INTERVIEW

National Science Review
5: 54–58, 2018
doi: 10.1093/nsr/nwx045
Advance access publication 2 May 2017

Special Topic: Machine Learning

Machine learning challenges and impact: an interview with Thomas Dietterich

By Zhi-Hua Zhou

Machine learning is the driving force of the hot artificial intelligence (AI) wave. In an interview with NSR, Prof. Thomas Dietterich, the distinguished professor emeritus of computer science at Oregon State University in the USA, the former president of Association of Advancement of Artificial Intelligence (AAAI, the most prestigious association in the field of artificial intelligence) and the founding president of the International Machine Learning Society, talked about exciting recent advances and technical challenges of machine learning, as well as its big impact on the world.

WHY MACHINE LEARNING IS IMPORTANT

NSR: Why machine learning is useful?
Dietterich: Machine learning provides a new method for creating high-performance software. In traditional software engineering, we talk with the users, formulate the requirements and then design, implement and test algorithms for achieving those requirements. With machine learning, we still formulate the overall goal of the software system, but instead of designing our own algorithms, we collect training examples (usually, by having people label data points) and then apply a machine learning algorithm to automatically learn the desired function.

This new methodology allows us to create software for many problems that we were not able to solve using previous software engineering methods. In particular, the performance of previous methods for visual object detection and recognition, speech recognition and language translation were not good enough to be usable. But with recent advances in machine learning, we now have systems that can perform these tasks with accuracy that matches human performance (more or less).

Machine learning is therefore providing a key technology to enable applications such as self-driving cars, real-time driving instructions, cross-language user interfaces and speech-enabled user interfaces. Machine learning is also valuable for web search engines, recommendation systems and personalized advertising. Many people predict that machine learning methods will lead to a revolution in medicine, particularly in the automatic collection and analysis of medical images. Machine learning is also a promising tool for many operational aspects of modern companies. For example, machine learning can help predict customer demand and optimize supply chains. It is also a key technology for training robots to perform flexible manufacturing tasks.

NSR: Why is machine learning important to the science community and to society?

Thomas Dietterich, professor at Oregon State University *(Courtesy of Thomas Dietterich).*

Dietterich: Machine learning methods can be helpful in data collection and analysis. For example, machine learning methods are applied to analyse the immense amount of data collected by the Large Hadron Collider, and machine learning techniques are

© The Author(s) 2017. Published by Oxford University Press on behalf of China Science Publishing & Media Ltd. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/nsr/article/5/1/54/3789514 by guest on 24 October 2020

> “With recent advances in machine learning, we now have systems that can [recognize objects in images, recognize speech, and translate languages] with accuracy that matches human performance.
>
> — Thomas Dietterich ”

critical to analysing astronomical data. Machine learning techniques can help scientists decide which data points to collect by helping design experiments. And robotic systems can then automatically perform those experiments either in the lab or in the real world. For example, there is an Automated Scientist developed by Ross King that designs, executes and analyses its own experiments. Ocean-going glider robots are controlled by AI systems. And machine learning techniques are starting to be applied to control drones that collect data in ecosystems and in cities.

My own research focuses on applying machine learning to improve our management of the earth's ecosystems. For example, in Oregon, we have frequent forest fires caused by lightning. These fires can destroy habitat for endangered species and burn up trees that could have been used to build houses. One cause of these large fires is that for many years, the USA suppressed every fire. This is very expensive, and it allows fuel to accumulate in the forests so that when a new fire is started, it burns very hot and is much more damaging. We are applying machine learning (reinforcement learning) methods to find good rules for deciding which fires should be suppressed and which fires should be permitted to burn. These rules save money and help preserve endangered species.

Machine learning methods can also be applied to map the location and population of endangered species such as the panda. In the USA, we have developed new machine learning methods for predicting and understanding bird migration. Similar problems arise in mapping the spread of new diseases, of air pollution and of traffic.

In business and finance, machine learning methods can help identify fraud and theft. My group has been studying algorithms for anomaly detection that can identify unusual transactions and present them to a human analyst for law enforcement.

Machine learning methods can also contribute to the development of 'Smart Cities'. I mentioned traffic management and pollution mapping. But machine learning techniques can also be applied to identify where new infrastructure is needed (e.g. water supply, electricity, internet). In the USA, machine learning has been applied to map the distribution of lead paint, which was used in older buildings in the 20th century and is a neurotoxin.

ABOUT DEEP LEARNING

NSR: Could you comment on the strength and weakness of deep learning?

Dietterich: The most exciting recent development is the wave of research on deep learning methods. Most machine learning methods require the data scientist to define a set of 'features' to describe each input. For example, in order to recognize an object in an image, the data scientist would first need to extract features such as edges, blobs and textured regions from the image. Then these could be fed to a machine learning algorithm to recognize objects. Deep learning allows us to feed the raw image (the pixels) to the learning algorithm without first defining and extracting features. We have discovered that deep learning can learn the right features, and that it does this much better than we were able to hand-code those features. So in problems where there is a big gap between the inputs (e.g. images, speech signals, etc.) and the outputs (e.g. objects, sentences, etc.), deep learning is able to do much better than previous machine learning methods.

However, there are still many problems where the features are easy to obtain. For example, in fraud detection, we might examine the number of credit card transactions, where and when they were initiated, and so on. These are already represented in high level features, and in such applications deep learning does not provide much, if any, benefit. Deep learning algorithms are also difficult to train and require large amounts of computer time, so in most problems, they are not the preferred method.

Deep learning is one particular method for machine learning. It is quite difficult to use, so in problems where features are available, it is generally much better to use methods such as random forests or boosted trees. These methods are very easy to use and require very little experience. They are also many orders of magnitude faster than deep learning methods, so they can run on a laptop or a smart phone instead of requiring a GPU supercomputer. An important goal of machine learning work is to make machine learning techniques usable by people with little or no formal training in machine learning. I am Chief Scientist of a company called BigML that has developed cloud-based machine learning services that are extremely easy to use. They are all based on decision tree methods (including boosting and random forests).

There are also interesting ways to combine deep learning with standard AI techniques. The best example is Alpha Go, which combines deep learning (to analyse the patterns of stones on the Go board) and Monte Carlo tree search (to search ahead into the future of the game to determine the consequences of alternate moves). Similarly, self-driving cars combine top-level software (for safety, control, and user interface) with deep learning methods for computer vision and activity recognition.

RESEARCH CHALLENGES FOR MACHINE LEARNING

NSR: Could you comment on the research challenges for machine learning?

Dietterich: There are many important research challenges for machine learning. The first challenge is to improve methods for unsupervised and reinforcement learning. Virtually all of the recent advances have been in so-called 'supervised learning',

Downloaded from https://academic.oup.com/nsr/article/5/1/54/3789514 by guest on 24 October 2020

> The most exciting recent development is the wave of research on deep learning methods.
>
> —Thomas Dietterich

where a 'teacher' tells the computer the right answer for each training example. However, there are many problems where we lack teachers but where we have huge amounts of data. One example is when we seek to detect anomalies or fraudulent transactions. There is work in developing 'anomaly detection' algorithms that can learn from such data without the need of a teacher. More generally, there are many classes of 'unsupervised' learning algorithms that can learn without a teacher.

Another area where more research is needed is in reinforcement learning. Reinforcement learning involves teaching a computer to perform a complex task by giving it rewards or punishments. In many problems, the computer can compute the reward itself, and this allows the computer to learn by trial and error rather than from a teacher's examples. Reinforcement learning is particularly valuable in control problems (such as self-driving cars, robots and the fire management problem I mentioned before). Methods for reinforcement learning are still very slow and difficult to apply, so researchers are attempting to find ways of speeding them up. Existing reinforcement learning algorithms also operate at a single time scale, and this makes it difficult for these methods to learn in problems that involve very different time scales. For example, the reinforcement learning algorithm that learns to drive a car by keeping it within the traffic lane cannot also learn to plan routes from one location to another, because these decisions occur at very different time scales. Research in hierarchical reinforcement learning is attempting to address this problem.

The second major research problem for machine learning is the problem of verification, validation and trust. Traditional software systems often contain bugs, but because software engineers can read the program code, they can design good tests to check that the software is working correctly. But the result of machine learning is a 'black box' system that accepts inputs and produces outputs but is difficult to inspect. Hence, a very active topic in machine learning research is to develop methods for making machine learning systems more interpretable (e.g. by providing explanations or translating their results into easy-to-understand forms). There is also research on automated methods for verification and validation of black box systems. One of the most interesting new directions is to create automated 'adversaries' that attempt to break the machine learning system. These can often discover inputs that cause the learned program to fail.

A related area of research is 'robust machine learning'. We seek machine learning algorithms that work well even when their assumptions are violated. The biggest assumption in machine learning is that the training data are assumed to be independently distributed and to be a representative example of the future input to the system. Several researchers are exploring ways of making machine learning systems more robust to failures of this assumption.

The third major challenge for machine learning is the question of bias. There are often biases in the way that data are collected. For example, experiments on the effectiveness of new drugs may be performed only on men. A machine learning system might then learn that the drugs are only effective for people older in 35 years. But in women, the effectiveness might be completely different. In a company, data might be collected from current customers, but these data might not be useful for predicting how new customers will behave, because the new customers might be different in some important way (younger, more internet-savvy, etc.). Current research is developing methods for detecting such biases and for creating learning algorithms that can recover from these biases.

ABOUT THE THREAT OF ADVANCED AI

NSR: With the rapid progress of machine learning, will human jobs be threatened by machines? Could you comment on the 'singularity theory' and the arguments about the risks of advanced AI?

Dietterich: Like all new technologies, machine learning is definitely going to change the job market. Jobs that involve simple repetitive activity—whether it is repeated physical actions (like factory work and truck driving) or repeated intellectual actions (like much work in law, accounting, and medicine)—will likely be at least partially replaced by software and robots. As with the Industrial Revolution, there is likely to be a large disruption in the economy as these new technologies are developed. The important question is whether machine learning and AI will also create new kinds of jobs. This also occurred during the Industrial Revolution, and I think it will happen again in the AI revolution. It is hard to predict what these jobs will be.

I think about what happened when the internet was developed. I was a graduate student in the early 1980s when the Internet Protocols were developed and deployed. They were designed to make it easy to move files from one computer to another and to log in to remote computers from local computers. We had no idea about the world wide web, search engines, electronic commerce or social networks! This means we also did not predict the new jobs that resulted (web page designers, user experience engineers, digital advertising, recommendation system designers, cyber security engineers and so on).

I think it is similarly very difficult today to predict what the jobs of the future will be. There will certainly be jobs involved in creating AI systems, teaching them, customizing them and repairing them. I suspect that it will not be cost-effective to completely automate most existing jobs. Instead, maybe 80% of each job will be automated, but a human will need to do the remaining 20%. That human thereby becomes much more valuable and will be paid well.

One aspect of many human jobs that I believe will be very difficult to automate is empathy. Robots and AI systems will have very different experiences than people. Unlike people, they will not be able to 'put themselves into a person's shoes' in order to

Downloaded from https://academic.oup.com/nsr/article/5/1/54/3789514 by guest on 24 October 2020

understand and empathize with humans. Instead, they will need to be taught, like aliens or like Commander Data in Star Trek, to predict and understand human emotions. In contrast, people are naturally able to do these things, because we all know 'what it feels like' to be human. So jobs that involve empathy (e.g. counseling, coaching, management, customer service) are least likely to be satisfactorily automated. This will be particularly true if human customers place a value on 'authentic human interaction' rather than accepting an interaction with a robot or automated system.

If most industrial and agricultural production becomes automated—and if the resulting wealth is evenly distributed throughout society—then humans are likely to find other things to do with their time. Traveling to visit other countries and other cultures will likely become even more popular than it is today. Sports, games, music and the arts may also become much more popular. One hundred years ago, it was hard to get a massage or a pedicure. Now these are available almost everywhere. Who knows what people will want to do and what experiences they will want to have 100 years from now?

There are two different popular notions of the 'singularity'. Let me talk about each of them. One notion, exemplified by the writings of Ray Kurzweil, is that because of the exponential improvement of many technologies, it is difficult for us to see very far into the future. This is the idea of a 'technological singularity'. A true mathematical singularity would be the point at which technology improves infinitely quickly. But this is impossible, because there are limits to all technologies (although we don't know what they are). There is a famous law in economics due to Herbert Stein: 'If something can't go on forever, it won't.' This is true for Moore's Law, and it is true for all AI technologies. However, even if a true mathematical singularity is impossible, we are currently experiencing exponential growth in the capabilities of AI systems, so their future capabilities will be very different from their current capabilities, and standard extrapolation is impossible. So I believe Kurzweil is correct that we cannot see very far into this exponentially-changing future.

There is a second notion of 'singularity' that refers to the rise of so-called superintelligence. The argument—first put forth by I.J. Good in an article in 1965—is that at some point AI technology will cross a threshold where it will be able to improve itself recursively and then it will very rapidly improve and become exponentially smarter than people. It will be the 'last invention' of humanity. Often, the threshold is assumed to be 'human-level AI', where the AI system matches human intelligence. I am not convinced by this argument for several reasons. First, the whole goal of machine learning is to create computer systems that can learn autonomously. This is a form of self-improvement (and often it is applied to improve the learning system itself and hence is recursive self-improvement). However, such systems have never been able to improve themselves beyond one iteration. That is, they improve themselves, but then the resulting system is not able to improve itself. I believe the reason for this is that we formulate the problem as a problem of function optimization, and once you have found the optimal value of that function, further optimization cannot improve it, by definition. To maintain exponential improvements, every technology requires repeated breakthroughs. Moore's Law for example is not a single process, but actually a staircase of improvements where each 'step' involved a different breakthrough. I believe this leads us back to the Kurzweil-type technological singularity rather than to superintelligence.

Second, it is very suspicious that the arguments about superintelligence set the threshold to match human intelligence. This strikes me as the same error that was exposed by Copernicus and by Darwin. Humans are presumably not special in any way with respect to the intelligence that computers can attain. The limits we encounter are probably dictated by many factors including the size and computing power of our brains, the durations of our lives, and the fact that each one of us must learn on our own (rather than like parallel and distributed computers). Computers are already more intelligent than people on a wide range of tasks including job shop scheduling, route planning, control of aircraft, simulation of complex systems (e.g. the atmosphere), web search, memory, arithmetic, certain forms of theorem proving, and so on. But none of these super-human capabilities has led the kind of superintelligence described by Good.

Third, we observe in humans that intelligence tends to involve breadth rather than depth. A great physicist such as Stephen Hawking is much smarter than me about cosmology, but I am more knowledgeable than he is about machine learning. Furthermore, experiments have revealed that people who are experts in one aspect of human endeavor are no better than average in most other aspects. This suggests that the metaphor of intelligence as rungs on a ladder, which is the basis of the argument on recursive self-improvement, is the wrong metaphor. Instead, we should consider the metaphor of a liquid spreading across the surface of a table or the metaphor of biodiversity where each branch of knowledge fills a niche in the rain forest of human intelligence. This metaphor does not suggest that there is some threshold that, once exceeded, will lead to superintelligence.

The claim that Kurzweil's view of the singularity is the right one does not mean that AI technology is inherently safe and that we have nothing to worry about. Far from it. Indeed, as computers become more intelligent, we are motivated to put them in charge of high-risk decision making such as controlling self-driving cars, managing the power grid, or fighting wars (as autonomous weapon systems). As I mentioned above, the machine learning technology of today is not sufficiently reliable or robust to be entrusted with such dangerous decisions. I am very concerned that premature deployment of AI technologies could lead to a major loss of life because of some bug in the machine learning components. Much like the HAL 9000 in '2001, A Space Odyssey', computers could 'take over the world' because we gave them autonomous control of important systems and then there was a programming error or a machine learning failure. I do not believe that computers will spontaneously 'decide' to take over the world; that is just a science fiction story line. I also don't believe that computers will 'want to be like us'; that is another story line that goes back at least to the Pinocchio story (and perhaps there is an even older story in Chinese culture?).

Downloaded from https://academic.oup.com/nsr/article/5/1/54/3789514 by guest on 24 October 2020

> I think people should be "in the loop" in all of these high-risk decision making applications.
>
> — Thomas Dietterich

People may not be as accurate or as fast as computers in making decisions, but we are more robust to unanticipated aspects of the world and hence better able to recognize and respond to failures in the computer system. For this reason, I think people should be 'in the loop' in all of these high-risk decision making applications.

MISCELLANEOUS

NSR: Many machine learning professors in the US have moved to big companies. Could you comment on this?

Dietterich: Yes, there has been a substantial 'brain drain' as professors move to companies. Let me discuss the causes and the effects of this. There are several causes. First, because many companies are engaged in a race to develop new AI products, they are offering very large salaries to professors. Second, because many machine learning techniques (especially deep learning) require large amounts of data and because companies are able to collect large amounts of data, it is much easier to do research on 'big data' and deep learning at companies than at universities. Third, companies can also afford to purchase or develop special computers for deep learning such as GPU computers or Google's Tensor Processing Units (TPUs). This is another thing that is very difficult to do in universities.

What are the effects of this? The main effect is that universities can't train as many students in AI and machine learning as they could in the past, because they lack the professors to teach and guide research. Universities also lack access to big data sets and to special computers. Industry and government should address these problems by providing funds for collecting data sets and for purchasing specialized computers. I'm not sure how governments can address the brain drain problem, but they can address the data and computing problems.

With the exception of work on big data and deep learning, all other forms of machine learning (and all of the challenges that I listed above) are easy to study in university laboratories. In our lab at Oregon State University, for example, we are studying anomaly detection, reinforcement learning and robust machine learning.

NSR: Could you comment on the contributions and impact to the field that are coming from China? Do you feel any obstacle encumbering Chinese researchers from making higher impact?

Dietterich: Chinese scientists (working both inside and outside China) are making huge contributions to the development of machine learning and AI technologies. China is a leader in deep learning for speech recognition and natural language translation, and I am expecting many more contributions from Chinese researchers as a result of the major investments of government and industry in AI research in China. I think the biggest obstacle to having higher impact is communication. Most computer science research is published in English, and because English is difficult for Mandarin speakers to learn, this makes it difficult for Chinese scientists to write papers and give presentations that have a big impact. The same problem occurs in the reverse direction. China is now the home of a major faction of AI research (I would guess at least 25%). People in the West who do not read Chinese are slow to learn about advances in China. I hope that the ongoing improvements in language translation will help lower the language barrier. A related communication problem is that the internet connection between China and the rest of the world is often difficult to use. This makes it hard to have teleconferences or Skype meetings, and that often means that researchers in China are not included in international research projects.

NSR: What suggestions will you give young researchers entering this field?

Dietterich: My first suggestion is that students learn as much mathematics as possible. Mathematics is central to machine learning, and math is difficult to learn on your own. So I recommend all students in university to study mathematics. My second suggestion is to read the literature as much as possible. Don't just read the deep learning papers, but study the theory of machine learning, AI and algorithms. New learning algorithms arise from deep understanding of the mathematics and the structure of optimization problems. Don't forget that ideas in other branches of knowledge (e.g. statistics, operations research, information theory, economics, game theory, philosophy of science, neuroscience, psychology) have been very important to the development of machine learning and AI. It is valuable to gain experience working in teams. Most research today is collaborative, so you should get practice working in teams and learning how to resolve conflicts. Finally, it is important to cultivate your skills in programming and in communication. Learn to program well and to master the latest software engineering tools. Learn to write and to speak well. This goes far beyond just learning English grammar and vocabulary. You must learn how to tell a compelling story about your research that brings out the key ideas and places them in context.

Zhi-Hua Zhou is a professor at Nanjing University in China.

Downloaded from https://academic.oup.com/nsr/article/5/1/54/3789514 by guest on 24 October 2020