On Welsh corgis, computer vision, and the power of deep learning

July 16 2014

Members of the team that worked on the asynchronous DNN project: (from left) Karthik Kalyanaraman, Trishul Chilimbi, Johnson Apacible, Yutaka Suzue

Can you tell the difference between the two breeds of corgis? If you're like many, you probably are barely even aware that such dogs exist, let alone the fact that there are two—and only two—kinds of corgis. Add the detail that those two breeds are both named after Welsh counties—the Pembroke Welsh corgi and the Cardigan Welsh corgi—and it's safe to say that the people who could correctly identify one from the other are few and far between..

But Project Adam can.

Project Adam, an initiative by Microsoft researchers and engineers, aims to demonstrate that large-scale, commodity distributed systems can train huge deep neural networks effectively. For proof, the researchers created the world's best photograph classifier, using 14 million images from ImageNet, an image database divided into 22,000 categories.

Included in the vast array of categories are some that pertain to dogs. Project Adam knows dogs. It can identify dogs in images. It can identify kinds of dogs. It can even identify particular breeds, such as whether a corgi is a Pembroke or a Cardigan.

Now, if this all sounds vaguely familiar, that's because it is—vaguely. A couple of years ago, The New York Timeswrote a story about Google using a network of 16,000 computers to teach itself to identify images of cats. That is a difficult task for computers, and it was an impressive achievement.

Project Adam is 50 times faster—and more than twice as accurate, as outlined in a paper currently under academic review. In addition, it is efficient, using 30 times fewer machines, and scalable, areas in which the Google effort fell short.

"We wanted to build a highly efficient, highly scalable distributed system from commodity PCs that has world-class training speed, scalability, and task accuracy for an important large-scale task," says Trishul Chilimbi, one of the Microsoft researchers who spearheaded the Project Adam effort. "We focused on vision because that was the task for which we had the largest publicly available data set.

"We said, 'OK, if we have a really scalable system, let's prove it works.' The challenge was: Can we do really well with this large-scale system to

train large models on this large data set? And the answer was: Yes, we could. Our system is general-purpose and supports training a wide variety of deep-neural-network [DNN] architectures. It also can be used to train large-scale DNNs for tasks such as speech recognition and text processing."

The project—which also included Johnson Apacible, engineering manager for Microsoft Research Special Projects, and engineers Yutaka Suzue and Karthik Kalyanaraman—addresses the suddenly surging potential offered by deep learning: hierarchical representation learning using big data. It's a game-changing approach that mimics the learning hierarchy used by the human brain.

"The machine-learning models we have trained in the past have been very tiny, especially in comparison to the size of our brain in terms of connections between neurons," Chilimbi says. "What the Google work had indicated is that if you train a larger model on more data, you do better on hard AI [artificial intelligence] tasks like image classification.

"We wanted to see if we could build a much more scalable, much more efficient system to train much larger models on larger amounts of data. Would we also see improvements in task accuracy? That was the overarching research goal, to build a scalable training system to prove that systems trained on large amounts of data is a promising path to follow and that you don't need to be a machine-learning expert to get good accuracy in some of these tasks. A system-driven approach by using brute-force computing, scale of model, scale of data is a viable approach."

What made that confirmation all the more satisfying is that there were many who said it couldn't be done.

"There was a lot of skepticism when we started out from machine-

learning experts around using the distributed system to do machine learning," Chilimbi says. "The fundamental machine-learning training algorithms are synchronous. They've typically been run on a single machine. They said, "Yes, you can do this distributed, but the synchronization cost will make it so slow that it's never going to be high-performance or scalable.

"One of the innovations we came up was saying that not only can we make it asynchronous, but we went whole hog and decided not to pretend it's synchronous in any way. We figured out a way to make the asynchrony not just learn but learn better, because it adds a level of robustness. Learning is not so much about optimizing on the training set of data. It's about generalizing well on unseen data."

The asynchronous technique offers an additional benefit.

"The asynchrony also helps us escape from ruts where the task accuracy does not improve much," Chilimbi says, "much like how humans learning a new task often find themselves plateauing after a period of rapid improvement."

Apacible expands on the Project Adam approach.

"As a child, you're shown pictures of an entire car," he explains, "but as an adult, sometimes in the corner of a window you see only part of a car, but you still know it's a car. You get trained. When the car is moving fast, then the picture becomes a little bit blurry, but you still know it's a car.

"This is what the system does. It allows it to train with different types of data, with different types of situations, and it makes the model more robust."

When the project began 18 months ago, its goals were far from modest. It was scoped to deliver the vision of a full-functionality system with an end-to-end scenario and a successful sustained operation spanning multiple days. And it needed to achieve world records in the size of the models, the speed of the training, and its accuracy in classifying the massive ImageNet collection.

But with such lofty ambitions also came plenty of support.

"This wouldn't have been possible," Apacible says, "if the lab under Peter Lee [head of Microsoft Research] had not invested in these types of disruptive projects and had Yi-Min Wang [managing director of Microsoft Research Technologies] not provided initial backing and acted as an angel investor.

"The goal was to come up with a very risky, highly disruptive project and just support it end to end. And because of the trust and the support provided to all of us, we were able to come up with a big success."

Not only that, but the project also underscored the fact that deep learning, previously shown to work effectively in the speech domain, also can perform wonders on vision tasks. And the researchers gained more understanding about how DNNs actually work.

"What we found," Chilimbi says, "was that as you add levels to the DNN, you get better accuracy—until a certain point. Going from two convolutional layers to three convolutional layers to five or six seems the sweet spot. And the interesting thing is that people have done studies on the human visual cortex, and they've found that it's around six layers deep in terms of layers of neurons in the brain.

"The reason it's interesting is that each layer of this neural network learns automatically a higher-level feature based on the layer below it.

The top-level layer learns high-level concepts like plants, written text, or shiny objects. It seems that you come to a point where there's diminishing returns to going another level deep. Biologically, it seems the right thing, as well."

To return to Project Adam's ability to identify our corgi friends, the layers would work like this: The first layer learns the contours of a dog's shape. The next layer might learn about textures and fur, and the third then could learn body parts—shapes of ears and eyes. The fourth layer would learn about more complex body parts, and the fifth layer would be reserved for high-level recognizable concepts such as dog faces. The information bubbles to the top, gaining increasingly complex visual comprehension at each step along the way.

Asked about the disruptiveness of DNNs in today's computing environment, Chilimbi refers back to what he calls the two ages of computing up to now, the first driven by computers getting faster and Moore's Law, and the second being the age of the Internet and communications and connectivity.

"These two eras were very transformative, and people would say that a lot of things came out of these," he says. "I think we're in the very early days of going to some form of true AI. I think it's going to be transformative in a similar sense, in that it needed the previous revolutions, the computing-power increase to be able to power it, and the connectivity and availability of data to be able to learn things that are interesting.

"Computers until now have been really good number crunchers. Now, we're starting to teach them to be pattern recognizers. Marrying these two things together will open a new world of applications that we couldn't imagine doing otherwise. Imagine if you could help blind people see by pointing a cellphone at a scene and having it describe the scene to

them. We could do things like take a photograph of food we're eating and have it provide us with nutritional information. We can use that to make smarter choices."

Point, click, and view: The technology can identify a breed of dog almost instantly--and the potential for future use can scarcely be imagined.

For Apacible, the transformation that deep learning enables is all about scale.

"When computers started," he says, "people were programming with vacuum tubes, and then there was assembly language, which helped a bit. Then there was the C language, which helped tremendously in terms of getting more and more code written. We're now in the age of data. We've got more and more data. A product like Bing requires hundreds of machine-learning experiences to be able to come up with good relevance models.

"If you look at the scale of millions of images, that would require hundreds, if not thousands, of machine-learning experts to even come up with a model. What this system has proven is that, with DNN, you could scale that. You don't need machine-learning experts trying to figure out what makes this look like a dog. The system learns that on its own. There is this promise of massive scale."

That scale could be used to train a system to represent, understand, and help explain the world around us by supplying the system with vast quantities of data across multiple modalities such as images, speech, and text.

The researchers are quick to note that these insights would not have been possible without the contributions from Suzue and Kalyanaraman.

"Yutaka and Karthik are both distributed-systems engineers," Apacible says. "Yutaka especially likes to work at the bit level, where he can optimize everything, and Karthik likes to think big in terms of how you design things. They make a great team, because Karthik can think about something complex, like the parameter server and how all the machines exchange data between them, while Yutaka deals with how he can optimize each box to the fullest."

Each time the subject turns to DNNs these days, the discussion rarely fails to refer to the mystery behind some of the wondrous things deep learning is able to achieve.

"The deep, mysterious thing that we still don't understand," Chilimbi says, "is how does a DNN, where all you're presenting it is an image, and you're saying, 'This is a Pembroke Welsh corgi'—how does it figure out how to decompose the image into these levels of features?

"There's no instruction that we provide for that. You just have training

algorithms saying, 'This is the image, this is the label.' It automatically figures out these hierarchical features. That's still a deep, mysterious, not well understood process. But then, nature has had several million years to work her magic in shaping the brain, so it shouldn't be surprising that we will need time to slowly unravel the mysteries."

The situation, though, is not entirely unique.

"It's like in quantum physics at the beginning of the 20th century," Chilimbi says. "The experimentalists and practitioners were ahead of the theoreticians. They couldn't explain the results. We appear to be at a similar stage with DNNs. We're realizing the power and the capabilities, but we still don't understand the fundamentals of exactly how they work.

"We tend to overestimate the impact of disruptive technologies in the short term and underestimate their long-term impact—the Internet being a good case in point. With deep learning, there's still a lot more to be done on the theoretical side."

Provided by Microsoft