Science | DOI:10.1145/2461256.2461262 Gary Anthes

Deep Learning Comes of Age

Advances on multiple fronts are bringing big improvements to the way computers learn, increasing the accuracy of speech and vision systems.

IMPROVEMENTS IN ALGORITHMS and application architectures, coupled with the recent availability of very fast computers and huge datasets, are enabling major increases in the power of machine learning systems. In particular, multilayer artificial neural networks are producing startling improvements in the accuracy of computer vision, speech recognition, and other applications in a field that has become known as "deep learning."

Artificial neural networks ("neural nets") are patterned after the arrangement of neurons in the brain, and the connections, or synapses, between the neurons. Work on neural nets ~~dates to the 1960s~~; although conceptually compelling, they proved difficult to apply effectively, and they did not begin to find broad commercial use until the early 1990s.

Neural nets are systems of highly interconnected, simple processing elements. The behavior of the net changes according to the "weights" assigned to each connection, with the output of any node determined by the weighted sum of its inputs. The nets do not work according to hand-coded rules, as with traditional computer programs;

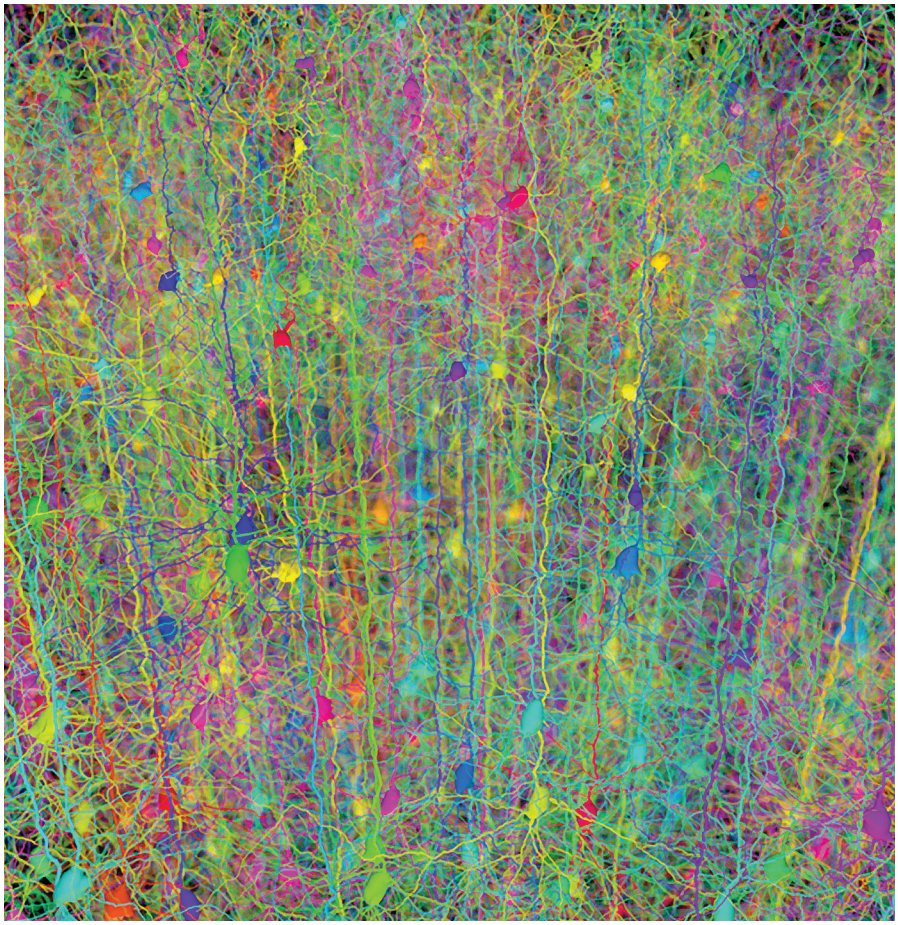

IMAGE BY HERMANN CUNTZ

Rainbow brainwaves made from a computer simulation of pyramidal neurons found in the cerebral cortex.

they must be trained, which involves an automated process of successively changing the inter-nodal weights in order to minimize the difference between the desired output and the actual output. Generally, the more input data used for this training, the better the results.

For years, most neural nets contained a single layer of "feature detectors" and were trained mainly with labeled data in a process called "supervised" training. In these kinds of networks, the system is shown an input and told what output it should produce, such as letters of the alphabet. (In unsupervised learning, the system attempts to model patterns in the unlabeled input without knowing in advance what the desired outputs are.) However, labeling data, particularly when there are many possible values, is labor-intensive and slow. Explains machine learning pioneer Geoffrey Hinton, a computer scientist at the University of Toronto, "The basic approach back then was, you hand-engineered a bunch of features, and then you learned what weights to put on the features to make a decision. For example: if it's red, it's more likely to be a car than a refrigerator."

In the 1980s, Hinton and others came up with a more powerful type of supervised learning, one that employed learning in multiple layers, combining low-level features into successively higher levels. However, Hinton says that with a few exceptions, these systems did not work as well as expected. The process of starting with very low-level features, such as the intensities of individual pixels, and learning multiple layers of features all at the same time, involved a huge amount of computation; computers were not fast enough, there was not enough labeled data, and system developers did not have a good way to initialize the weights.

Recent Developments

Since about 2005, the picture has changed dramatically. Hinton (with Yann LeCun, a professor of computer and neural science at New York University, and others) made a number of fundamental advancements in neural nets, principally with unsupervised learning and multilayer learning. Neither concept was new, but improved algorithms, more data, and faster computers make them work much better than early versions. Of unsupervised learning, Hinton says, "That's much more like what people do—you are not given labels. It's much easier to get unlabeled training data; just take a video camera and wave it at the world."

"A wave of excitement today comes from the application of unsupervised learning to deep neural nets."

While the idea of unsupervised learning had obvious appeal for years, it had not been clear how to put it inside a hierarchical, multilayer system, says research collaborator LeCun. "You want four or five or six layers, because that's how you go from edges to textures to parts of objects to whole objects in particular configurations to whole objects regardless of configuration." And that is conceptually how the brain works, research has shown.

"A wave of excitement today comes from the application of unsupervised learning to deep neural nets," LeCun says, adding that another wave of excitement surrounds the use of the more-traditional supervised training of multilayer systems called "convolutional" neural nets. LeCun developed convolutional neural nets at Bell Laboratories in the late 1980s; they were among the earliest to employ multilayer learning.

Convolutional nets, a biologically inspired model for image recognition, can recognize visual patterns directly from pixel images with minimal preprocessing. The system processes a small—say 10 by 10 pixels—portion of an image, looking for small features such as edges, then slides the 10-by-10 window over one pixel and repeats the operation until the entire image has been processed. It produces output values that are sums of the inputs weighted by the synaptic weights in each small window.

The reason these nets succeeded in an era of relatively weak computers is that by working on a small window—100 pixels, say—instead of millions of pixels, the number of connections and weights, and hence the computational workload, was greatly reduced.

Today, faster computers, more data and some "simple architectural tricks" applied by LeCun have allowed multilayer convolutional neural nets to become practical with unsupervised learning, where the net is trained, one layer after another, using only unlabeled data. The process involves initial unsupervised training to initialize the weights in each layer of the network, followed by a supervised global refinement of the network. "This works very well when you have very fast machines and very large datasets, and we have just had access to those recently," LeCun says. Convolutional neural nets are well suited to hardware implementations, he says, and we will see many embedded vision systems based on them in the coming years.

This approach — initializing weights in each layer at first using a large amount of unlabeled data, followed by fine-tuning the global net using a smaller amount of labeled data — is called "semi-supervised" learning. "Before, if you initialized purely randomly, the systems would get lost," says John Platt, manager of the Machine Learning Groups at Microsoft Research (MSR). "This gives the network a shove in the right direction."

Results

The use of semi-supervised learning and deep neural nets is the basis for some of the more dramatic results seen recently in pattern recognition. For 20 years, most speech systems have been based on a learning method that does not use neural nets. In 2011, however, computer scientists at MSR, building on earlier work with the University of Toronto, used a combination of labeled and unlabeled data in a deep neural net to lower the error rate of a speech recognition

system on a standard industry benchmark from 24% to about 16%. "Core speech recognition has been stuck at about 24% for more than a decade," Platt says. "Clever new ideas typically get a 2% to 5% relative improvement, so a 30% improvement is astounding. That really made the speech people sit up and take notice."

In last year's ImageNet Large Scale Visual Recognition Challenge, Hinton's team from the University of Toronto scored first with a supervised, seven-layer convolutional neural network trained on raw pixel values, utilizing two NVIDIA graphics processing units (GPUs) for a week. The neural network also used a new method called "dropout" to reduce overfitting, in which the model finds properties that fit the training data but are not representative of the real world. Using these methods, the University of Toronto team came in with a 16% error rate in classifying 1.2 million images, against a 26% error rate by its closest competitors. "It is a staggeringly impressive improvement," says Andrew Zisserman, a computer vision expert at the University of Oxford in the U.K. "It will have a big impact in the vision community."

Also last year, researchers at Google and Stanford University claimed a 70% improvement over previous best results in a mammoth nine-layer neural network that learned to recognize faces without recourse to any labeled data at all. The system, with one billion connections, was trained over three days on 10 million images using a cluster of machines with a total of 16,000 cores.

The different models for learning via neural nets, and their variations and refinements, are myriad. Moreover, researchers do not always clearly understand why certain techniques work better than others. Still, the models share at least one thing: the more data available for training, the better the methods work.

MSR's Platt likens the machine learning problem to one of search, in which the network is looking for representations in the data. "Now it's much easier, because we have much more computation and much more data," he says. "The data constrains the search so you can throw away representations that are not useful."

Image and speech researchers are using GPUs, which can operate at teraflop levels, in many of their systems. Whether GPUs or traditional supercomputers or something else will come to dominate in the largest machine learning systems is a matter of debate. In any case, it is training these systems with large amounts of data, not using them, that is the computationally intensive task, and it is not one that lends itself readily to parallel, distributed processing, Platt says. As data availability continues to increase, so will the demand for compute power; "We don't know how much data we'll need to reach human performance," he says.

Hinton predicts, "These big, deep neural nets, trained on graphics processor boards or supercomputers, will take over machine learning. The hand-engineered systems will never catch up again." C

Further Reading

Le, Q. and seven others
Building High-level Features Using Large Scale Unsupervised Learning, *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland, 2012

Dahl, G., Yu, D., Deng, L., and Acero, A.
Context-Dependent Pre-trained Deep Neural Networks for Large Vocabulary Speech Recognition, draft accepted by *IEEE Transactions on Audio, Speech, and Language Processing*, http://research.microsoft.com/pubs/144412/dbn4lvcsr-transaslp.pdf

Hinton, G.
Brains, Sex, and Machine Learning, GoogleTechTalks (video), June 22, 2012 http://www.youtube.com/watch?feature=player_embedded&v=DleXA5ADG78#!

Krizhevsky, A., Sutskever, I., and Hinton, G.
ImageNet Classification with Deep Convolutional Neural Networks, paper to appear in *Proceedings of the Neural Information Processing Systems Foundation 2012 conference*, Lake Tahoe, NV http://www.image-net.org/challenges/LSVRC/2012/supervision.pdf

LeCun, Y., Kavukvuoglu, K., and Farabet, C.
Convolutional Networks and Applications in Vision, *Proc. International Symposium on Circuits and Systems*, IEEE, 2010 http://yann.lecun.com/exdb/publis/pdf/lecun-iscas-10.pdf

Gary Anthes is a technology writer and editor based in Arlington, VA.

In Memoriam

David Notkin, 1955–2013

David Notkin, Professor and Bradley Chair of the Department of Computer Science & Engineering (CSE) of the University of Washington (UW), and associate dean of Research and Graduate Studies in UW's College of Engineering, died April 22 at the age of 58.

Notkin's research focus was in software engineering; as he explained on the UW website, "understanding why software is so hard and expensive to change, and in turn reducing those difficulties and costs." These interests underscored Notkin's belief "that the ability to change software—that is, the 'softness' of software—is where its true power resides."

Eugene Spafford, chair of the ACM Public Policy Council, said Notkin "was the personification of 'considerate.' He was deeply thoughtful about everything he did and the people with whom he came in contact. He cared if people could succeed. He cared that tomorrow will be better than today, for more than himself. And most of all, he acted on his caring in ways that did, indeed, make a difference for many, many people."

Colleague and friend Ed Lazowska, Bill & Melinda Gates Chair in Computer Science & Engineering at UW, said Notkin was an accomplished researcher, but "he was all about people and relationships—his family, his friends, his professional colleagues, and most importantly, his students. At the University of Washington, and more broadly in the field, he was our compass: he could always be counted upon to point us in the right direction."

For more information on Notkin, visit the web page of Notkinfest (http://news.cs.washington.edu/2013/02/01/honoring-david-notkin/), an event held in February 2013 at UW CSE honoring Notkin for his contributions to the field of computer science, and to announce a fellowship in his name.