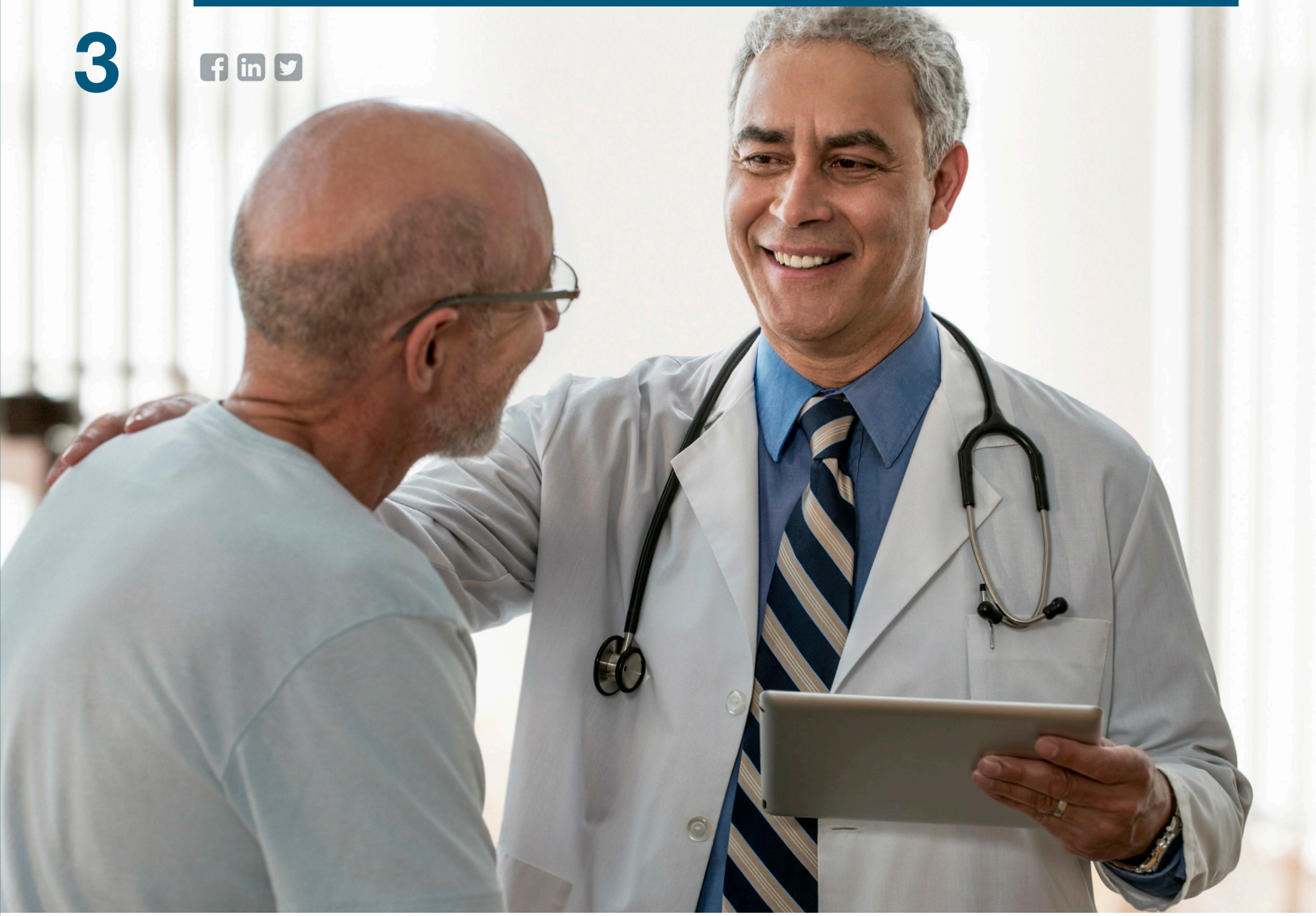
Reducing the clinical documentation burden with artificial intelligence

NUANCE

Amplifying human intelligence.

Physicians become physicians to care for people—to help healthy people stay healthy, and to help sick people get well. Clinical documentation, although an essential aspect of patient care, can drain physicians' time, energy, and attention away from the patient. The goal, instead, is to create technologies that enhance the physician and patient experience with documentation, helping physicians create more effective documentation more quickly while advancing patient care and satisfaction.

In this eBook, you'll hear from our experts about how we're creating technologies that use artificial intelligence and machine learning to mimic the way the human brain works—and what that means for the future of healthcare.



Contents

How many neural nets does it take to catch the big fish in machine learning?

By Nils Lenke, Senior Director Corporate Research, Nuance Communications

Deep neural nets (DNNs) have taken over machine learning these past few years, driving headlines and discussion within the industry and the media. That said, we're just scratching the surface with neural nets, which are evolving and changing, with many different approaches and challenges to address.

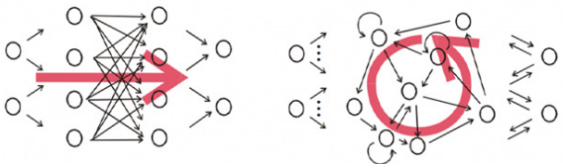
"Standard" DNNs are unidirectional: information flows in just one direction, from the input layer through the hidden layers to the output layer. In machine learning lingo, these DNNs are of the "feed-forward" type. They are best when all the information needed to learn is available at the same time. Think of image recognition: the image is available at once and the network can decide what it sees in it in one look or, in this case, in one pass through the network.

The teams here at Nuance are applying DNNs to advance speech recognition and natural language understanding as part of our mission to better facilitate communication between people and technology.

One of the interesting challenges with speech is that, unlike vision, it is embedded in time. An utterance unfolds over a number of seconds. And what happened a few minutes ago or even a few seconds ago helps clarify what is happening right now—better known as context. Technically speaking, of course, if you waited until the end of an utterance, you could make the whole utterance available to a DNN at once and a feed-forward network could access all the info it needs to do the recognition job in just one go. The problem is that for dialogue systems, like personal assistants, you cannot do that. As speech recognition is a heavy-duty computing job, engines start working as soon as an utterance begins and try to keep up with the speaker, so as to quickly offer up a response once the speaker is done talking, just like in a conversation between people.

As a result, the speech recognition engine will look at a slice of speech at a time. And to establish the context, we at Nuance use a special variant of DNNs, so-called recurrent neural nets (RNNs).

Their neurons not only take input from the left (as shown below), but they also have access to their own *previous* state (or in variants, even that of other neurons—see below). These feedback loops form a kind of memory.



Let's look at language modeling to illustrate that: language models (LMs) predict the next word based on the last so many words (where ideally we would not have to define a fixed number of words—it should be variable). For example, if you have already heard "God save the," then "queen" is a much more likely continuation than most other words. What we have found is that LMs based on RNNs work significantly better than traditional LMs.

Now let's look at natural language understanding (NLU), or mapping the recognized words to meaning via speech recognition.

One subtask is to identify "named entities." For example, in an inquiry like "Is there free capacity at the *parking garage* next to *Boston South Station*?," the two blocks of italicized words are such named entities. So in a first step, we want to label the words of the utterance as either belonging or not belonging to such named entity expressions. A decade or two ago, such a task would have been handled by, for example, HMMs (Hidden Markov Models) – the old workhorse of machine learning, also used in voice recognition before DNNs. But since then, another mathematical model has taken over, and it is especially good at such labeling or tagging tasks—in our case, a sequence of items being mapped to a set of labels.

This model is called Conditional Random Fields (CRFs). In contrast to the previous task we looked at (speech recognition) for NLU, we can afford to wait for the entire utterance to be available. The benefits of being able to look at all the words at the same time outweigh the little delay caused by the NLU processing step, which is very fast compared with automatic speech recognition (ASR). CRFs easily outperform HMMs on tasks like NER.

They have one soft spot, however. It takes a bit of manual work to tell them what they should look for in the input data (the sequence of words). Should they look only at the words' face value or also at their grammatical type? The neighbors left and right, and how far out? But this so-called feature selection is something that neural nets are good at: they evolve to learn on their own what the most valuable features are.

So why not combine CRFs with neural nets? That is exactly what our NLU team at Nuance has done.

In this model—NeuroCRFs—the neural nets do the feature induction part and the CRFs do the "rest."

We found that RNNs, with their built-in memory function, work especially well in combination with CRFs. This is because they can "remember" a context of variable length, whereas other neural nets would force us to arbitrarily define a context window size. Together with some clever tricks and optimizations, the resulting models can outperform an already good CRF accuracy baseline by more than 10 percent. (Two of my colleagues spoke about this in [much] more detail at ASRU 2015 in December—Marc-Antoine Rondeau and Yi Su: "RECENT IMPROVEMENTS TO NEUROCRFS FOR NAMED ENTITY RECOGNITION," in Proc. of ASRU 2015, pp. 390-396.)

The takeaway

First, while it is true that machine learning models, especially DNNs, are good at many tasks, it doesn't mean that the exact same type of DNN is the best answer in all cases. For that reason, there is a lot of hard work in research to find the best "net" to catch each proverbial fish—or in this case, a task.

Second, as an end user, you have no way of knowing which technology you are talking to. When you have called various ASR- and NLU-powered systems over time, you may have spoken to the different generations of technology. But the only difference you would have noticed was how the systems were becoming more accurate and more powerful all the time. And in the era of machine learning, with ever more data being turned into "big knowledge," it will not stop there.

Getting “deep” about “deep learning”

By Nils Lenke, Senior Director Corporate Research, Nuance Communications

Metaphors are everywhere—in popular culture and in science. Take “Elvis is the king of rock and roll.” Strictly speaking, rock and roll is no kingdom, but by applying the word “king” to it we mentally form it into one, with different ranks of characters and huge masses of underlings hailing their betters (i.e., the fans). Now take “deep learning,” a phrase tangled in a web of metaphors. Are there really machines that are complex enough to display true learning abilities, such as humans do?

How can something as abstract as learning be “deep” (or “shallow” for that matter)? These are some of the questions I explore in this post.

As movies like “Ex Machina,” “Her,” and “The Imitation Game,” continue to hit the big screen, we are also seeing a lot of excitement around “deep learning.” Just for fun, I entered “applies deep learning to” into a well-known search engine. Among the hundreds of results, “deep learning” is being applied to “satellite images to gain business insights,” “differentiate disease state in data collected in naturalistic settings,” “the task of understanding movie reviews,” “emotion prediction via physiological sensor data,” “natural language,” and—probably my favorite—“the tangled confusion of human affairs” (I guess I am not the only one who would claim that the two phenomena are related). Deep learning can seemingly be applied to all these different areas, and yet we must first answer the question of where this “deep learning” idea came from. What does it mean? To start with, I think it has to do with metaphors.

Today, we will take a *deep dive* and *see* how metaphors can be *powerful tools* to *guide* our minds into new *insights*—but also to *lure* them into *fresh* misconceptions.

Metaphors are not a decorative element in elaborated speech, but instead an **economical instrument to save words and effort** by recycling old words into a new context.

Metaphors are everywhere; I easily crammed seven into the last sentence (marked in italics). To recap, a metaphor applies words and concepts belonging to a certain field in order to talk about a quite different field. With “Elvis is the king of rock and roll,” one could say instead “most dominant artist in” or invent a new phrase with that meaning. But the first alternative is clumsy and the second leaves us with an abundance of words. We would then be faced with a similar quandary when trying to find a new phrase for “king of pop” to describe Michael Jackson. Evidently, metaphors are not a decorative element in elaborated speech, but instead an economical instrument to save words and effort by recycling old words into a new context *en suite* with all the associations they bring with them.

Part I: The “learning” in “deep learning”

According to most dictionary definitions, “learning”—defined by Merriam-Webster’s as “to gain knowledge or skill by studying, practicing, being taught, or experiencing something”—is something that humans do. So when attaching the word “learning” to things such as animals, substances, or even device systems in our Internet of Things world, this is already metaphorical, as it applies a human concept involving consciousness—something that these other things don’t have. You may have heard about so-called shape-memory alloys. Things made from these metals have an interesting feature: when you bend them from their current form into a new one and then heat them up, they will revert back to their original form. It’s tempting (and also helps with conceptualizing) to describe this behavior in metaphorical terms. Wikipedia employs this method to describe shape-memory alloys, conveniently marking its use of metaphors with quotation marks:

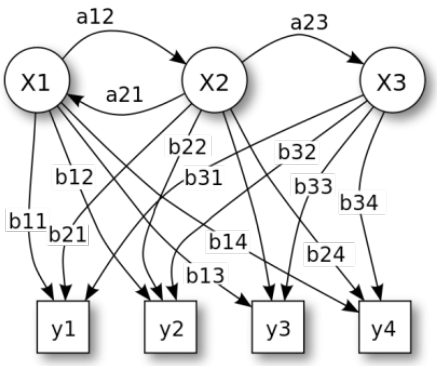
Training implies that a shape memory can “learn” to behave in a certain way. Under normal circumstances, a shape-memory alloy “remembers” its low-temperature shape, but upon heating to recover the high-temperature shape, immediately “forgets” the low-temperature shape.

I don’t think anybody would take any of this at face value and assume the metal atoms have little brains “learning” and “remembering” something. But how about computer programs that do “machine learning?” Is it also purely metaphorical “learning” that they do? Or are they complex enough to display true learning, like humans do? And why don’t we reject this latter idea right away, like we do for the memory alloy?

One reason, of course, is that computers are more complex and many people don’t understand them well. Computers were being spoken about in metaphorical terms right from the start, referred to by the media as “electronic brains” in the 1950s and onward. Then, science fiction took over and, not being tied by “technical feasibility” and other boring details, presented us with an abundance of “thinking” machines and robots that subsequently took hold in popular culture. There, they met concepts of “artificial life” rooted in western culture, from the golem made from clay and the homunculus of the medieval alchemists all the way to Mary Shelley’s Frankenstein’s monster.

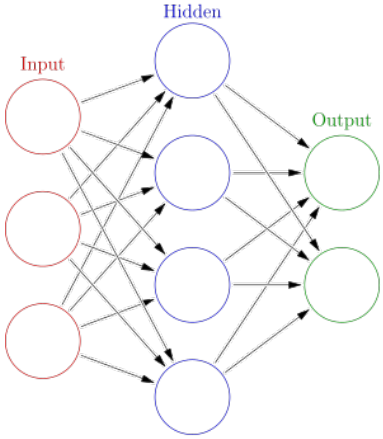
In order to decide whether machine learning (ML) really learns or just “learns,” here’s a quick primer on the subject.

Let's start with a mathematical model that formed the backbone of many ML systems for many years, Hidden Markov Models, or HMMs.



The image above seems to be of modest complexity, made up of states (x) and possible transitions between them (a), which come with certain probabilities and mappings (b) to input states (y). Probabilities are "learned" by the model in "training" on many samples of whatever the models are supposed to represent, for instance words or their acoustic building blocks, which are called phonemes. I think we can agree that "learning" is metaphorical here, as these models aren't really that different from the atoms of the memory alloy above.

However, a few years ago in mainstream ML, HMMs were swapped out and replaced by a different model type, one that first gained popularity in the 1990s. This replacement nearly disappeared after a while but is now making a forceful return to the stage (we'll discuss why a little later). Problems start with the name of the model: neural network (NN). As you can see in the figure below, it comprises layers of nodes, and these nodes are supposed to be "inspired" by neurons, such as what we find in a brain: they have input coming in through the arrows on the left, similar to how neurons get (electrical) input through their dendrites, then some calculation happens, and the resulting output leaves to the right (and becomes input for the next layer), such as through the axon of a neuron.



The calculation in the body of the "neuron" is typically rather trivial, like taking the maximum of inputs or summing up inputs. In order to use this for an ML task—for example, image recognition—you assign each input node to a pixel of a, say, black-and-white picture, and each output node to a category of an object you want it to be able to recognize ("tree," "cow," etc.). Then, like with HMMs, you train the model with pictures and known correct results (the picture shows a cow) by setting input and output nodes to the appropriate values. Then you apply a method called "back propagation" that calculates from right to left (at runtime, your model works left to right, from input to output), adjusting probabilities attached to the arcs, so that if input of this nature occurred, it would result in triggering the correct output when calculating left to right.

As you can see, the whole model is not much more complex than HMMs, or at least not so much as to justify that all of a sudden we should accept that NNs can think or learn. Granted, real models have more nodes (several thousands), but still, the differences to real neurons in real brains are fundamental: the latter still has many more neurons, works in an analog way, and combines electrical with chemical and even genetic effects, and yet we don't know how things like consciousness come about. In my view, NNs aren't any more (or less) likely to mimic brains than are HMMs. But because of the "neural" in the name (unfortunately, alternative names such as "perceptron," with "perception" in it, aren't much better), the model carries a big rucksack of metaphorical meaning like we saw above: isn't it natural that an "electronic brain" using artificial "neurons" can "learn?"

Part II: The "deep" in "deep learning"

But wait—that doesn't even take the "deep" in "deep learning" into account. In its verbatim meaning, "deep" is tied to spatial relations. Water can be deep, especially in lakes and seas. Other uses are nearly always metaphorical, like with the colors "deep red" and "deep blue" that are intense and/or dark variants, nothing more. And of course combinations with "thinking" are quite popular: you have "deep thinkers" coming up with "deep thoughts" after a "deep dive" into the matter.

In his "Hitchhikers Guide to the Galaxy," Douglas Adams names a computer (it works for 7.5 million years to return "42" as the answer to the only question it can answer) "Deep Thought." Later, a student and future IBM employee borrowed that same name for his real-life chess-playing computer, which gained fame in 1996 by beating world champion Gary Kasparov. And then someone in IBM's marketing department rebranded it "Deep Blue"—here you see the whole metaphorical path from "deep sea" to "deep blue" (like the sea, but also

IBM's logo) into "deep thinking" in just two words. In actuality, there isn't much that is "deep" about a chess-playing computer: the algorithm is fairly "shallow" and brute force; Deep Blue was successful because it used a lot of hardware for calculating possible moves and chips cleverly designed to help with evaluating positions. Nevertheless, the "deep" was here to stay. (For what it's worth, Deep Fritz is a competing chess computer that is still commercially available today.) DeepQA was another foray into the world of "deep" things: text-passage retrieval of mostly Wikipedia-originated text collections and ontologies packaged so that a machine could beat the human challengers in "Jeopardy" (under the name Watson).

The scene was set years ago when ML researchers decided they would expand the middle, "shallow," hidden layer in a neural net to multiple layers and make the nodes a bit more complicated, naming the result deep neural networks (DNNs) or deep belief networks. How would the unassuming, nontechnical person hearing about deep belief networks not apply the metaphorical associations they have been used to and connect this with "deep thinking" and artificial intelligence?

OK. I think we have dissected this and demystified it sufficiently to see that what we have in front of us is a mathematical way of modeling that is not completely different from HMMs. And hence, despite the name, these systems can only "learn" in a metaphorical way. So, does it mean we should think lowly of it? Not at all!

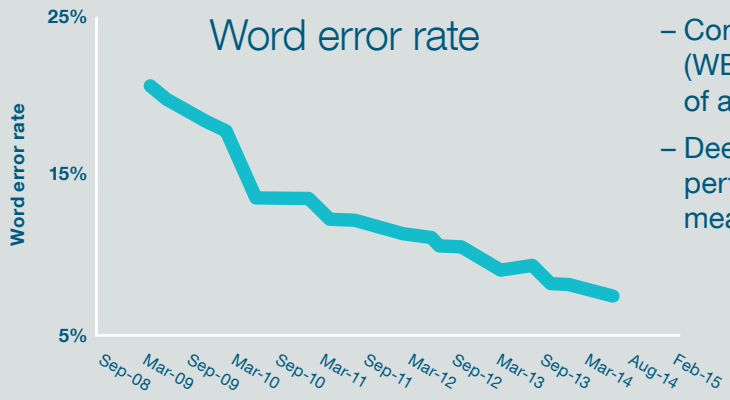
First of all, DNNs help us at Nuance drive accuracy up (and error rates down) for our core Automatic Speech Recognition (ASR) engine—the technology behind our cloud-based offerings as well as inside Dragon® NaturallySpeaking, which is currently in its 13th generation. Over the past 20 years, error rates continue to decrease, version after version. Achieving this within the HMM framework was getting increasingly difficult in the end, as this framework had been tuned and improved over decades and headroom was getting smaller.

DNNs have been the **single-largest contributor to innovation** across many of our products in recent years.

So, not only did DNNs drive error rates down at once, but because there is such a huge space of largely untested possibilities under the umbrella of DNNs—different topologies, the numbers of layers and nodes, how the nodes are structured, how the nodes are trained, etc.—they also promise a lot of potential for the years to come. And in speech synthesis, DNNs improve the mapping from the linguistic features of the text to be synthesized into the acoustic parameters of the target speech, like prosody. In voice biometrics, they help improve the accuracy of speaker authentication. With all this in mind, it is no overstatement to say that DNNs have been the single-largest contributor to innovation across many of our products in recent years.

When I spoke of DNNs as not being complex (in the sense that it is hard to see how consciousness and true intelligence would hide in them), I did not mean that they were easy to find, or better, easy to get them to work. Again, quite the contrary. As mentioned before, neural nets were around already in the 1990s, but two problems limited their success back then.

For one, when you wanted to train them on large data sets, and when the number of nodes and layers was nontrivial, the training would take very long—prohibitively long on hardware as it existed back then. Moreover, the training could end up in a model that was better than similar models "in the vicinity," but if you looked at the global search space, there would have been quite different and much better configurations. Whether or not



– Continuous server ASR word error rate (WER) reduction ~18%/year: combination of algorithms, data, and computing

– Deep learning (DNNs) is driving recent performance improvements in ASR and meaning extraction.

your training ended up in such a “local optimum” or truly found the global optimum depended on random factors during the early stages of the training process.

The breakthrough of DNNs was made possible when both problems were solved by pioneers such as Geoffrey Hinton and Yoshua Bengio. Of course, better hardware helped them solve the first problem, but it was clever ideas about how to parallelize the work better and how to use graphics processing units, or GPUs (i.e., special chips originally developed for computer graphics), that took them further. Even before, the problem of local minima was solved by introducing the concept of pretraining, that is, a processing step that presets the model into a state that is more likely (and faster) to end up in a global optimum than when starting from scratch.

What’s next?

The great thing is not only that these problems were solved and that NNs now work in general, but they have also opened up fields for additional research that promise more improvements for the future. GPUs get ever more powerful, driven by the games industry, and DNNs get a free ride. Speeding up training times is not only important for practical applications, but indirectly it also helps progress on the algorithmic side: when DNN trainings took several weeks or months to complete on meaningfully sized data sets (as they did until a few years ago), experiments were very costly and progress was slow. Now that you can turn around these trainings in days or even hours, it is much easier to test new ideas.

Even with all this progress, I, alongside other researchers, acknowledge that more work needs to be done. For example, using GPUs for all training steps of a DNN is a challenge because of the intertwined nature of the network. Because the output of a “neuron” potentially depends on many other neurons and the input data, and the training is not a purely local matter (and hence easily parallelizable), a lot of data needs to be transferred between compute nodes, potentially eating up the time advantage of the GPUs. How will we solve that? Also, when DNNs first took over the “backbone” of speech recognition, the speaker-independent model trained off a large quantity of data that reflected nearly all the variety of dialects and individual speaking styles possible. The challenge here is that most practical systems use a second, speaker-dependent training method that adapts the base model to the specific speaker. Depending on whether you have only a few seconds to train or hours of speech samples to pull from, different methods have been used. As all these were developed for HMM-like base models, they now need to be adapted to DNNs.

And so on, and so on.

Clearly, a lot of work still awaits us in the field of DNNs, but with that a lot of excitement too. Even if we don’t get carried away by the metaphors around “deep learning.”

Sources:
– deep-learning-reduces-ASR-error-rate © artificial-neural-network CC BY Wikipedia
– hidden-markov-model CC BY Wikipedia

The future of healthcare: how ambient clinical intelligence will drive better care

By Joe Petro, Senior Vice President,
Research and Development, Nuance Communications

Every day, people rely on technology to stay connected with other people, information, and content. For all the positives that technology has brought to healthcare, it's also been a curse when it forces physicians to do something unnatural. Technology should be used to support physicians and their workflows—not the other way around. We envision a future where amplifying human intelligence and leveraging existing capabilities will create ambient clinical intelligence that will finally decrease the time physicians spend doing tedious tasks such as clinical documentation, and instead allow them to spend more time face-to-face with their patients.

As we close out another year and look toward a new one, it's only natural to wonder what the future has in store for healthcare. Whether thinking about the changes to come in 2017 or looking at the bigger picture over the next few years, it's easy to see that major advances in technology are driving real improvements in healthcare now and into the future.

One of our goals is to improve both the physician and patient experiences. Electronic health records and similar technologies are starting to provide more value, but the use of these platforms has increased the time physicians spend on clinical documentation and has complicated their workflows, causing major physician burnout. We need to look at solutions that will return physicians to the job of providing care, which in turn will improve the care patients receive.

At Nuance, the way we envision advancing these areas in the future is by amplifying human intelligence and leveraging our existing capabilities to create ambient clinical intelligence to anticipate and assist humans (physicians) while technology operates unobtrusively in the background. Creating this type of intelligence will ease the documentation process and support physicians and patients in a personal healthcare experience, allowing them to engage and avoid workflow interruptions. We think that the support of intelligent and informed decisions using technology and content to amplify human intelligence will change how healthcare is practiced and will improve outcomes for both patients and health systems.

By saving physicians from tedious and complex labor, and amplifying the considerable capabilities they already have, we will achieve our goal of allowing physicians to get back to what's truly important: caring for their patients.

"We envision a future where amplifying human intelligence and leveraging existing capabilities will create ambient clinical intelligence that will finally decrease the time physicians spend doing tedious tasks such as clinical documentation, and instead allow them to spend more time face-to-face with their patients."

Joe Petro, Senior Vice President
Research and Development, Nuance

Visit nuance.com/healthcareinsights for resources, trends, and insights that explore the connection between satisfied clinicians, productive organizations and positive patient outcomes.

About Nuance Communications, Inc.
Nuance Communications, Inc. is a leading provider of voice and language solutions for businesses and consumers around the world. Its technologies, applications and services make the user experience more compelling by transforming the way people interact with devices and systems. Every day, millions of users and thousands of businesses experience Nuance's proven applications. For more information, visit www.nuance.com/healthcare or call 1-877-805-5902. Connect with us through the healthcare blog, What's next, Twitter and Facebook.

NUANCE

HC_4144 FEB 2017