Machine Humanity: How the Machine Learning of Today is Driving the Artificial Intelligence of Tomorrow

ANDREW PEASE

Principal Business Solutions Manager, Global Technology Practice, SAS Institute

JOSEFIN ROSÉN

Principal Advisor Analytics, Nordic Government, SAS Institute

December 2016

Interviewed by Robert Morison, IIA Lead Faculty

DISCUSSION OVERVIEW

Machine learning is hot and for good reason. The components - big data, computing power, analytical methods - are in place, and compelling applications are multiplying. To capitalize on the technology, organizations must build experience. They must also proceed pragmatically with one eye on the business and the other on the ethical implications of the algorithms deployed and the decisions automatically made. To explore the opportunities, challenges, and success factors of machine learning today and tomorrow, IIA spoke with Andrew Pease, Principal Business Solutions Manager, Global Technology Practice at SAS Institute and Josefin Rosén, Principal Advisor Analytics, Nordic Government at SAS Institute.

How do you define machine learning?

Andrew Pease (AP): Let's start with the definition that Arthur Samuel coined in 1959. He called it a "field of study that gives computers the ability to learn without being explicitly programmed." That relates directly to what we're trying to do with machine learning today. We're trying to take data, intelligently filter it, and come down to insights that help us to make better business decisions. My working definition of machine learning is "using algorithms to intelligently filter data for an actionable insight."

That said, here's a provocative and cautionary description I heard recently: Machine learning can be "money laundering for bias." If we're not careful, our machine learning algorithms can reinforce societal biases reflected in the data. Those algorithms can provide a rationale — the computer says so, so it must be okay — for making arbitrary or socially unfair decisions about creditworthiness or insurability, for example. As data scientists, we have to think about the effects of machine learning outputs on the organization, and be alert to the possibility that the algorithms may be doing things that would conflict with our personal and organizational ethics. We have to be cognizant of what data is going in and what's being done with that data, and make sure that's in line with our organizational objectives.

Josefin Rosén (JR): Let me expand on the learning process. Machine learning enables computers to find hidden insights without being explicitly programmed where to look. They can change and improve their algorithms by themselves, and every time an error is made and recognized, the algorithm corrects itself and begins another iteration of the learning. With the help of feedback loops, machine learning automates analytical model building and improvement.

We have to remember that the algorithms will just learn what we show them. They don't have their own will, and the insights gained are directly related to the data they are exposed to. There is still a need for humans to assess and make decisions on how the machine should respond and if and when it should take actions.

How does machine learning fit into the field of artificial intelligence?

JR: Machine learning is a subset of artificial intelligence. In layman's terms, artificial intelligence is like an umbrella including everything that enables computers to mimic or imitate human intelligence.

This means that simple if-then rules could be artificial intelligence, but also more advanced machine learning, like deep learning.

AP: I look at artificial intelligence as a flexible, rational application that can perceive the current situation and then take actions that maximize a chance of success at some stated goal. Machine learning is one of the main tools for building artificial intelligence. Not the only one, but it is certainly a very important one. Intelligently filtering data is a way to learn about the current situation and assess what actions make sense. The essential thing to recognize is that what we do today with machine learning is setting up a framework for how we're going to interact with artificial intelligence and how artificial intelligence is going to play a part in our lives. We have to be very, very cognizant of that.

How are artificial intelligence and machine learning changing our lives?

JR: Machine learning is already a large part of our lives, even though we might not recognize it. It is everywhere — the online recommendations we receive for products and services and potential friends, the spam filters in our email, the smart voice assistants like Siri and Cortana, Google's translation function. The technologies also assist in medical diagnostics, fraud detection, and various "smart" applications like smart homes, to mention only a few. As computing capabilities and available data continue to grow, all these applications will naturally become even smarter and more efficient.

At the same time, it will be exciting to see what happens at the intersection of machine learning and cognitive computing, including natural language processing and generation. That will profoundly reshape human-machine interactions and open new possibilities in fields ranging from remote medicine to public security to how we as individuals manage our time, organize our space, and connect with others.

AP: Machine learning is getting a lot of attention right now. It's important that people are really starting to focus on what machine learning can mean for their businesses, for their lives, and for society.

At a basic level, these technologies have been enabling a lot more automation of repetitive tasks and often mindless work. Computers can execute such tasks in a much more error-free way because they don't get bored. Now, however, we have machines performing much more complex and interrelated tasks, such as driving a car. Over the next decade, driverless cars are going to transform the transportation and travel industries and pave the way for smarter cities that need less space for cars. We have to closely monitor such developments.

People most often think in terms of what intelligent machines can do for us, but we must also be thinking about what *we* can do for *them*. How will we continue to add value to what machines can do? As machines perform more sophisticated activities, humans need to add value where artificial intelligence struggles — with creativity and sociability, for example. Our education systems need to evolve to make sure that we can continue to add value on top of what intelligent machines will be able to do for us.

How are artificial intelligence and machine learning changing business models?

AP: The computer chip industry provides a revealing example. High-tech manufacturing already deploys robotics extensively. The new addition is 3-D printing to generate chips. The 3-D printer can be accurate to a tolerance of one in a billion droplet errors. That

sounds like error- free, except that when we're talking about upward of 50 million droplets per second, we can have an error every 20 seconds. All kinds of preemptive analytics, including machine learning algorithms working on photographic image data from the production lines, are at work to spot and minimize errors. That may sound like an extreme example, but it is a reality with today's technology and machine learning will continue to enhance quality control in manufacturing.

Machine learning is also transforming medical and legal practices. I have a good doctor, but I'm more confident if she can consult a wider body of cases in determining my diagnosis and prescribing treatment. That's most important with serious conditions like cancer, where medical knowledge and treatment options are changing rapidly. Algorithms can sort through and make sense of up-to-date medical data much more quickly and thoroughly than an individual doctor can. There's a similar situation in legal, where the number of potentially relevant precedents can be large. Algorithms can process the often lengthy descriptions and help lawyers to find the right precedents concerning a particular case.

JR*:* Customer experience is an example that cuts across industries. Machine learning is embedded in day-to-day business processes to identify prospects, shape interactions with customers, and personalize services and offers. Technology can provide services on demand when people are too slow, add capacity when there's a shortage of skilled people, and conduct interactions in ways that are more convenient to the customer. Every organization has got to be thinking about how to interact with customers in the right ways via the right channels at the right times. Customers expect it, and machine learning is essential to keeping pace with their expectations.

We're also seeing more and more artificial intelligence based advisors, both general purpose such as the assistants in our smartphones and very specialized such as the robo-advisors in financial services. Tieto, a Nordic software and services company, has appointed an artificial intelligence application called Alicia T to its leadership team, with the capacity to cast votes. The goal is to drive the whole team to be thoroughly data driven and innovative in how it uses data in the business.

Please say more about why machine learning is so important today. Why should organizations be focusing on it now?

JR*:* Compared to only a few years ago, we are operating in a new context. We have accelerated the generation of data through the Internet of Things, social media, and advancement in imaging technologies. We also have virtually unlimited opportunities for data storage and computing power to work with that data. We can handle unstructured data as never before and combine it with traditional structured data with innovative results. We can employ algorithmic techniques such as deep learning that perform best with millions of training examples. Together these factors make machine learning an appropriate tool for more and more tasks. There are many more use cases, and the added value of this new information to organizations is obviously vast. Organizations that embrace machine learning have larger potential to identify opportunities and also to avoid risk, and they will get well ahead of those that lag technologically.

AP*:* Josefin framed it well. We have a convergence of more data and more processing capability, including in-memory technology. That means we can apply more advanced algorithms to more and more complex

use cases. And we can think beyond the realm of what was possible just a few years ago. Organizations that invest in know-how, gain experience, and approach machine learning pragmatically today will be ready as the technology continues to advance and new opportunities arise. Organizations that don't are going to fall behind very, very quickly.

There's a second fundamental reason to be working with machine learning today — to build experience with "artificial psychology" alongside artificial intelligence. We have to recognize the implications of having machines make more decisions. How can we get algorithms to represent our values and ethics as well as the facts and figures?

Let's take the example of a bank's credit-scoring model. They want to use all available data to identify who should receive credit. The data screams out that single mothers are high credit risks. Does the bank want to reject all applications from single mothers? I don't think so. They are more likely to need and value the opportunity to build up their credit profiles. So what's the right decision and action for the bank? It may involve adjusting the credit terms. But if it is to arbitrarily deny credit, or make the credit prohibitively expensive, then the bank and its algorithms are reinforcing social stereotypes and prejudice against single mothers.

The algorithm is not going to think about these implications. As producers of the algorithms, data scientists have to. So do business analysts and the business people making decisions or approving the automated decisions. They all have a responsibility to understand the data in play, the techniques at work, and how the analytics are applied in the organization.

This is why I said earlier that what we do with machine learning today will shape our use of artificial intelligence going forward. We have to wrestle with

these issues now and learn to avoid unanticipated outcomes as we work with increasingly intelligent machines.

What does it take to succeed with machine learning - in terms of people, processes, and technology? What do organizations need to get right?

JR*:* The people side can be challenging because the skills and talents needed are in high demand and often hard to come by. The key role is the data scientist, but it really helps to have versatile business analysts who can work effectively with the data scientists. If the organization is trying to gain traction with advanced analytics, it may be time to appoint a chief analytics officer. This person should focus on the business strategy and use of analytics, as well as have the technical background to move initiatives forward. Ideally, analytical attitudes and capabilities also extend to the top management of the organization.

Key processes support access to and management of huge amounts of data. It cannot just be any data, but rather data that is relevant and reliable for the use case. The machine learning algorithms need a lot of data and examples to train on. Many organizations do not have the historical data needed and must first adjust their processes to capture this data to begin with. Also essential are processes for assessing the performance of analytical models, keeping them relevant, and providing them the feedback to learn and improve over time.

Essential technology capabilities include data preparation, management, and exploration, plus software for building and managing both basic and advanced algorithms and models. An integrated end-to-end platform for model development, deployment, and management will accelerate the time to business results and help maintain the integrity and value of

models. Finally, given the explosive growth of data to work with, the whole platform has got to be scalable.

AP: Let me emphasize the point on the platform. Having the technology and processes to automate as much of the model life cycle as possible is a huge success factor. This includes being able to orchestrate and automate the movement from data to model to deployment in a production environment to maintaining and enhancing and eventually retiring the model and recycling its components. There's a lot of room for error in there, but we can minimize the risk.

In the people realm, retaining top-caliber data scientists can be more difficult than attracting them in the first place. You've got to make sure they're working on a variety of interesting challenges, and not spending too much time on mundane tasks. They have to pay close attention to the business implications and effects of their models, but they shouldn't have to do very much of the day-to-day monitoring and management of model performance. It also helps to have technical or perhaps nontraditional career paths for data scientists and other professional analysts. And because data scientist unicorns — proficient in business, data, algorithms, and communication — are so rare, it really helps to have a strong surrounding cast.

That leads to what may be the biggest success factor of all — the collaboration among business experts, business analysts, and data scientists. Collaboration that is based on transparency about what's happening and why. The more business analysts and experts understand the possibilities in the data and algorithms, the more they can exercise their curiosity, creativity, and due diligence. The business analysts should be saying things like, "What if we looked at this additional data as well?" "If we can do this with the data, we can also do that." "Wait a minute. We should think about the legal and ethical ramifications of using this data in this way." The technology platform should

support this collaboration and facilitate regular dialog by sharing visuals and information about data very quickly and easily.

How can organizations best ramp-up their machine learning capabilities?

AP: For an organization's first forays into machine learning, it always makes sense to find use cases that combine high value with good feasibility. The initiative should really solve a key challenge within the business, because success earns visibility, management attention, and forward momentum. On the flip side, you want the initiative to be feasible in that there is already a lot of data available, or it would be relatively easy to collect the necessary data and start applying machine learning on it.

Once an organization is getting experience in machine learning and has generated those visible wins, the key ingredient is top management support. That raises interest across the organization; drives investment in people, processes, and technology; and opens up opportunities to compete on analytics with innovative use cases for machine learning that are ahead of the competition.

In the most advanced organizations, analytics are everywhere. A great example is DirectPay, a credit management business based in the Netherlands. They buy and try to collect upon loans that banks are writing off. Machine learning algorithms inform all their key decisions: Does this loan have a high propensity for successful collection? At what level — 20 percent — 60 percent — more? What should we be willing to pay to take over this loan? On the outbound side, they optimize channels and collection methods. Their investment in machine learning and analytics infrastructure keeps paying off, as the algorithms get smarter and contribute more to business profitability.

JR*:* I'll simply summarize that most organizations do well to start relatively small with a specific business need with measurable results. Make every success very visible so that business people see how machine learning results in concrete improvements. Note that "visible" demands that we make analytics approachable as well as accurate, with the help of intuitive interfaces and built-in explanations behind automated decisions and recommendations. And be sure that technological skills development is keeping pace with business demand and opportunity.

What should organizations be watching out for with machine learning? We've talked about ethical issues. What are other pitfalls or unanticipated consequences?

JR*:* We've touched on several. Models can be hampered by lack of transparency, interpretability or auditability. Algorithms may have very high accuracy, but if they are complex and difficult to understand, they become a "black box" to business people — not fully trusted and not fully used. In regulated industries like banking, models are required to be explainable.

Another growing consideration is data privacy. We are generating and collecting more and more data, including all kinds of personal data. Privacy restrictions may limit how granular machine learning algorithms can be, and model output in domains like medical diagnosis may immediately become protected data. Privacy must be protected according to different regulations in different countries or regions. And the increasing value of data has raised the question of its ownership and use. Who should benefit from the use of online data about me?

An anticipated consequence and worry for people when discussing machine learning and artificial intelligence is job loss. What happens if our jobs, even forms of knowledge work, become increasingly

automated? We need to remember that computers can simulate intelligent behavior and work autonomously, but only within the domains where they are trained. Machines will not replace human knowledge and capability; they will augment it and enable us to be more efficient. Of course, this works to society's detriment if the machines are controlled by the wrong hands and programmed to do bad things better.

AP*:* I'll give you four warnings. First, in the enthusiasm over what machine learning can do, don't oversell it. Be pragmatic about what you can achieve, scrutinize what your data landscape looks like, and have a good idea of the business issues and the change management challenges. Then match potential use cases against those factors, and determine how to really put machine learning on the organization's map.

Second, avoid the one-shots, the applications that answer one question, change one thing, and then you're done. That may deliver valuable data-driven insight, and that's fine. But it's not taking advantage of the essence of machine learning, which is to use evidence to improve over time. Instead, look for applied cases where you're influencing operational decisions and the influence and value of machine learning can grow. And maximize the uptake by making sure that the intelligence and insights are pushed out to the decision points.

Third, focus on the feedback loops. It can be easy to get caught up in mastering the data and developing the algorithms, and pay too little attention to the design and robustness of the feedback that enables the algorithms to learn. You can fit the algorithm once on all the existing data, but trends change. Business constructs evolve. Feedback helps the model remain relevant, and the feedback process can often be fully automated, so the model reconfigures and redeploys itself.

Number four relates to the ethical side of things. Be careful when dealing with "hot" variables, data elements that are too sensitive or politically incorrect to incorporate directly into predictive models. Many times, however, you can find another variable that behaves similarly and can serve as a surrogate to represent exactly the same thing. But is that the right thing to do? Should we use that as a justification for what we're doing with the algorithm? Are we simply skirting the sensitive issue and possibly "money laundering bias"? Again, we as machine learning practitioners have to be thinking hard about these things.

To wrap up, what are the top things that business leaders need to know and do about machine learning?

JR: Become familiar with machine learning, not so much technologically as in terms of its business potential. You need to know enough to find the opportunities to leverage the technological capabilities in your business strategy. That's the standard to meet.

Also recognize the importance of data and support data generation, management, and quality. Machine learning needs large volumes of data to train on. Much the same can be said of a business at large.

Finally, machine learning algorithms are built by very skilled people. Invest in them and in the technology platform and processes that support their work.

AP: I'll expand on Josefin's first point. The way to become familiar with machine learning is to talk with your data scientists. And hold them accountable for explaining in business terms what the algorithms do, what data they use, what the outputs try to say, and what the implications are of using them. It may take some back-and-forth, but as a business leader you have responsibility to be informed on technological

opportunities, and that may involve learning a little bit about the underlying processes along the way.

Then recognize that you don't have time to play a waiting game anymore. If your organization is not looking at how machine learning is going to impact your industry and what you need to be doing about it, you'd better get going. The data, the technology, and the business use cases are all here. It's all happening now. Organizations that aren't getting up to speed might not be around for that much longer. And don't wait to find the perfect data scientist to get you started. There are plenty of resources in the marketplace to assist you. So don't hesitate — explore machine learning and seize the opportunities.

Additional Information

To learn more about this topic, please visit: http://www.sas.com/en_us/insights/analytics/machine-learning.html.

About the Interviewees

ANDREW PEASE

For 18 years at SAS, Andrew has worked to evangelize customers to the power of analytics, while keeping a pragmatic view on project execution. He is currently leading the EMEA analytics team and is passionate about maximum data analysis for tackling the most challenging business issues. He helps financial institutions, major retailers, pharmaceuticals, manufacturers, utilities and public sector organizations to understand and exploit powerful analytic techniques such as decision management, predictive modelling, time-series forecasting, optimization, and text mining.

Andrew currently spends time helping organizations develop analytic roadmaps for dealing with "Big Data," exploiting Hadoop and using High Performance Analytics to embed analytics in core business processes for strategically consistent decisions across the enterprise. He also thrives in high stakes, customer-facing presales engagements. Finally, he works to spread the word about the latest analytic technologies both internally to SAS, as well as frequently speaking at external events.

Andrew has a master's degree in clinical psychology with a specialization in experimental design. He lives in Leuven, Belgium and enjoys playing the guitar and piano, traveling with his family, and getting in the occasional run to keep in shape.

JOSEFIN ROSÉN

Josefin is a curious, enthusiastic, and data driven analytics expert with an intuitive understanding of applied advanced analytics. In her role as Principal Advisor Analytics, Nordic Government, she supports customers from a wide variety of businesses, industries, and academic fields in getting value from their often huge amounts of both unstructured and structured data and drive smarter, more strategic business through advanced analytics. Specialties include: machine learning, data mining, and text analytics.

Josefin holds a Ph. D. in Computational Chemistry from the Faculty of Pharmacy, Uppsala University.

IIANALYTICS.COM