Deep Learning in Depth: Deep Learning versus Machine Learning

featuring Ritwik Gupta and Carson Sestili as Interviewed by Will Hayes

---

Welcome to the SEI Podcast Series, a production of the Carnegie Mellon University Software Engineering Institute. The SEI is a federally funded research and development center sponsored by the United States Department of Defense and housed here on the campus of Carnegie Mellon University. A transcript of today's podcast is posted on the SEI website at sei.cmu.edu/podcasts.

Will Hayes: Why don't we start with a little bit of a definition of what deep learning is, and perhaps maybe a little bit about what it isn't? Because I think there is a lot of material out there on the web for people to look at. Why don't we start with you, Ritwik and then go to you, Carson.

Ritwik Gupta: Sure. So, like you said, deep learning—there is a lot of information about it out there on the Internet— sometimes what people get confused about is *is deep learning different than what machine learning is?* I think it is very important to start at the start of things, which is, deep learning is a subset of the wider field of machine learning.

Generally it has been the case that a lot of people have heard of things such as SVMs and linear regressions, some basic—people call them *shallow models.* Now they are saying, *Oh, deep learning. That sounds deeper and better. So, it must obviously be more accurate and the end all, be all to everything*. That is not the case.

While traditional shallow learning, and what happens in that sense is, *I want to make some inference about some environment, or some state of the world that I want to learn about*. What I do is, I as a scientist, go out and I collect a set of features, right? An example of a set of features in this situation might be the amplitude of my voice, the tone, the pitch. I would have to specifically extract those features and put them in a data set that a computer can understand. Then we would have a simple model, or even a complicated model, that would then learn some stuff on that and do inference, regression, classification, whatever.

Compared to that, what deep learning is…Deep learning is traditionally three main things. One is that it is a composition of a series of non-linear filters. So, filter being the statistical term. So, any kind of non-linear function composed together end to end to end to end, that learns its own representation of the world.

So, unlike shallow learning, in which I would have to physically give you a set of features, "*Here is the features for you to learn on top of*". The point of deep learning is that it will learn its own representation of the world that it's supposed to be learning. So, it would learn whatever is important for that specific environment and then do inference on top of it.

Will: So in the shallow learning example you started with, I would give you a recording of a voice and then I would indicate to you that, *when the voice sounds like this, it means this emotion is present. Or when the voice sounds like this, it means this communication is being intended*. Whereas in deep learning, you don't make that link for the system, the system derives that link?

Ritwik: So, that's a different part. What you are talking about in general is just the whole topic of dataset labeling and supervised versus unsupervised learning. What this is – I will build a simpler example. Let's say I'm observing bees flying back and forth from a hive to a flower, back and forth from a hive to a flower. I would tell the machine learning algorithm a feature set, such as the height of the bees' flight.

I would give it meters off the ground at certain points of the flight. I would tell it the color of the bee at different points – if it changes. It is a magical bee. I would tell it the distance to the flower. It is not examples of different bee flights, it is specific metrics about that environment. It can be metrics. It could be something more abstract, but I define a feature space for the environment, and then give that to the shallow learning model.

Carson: If I can unpack that a little bit with your audio understanding example. To think about the idea of what a feature is here, it would be reasonable for a human engineer to want to make a claim to the computer about maybe, *When the pitch of your voice takes this certain contour. Here is a set of 20 different contours that we, the humans, believe is meaningful for this language of speech*. Or maybe *this kind of filtering of white noise or something is useful for speech in this way. This kind of sentence structure is useful in this particular way*.

You are still going to give data to a machine learning algorithm or a deep learning algorithm. You are going to say, *Here is the data set, and here is the right answer*. But the difference is going to be, in a deep learning system, the reason they exist and the reason they shine is what is very hard to describe what the right features might be.

In this case, say I don't know anything about the language that I'm studying. I don't know that maybe a rising pitch at the end of the sentence has any kind of semantic meaning. But I do have

a million labeled examples of a sentence that ends like that, and I know that there is something in common between them. The point of a deep learning algorithm is that it can infer the fact that that rising pitch was important. You did not have to tell the system that in the first place.

Will: Perhaps another example to make it even more obvious for our audience, when my mother saw my firstborn child, she looked down at the crib and said, "*Yes, that's one of mine"*. Was it because my child has the same nose I have? Was it the shape of the eyebrows? Was it the shape of the jawline? Those are things that we could talk about as features.

The difference between shallow learning and deep learning is, in shallow learning, we would tell the system, "*These are aspects that you need to care about"*. In deep learning, that's inferred. Just as my son had to infer what his grandmother looks like, never being told before he was verbal, that these are features that you care about. His way of learning about who that person is looking down at the crib was much more in line with deep learning than shallow learning, where I say, *When you're studying for this book report*, now that he's in sixth grade, *these are the kinds of thing you should attune to* and so I'm telling him features. It's a different kind of learning.

Ritwik: Correct. And what your son did, was learn a representation of his grandma, AKA representation learning, which is what deep learning models do, representation learning.

Carson: There's something else that Ritwik mentioned in his characterization of deep learning that I think is really good, which is this sequence of non-linear transformations. For me, as a mathematician, I perfectly understand what that means, but in case that isn't clear, what you are doing is…Passing through a filter will take your input data and change it slightly to exaggerate important parts of it. In an image, what that might do is to find edges, horizontal edges or vertical edges, to pick out what is the boundary of things. Or, it might actually be a blurring. If it turns out that your image has a lot of noise in it like TV static, blurring might be a really useful way to transform that image to get rid of the stuff that doesn't matter. You can do that once, but you can also do it 100 times. Every time that you are applying this transformation, you are pulling out the parts of the data that are interesting, that matter for the problem at hand. So that is really what a transformation is, but the non-linearities really gives it mathematical power to create good computation.

Will: So you get a wider range with each of the filters than you would with the linear application?

Ritwik: Correct. That is not to say that shallow learning doesn't do non-linear transformations. It is just that, let's say I have one non-linear transformation, like a sigmoid function, right. It kind of looks like that. Let's say we just have that as a shallow learning model. It can only represent data that kind of looks like that. But imagine I stacked, composed, a whole bunch of those

together. The more non-linear filters that are composed together…you can imagine me putting kinks in a ruler, right? The more kinks I put in, the more I can approximate much more convex, very varied shaped functions, right?

That is the idea. As you compose more and more non-linear functions together, you can represent a much wider function space than you could with just one non-linear function. That is why deep learning is different from shallow learning. Shallow learning doesn't compose multiple things together. Deep learning does.

Will: As I was preparing for this podcast, one of the things I thought of is the process of communicating to somebody how someone looks. So if I was describing over the telephone to a distant cousin of mine what my mother's appearance is, because this distant cousin is meeting her at a train station, I would start by talking about features of the face that are known to be germane to such a description. That might be shallow learning. How does deep learning differ from that?

Ritwik: So in shallow learning—if I am using something like shallow learning to explain to you what my mom looks like—I would say, *She has a chin that looks like a 75-degree V. She has eyebrows that are about three centimeters wide and two centimeters thick, and a nose that is about this sharp*. In deep learning, I would say, *Oh, she looks like me. She looks like my brother. She looks like my sister*.

Will: Ahh. Reference points.

Ritwik: Yes. What she would have to do is pull all of you together in her head and build her own representation of what your mom should look like. What are common features between you, your siblings, etc. that your mom would share that she can then identify your sister. That is shallow learning versus deep learning.

Will: You gentlemen both work for different elements of the Software Engineering Institute. You [Carson] are at CERT, and you [Ritwik] are with the Emerging Technology Center. But deep learning has applications in both of these. I imagine you are not necessarily just working on one project together. Could you talk a little bit about maybe what is happening at CERT, and then what's happening at ETC?

Carson: Yes. So, I think it is actually really important to impact the distinction because that took a long time for me to understand the meaning. CERT was actually the—please fact check me on this—but CERT was the United States' first cybersecurity group. It was formed in response to the Morris worm, which was a super bad computer virus.

All the projects that we do, in order to get funding, need to be pitched under the lens of, *This is useful for software security for the defense of the United States*. So all of my data science work, and all my machine learning work is involved with a cybersecurity application. Right now I am working on projects that investigate the utility of machine learning for code in various parts of its development cycle, for analysis of software in various parts of its development cycle.

Will: So, new frontiers that you push are frontiers relating to cybersecurity and really have a nice focusing effect on what you are doing. And you still get to work with a colleague who has a different filtering effect based on his affiliation.

Carson: Exactly. I think because I had this image-processing background from the brain imaging lab, I was fortunate to have the opportunity to work with Ritwik who can now tell you about what the ETC does.

Ritwik: Sure, yeah. Again, CERT does some really, really cool world class cybersecurity stuff. So, the ETC [SEI Emerging Technology Center] is a bit different. We are very new to SEI. We were founded in 2013. Our focuses lie in three main areas, them being human machine interaction/machine emotional intelligence *(How do we get a machine to better understand emotion, the emotional state of a human, we can work better together) and* applied artificial intelligence/machine learning. So, this is AI applied to some task in the world. So, that is a very broad description. We tend to not do cybersecurity, because that in CERT's domain. So, we do things like satellite imagery, voice articulometry, etc. And then, the third one is advanced computing. So, *how do we push the fabric of computing further*? So, as the paradigms of computing change from CPUs to SIMD architectures and GPUs and TPUs, what's next? We are working on the *what is next* as well. So, that is our focus areas. You can imagine, there is a lot of applications for deep learning in all of those aspects.

Thank you for joining us. Links to resources mentioned in this podcast are available in our transcript. This podcast is available on the SEI website at sei.cmu.edu/podcasts and on Carnegie Mellon University's iTunes U site, as well as the SEI's YouTube channel. And as always, if you have questions, please do not hesitate to send us an email at info@sei.cmu.edu