients to possible unknown side effects. This is a part of a more broad research thrust into learning systems that take more active control over the learning setting, rather than passively using data collected by others.

- *To what degree can we have both data privacy and the benefits of data mining?* There are many beneficial uses of machine learning, such as training a medical diagnosis system on data from all hospitals in the world, which are not being pursued largely because of privacy considerations. Although at first it might seem that we must choose between privacy and the benefits of data mining, in fact we might be able to have both in some cases. For example, rather than forcing hospitals to sacrifice privacy and pass around their patient records to a central data repository, we might instead pass around a learning algorithm to the hospitals, allowing each to run it under certain restrictions, then pass it along to the next hospital. This is an active research area, building both on past statistical work on data disclosure and on more recent cryptographic approaches.

2.3.1 Longer Term Research Questions

The above research questions are already being energetically pursued by researchers in the field. It is also interesting to consider longer term research questions. Below are some additional research topics which I feel hold the potential to significantly change the face of machine learning over the coming decade.

- *Can we build never-ending learners?* The vast majority of machine learning work to date involves running programs on particular data sets, then putting the learner aside and using the result. In contrast, learning in humans and other animals is an ongoing process in which the agent learns many different capabilities, often in a sequenced curriculum, and uses these different learned facts and capabilities in a highly synergistic fashion. Why not build machine learners that learn in this same cumulative way, becoming increasingly competent rather than halting at some plateau? For example, a robot in the same office building for months or years should learn a variety of capabilities, starting with simpler tasks (e.g., how to recognize objects in that dark end of the hallway), to more complex problems that build on previous learning (e.g., where to look first to find the missing recycling container). Similarly, a program to learn to read the web might learn a graded set of capabilities beginning with simpler abilities such as learning to recognize names of people and places, and extending to extracting complex relational information spread across multiple sentences and web pages. A key research issue here is self-supervised learning and constructing an appropriate graded curriculum.

- *Can machine learning theories and algorithms help explain human learning?* Recently, theories and algorithms from machine learning have been found relevant to understanding aspects of human and animal learning. For example, reinforcement learning algorithms and theories predict surprisingly

well the neural activity of dopaminergic neurons in animals during reward-based learning. And machine learning algorithms for discovering sparse representations of naturally occurring images predict surprisingly well the types of visual features found in the early visual cortex of animals. However, theories of animal learning involve considerations that have not yet been considered in machine learning, such as the role of motivation, fear, urgency, forgetting, and learning over multiple time scales. There is a rich opportunity for cross fertilization here, an opportunity to develop a general theory of learning processes covering animals as well as machines, and potential implications for improved strategies for teaching students.

- *Can we design programming languages containing machine learning primitives?* Can a new generation of computer programming languages directly support writing programs that learn? In many current machine learning applications, standard machine learning algorithms are integrated with hand-coded software into a final application program. Why not design a new computer programming language that supports writing programs in which some subroutines are hand-coded while others are specified as "to be learned." Such a programming language could allow the programmer to declare the inputs and outputs of each "to be learned" subroutine, then select a learning algorithm from the primitives provided by the programming language. Interesting new research issues arise here, such as designing programming language constructs for declaring what training experience should be given to each "to be learned" subroutine, when, and with what safeguards against arbitrary changes to program behavior.

- *Will computer perception merge with machine learning?* Given the increasing use of machine learning for state-of-the-art computer vision, computer speech recognition, and other forms of computer perception, can we develop a general theory of perception grounded in learning processes? One intriguing opportunity here the incorporation of multiple sensory modalities (e.g., vision, sound, touch) to provide a setting in which self-supervised learning could be applied to predict one sensory experience from the others. Already researchers in developmental psychology and education have observed that learning can be more effective when people are provided multiple input modalities, and work on co-training methods from machine learning suggests the same.

2.4 Ethical Questions

Above are some of the problems that will shape the field of machine learning over the coming decade. While it is impossible to predict the future, further research in machine learning will almost certainly produce more powerful computer capabilities. This, in turn, will lead on occasion to ethical questions about where and when to apply the resulting technology. For example, consider that today's technology could enable discovering unanticipated side effects of new drugs, if it were applied to data describing all doctor visits and medical records in the country along with all purchases of drugs. Recent cases in which new drugs were recalled following a number of unanticipated patient deaths might well have been ameliorated by already available machine learning methods. However, applying this machine learning technology would also have impacted our personal privacy, as our medial records and drug purchases would have had to be captured and analyzed. Is this something we wish as a society to do? Personally, I believe there are good arguments on both sides, and that as a society we need to discuss and debate these questions in an open and informed fashion, then come to a decision. Related questions occur about collecting data for security and law enforcement, or for marketing purposes. Like all powerful technologies, machine learning will raise its share of questions about whether it should be used for particular purposes. Although the answer to each of

these questions will have a technical component, in some cases the question will also have a social policy component requiring all of us to become engaged in deciding its answer.

3 Where to Learn More

To find out more about Machine Learning, see the top conferences and journals in the field, including:

- *International Conference on Machine Learning (ICML).*
- *Conference on Neural Information Processing Systems (NIPS).*
- *Annual Conference on Learning Theory (COLT).*
- *Journal of Machine Learning Research (JMLR).* This top journal is freely available online at www.jmlr.org.
- *Machine Learning*. Published by Springer.

4 Acknowledgments

Many of the ideas presented here have arisen from discussions with others.

I would like to acknowledge many stimulating discussions with students and faculty of the Machine Learning Department at Carnegie Mellon University, for helping to shape my own view of the discipline of machine learning. I would also like to specifically thank Avrim Blum, Stephen Fienberg, Carlos Guestrin, Michael Jordan, and Andrew Ng for helpful comments on earlier drafts of this document.

It would be impossible to do our work without the generous support of funders. I am particularly grateful for research support from Darpa, NSF, NIH, the Keck Foundation, and Lockheed Martin Corporation.