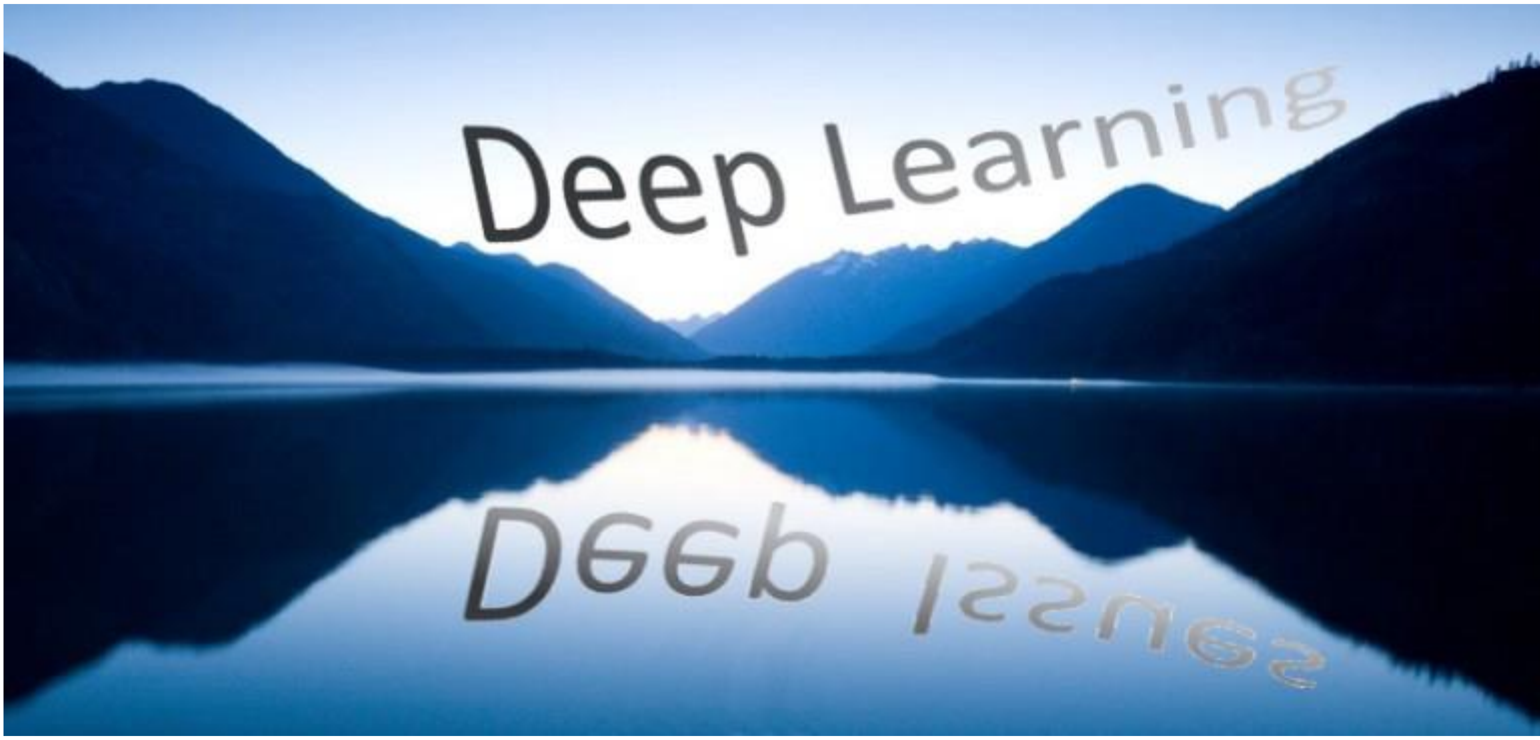
Deep Issues Lurking Under Deep Learning:

Reflections on Coursera Specialization on Deep Learning by Andrew Ng



Richard Hackathorn, Apr 3, 2018
https://towardsdatascience.com/deep-issues-lurking-within-deep-learning-f923a96564c7

> TL; DR—I highly recommend this MOOC specialization. However, be aware! It is more than yet-another certificate. Deeper issues lurk beneath the surface, causing you to think deeper about your responsibilities using this new technology.
>
> Note to reader (2018–06–07): Article has two parts. First part describes the MOOC course for potential students of DL. Second part explores the "So What?" implications, especially the responsibilities for practitioners of DL.

I recently finished the Coursera specialization on Deep Learning (DL) with its 5 courses, 14 class weeks, 180 videos, 29 labs, and 5 'hero' interviews. [01]

My motivation was to learn a new skill, which I did. However, I also came away with many questions and concerns, along with a list of projects to investigate. If you commit to this specialization, you will receive more than yet-another certificate! Here are my reflections about the experience and its significance within the larger context.

My Experience

Andrew Ng of DeepLearning.ai and Stanford taught the course in a personable style, incrementally revealing details as appropriate. He has the gift of understanding the mind of students as they struggle with this complex subject. His mission is to train millions of professionals in the use of AI tools "so they can go and invent the things that no large company … could do." [02]

© Bolder Technology, Inc., 2019 – Attribution 4.0 International (CC BY 4.0)

I had taken Ng's Machine Learning course in 2013, using Octave, and was excited to go deeper with Python. In September 2017, I started the first of five courses. I immediately noticed that Coursera had made many learning improvements in their MOOC infrastructure over those four years. The following are the memorable elements of my learning experience.

Know Thy Math!

Like the first course, the topic flow started with the basic math and its translation into working code, using the classic example of logistic regression. However, I noticed more emphasis on notation (especially weird superscripts) like we were building something more complex. *I was correct!* I soon learned that NumPy was your friend, which enable restructuring code for vectorization. *Bye-bye FOR loops. Well, not entirely.* I eventually learned that vectorization is at the heart of smashing tensors for deep neural networks, enabling computational power necessary for practical applications.

Retro Chalk-Talk

The videos were short, averaging seven minutes. Ng conducted the lecture in the retro chalk-talk style—start with a clean slate and write everything in detail *…with barely legible script.* I soon adapted to his style by capturing the final screen at the end of each segment, pasting into PowerPoint, and adding more comments. Also provided is the MP3 video and rough TXT translation.

Just-Enough Quizzing

Each week, the lectures ended with a short multi-choice quiz. Initially the quizzes were annoying, but they proved to hit the key points and was just enough to ensure that I was paying attention to the videos.

Hero Interviews

A delight was the 'hero' interviews in which Ng would interview DL researchers about the history and milestones of deep learning. *I missed the hero videos in the latter courses! Did Ng run out of heroes?*

Fill-in-The-Blanks Labs

The best part was the labs, consisting of fill-in-the-blanks within a specified structure. This was also an excellent example of using Jupyter notebooks for teaching, with clear detailed instructions inspiring the student to code and receive immediate feedback. Thus, the student was motivated to learn good practices for meticulously inserting debugging statements and for careful attention to unit testing.

The weak link with labs was the auto-grading procedure, which was curt and sometimes strange. *A great future application of DL!* This glitch forced the student to spend extra time understanding the math-to-code translation and data flow within the algorithms. With much patience, forum comments, and SlashDot, I was able to find the solutions, usually in double the estimated time.

Soft Deadlines

I was occasionally faced with a looming deadline for submitting a lab, adding to the drama and satisfaction of completion. These deadlines were 'soft', just delaying course completion (without loss of prior work) until the next monthly cycle. I found that the deadline crunch to be beneficial for me, as with any busy professional.

To Pay or Not to Pay

You can take this entire specialization as an auditor for FREE! I started this way, until the first quiz. Auditors do not get their quizzes or labs graded. You can do the labs, but you just do not receive auto-grader feedback. At my professional stage, I do not need another certificate. But, I wanted the FEEDBACK! And, that feedback did make a difference in the quality of learning for me.

Best Practices

After the first basic course, I was anticipating the exciting stuff of processing images and natural language with sexy CNN and RNN models. To my disappointment, Ng focused the next two courses instead on practical aspects of neural network tuning (test/dev, regularization, bias/variance, mini-batches, hyperparameter tuning) and structuring development projects (evaluation metric, human-level performance, error analysis). *Bummer!*

In hindsight, these two courses were the more insightful of the entire specialization. Ng was mentoring us in the artistic side of deep learning, stressing the evolving best practices and its experimental nature. Combined with the 'hero' interviews, Ng was introducing me to the community of DL professionals. *Nice touch!*

Full Stack

At the end, this specialization provided me with the skills to design and implement deep neural networks on a full stack of tools from Python, NumPDoney, TensorFlow and Keras, along with familiarity with cloud-based programming environments, like Jupyter notebooks. Often I copied the lab notebook files (and other related files) into my local Anaconda3 environment and executed without changes. For addition compute power, I am installing a GTX1060 card. I am amazed at how convenient and approachable DL technology had become, both locally and cloud-based.

So-What?

Upon completing the DL specialization, I felt relief that my time was freed up and sadness that Andrew's chalk-talks were over. I had spent many hours learning this complex subject. And, yet, what did I learn of enduring significance? Was it just cute methods for classifying dogs from cats? Applying Van Gogh's style to photos of friends? Does Deep Learning have practical value and the potential to be socially redeeming?

After all this effort, so-what?

The question of significance kept churning in my mind. So, I went back to first lecture and compiled a complete outline of the specialization (in ten single-space pages!). Yes, there was a ton of material. But then, it hit me… *my 'Oh-My-God' moment!*

The significance of DL lies in exposing the fundamentals about thinking and learning, not by mimicking the human brain (which it does only superficially). But rather, DL enables humans to create algorithms that exceed 'human-level performance' in several challenging intellectual tasks. In other words, …

We humans have created tools smart enough to out-smart ourselves!

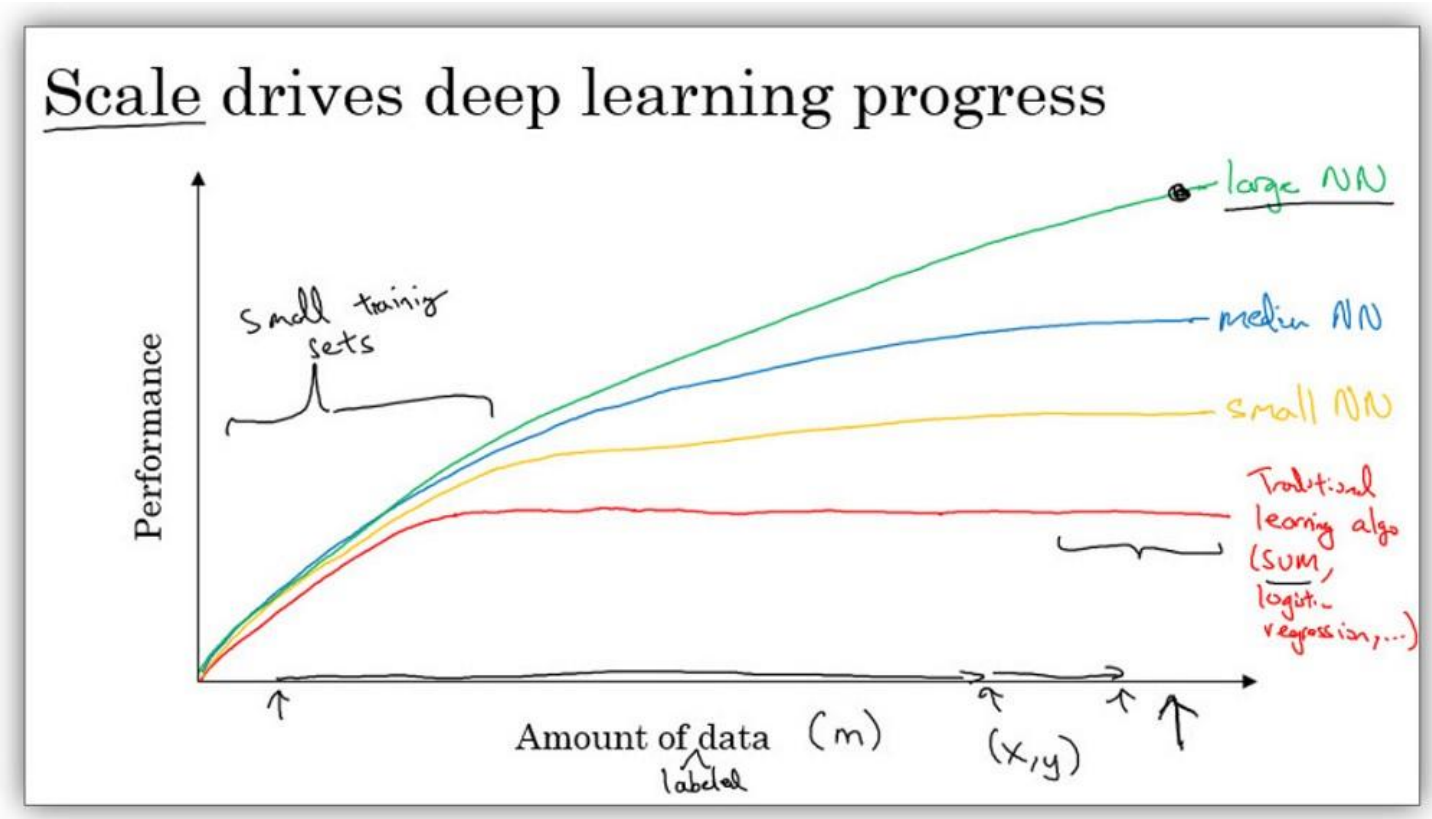
This is not the usual sci-fi story where super-intelligent robots decide to eliminate humans because of our inferiority. The real problem is the unintended consequences of embedded DL applications in larger systems, like Facebook, Amazon, Google, and other companies who impact many lives. Further, these DL applications will greatly multiply over the coming years!

My Insights

Here are several insights that I drew from the above?

Starts and Ends with the Data

A simple, yet profound, slide of Ng is titled, “Scale Drives DL Progress”, as shown in the following chart of Performance versus Amount of Data.



From Coursera Course “Neural Networks & Deep Learning” by Andrew Ng

For any predictive algorithm, the performance (or accuracy) of prediction is critically influenced by data volume. The data must be quality labeled data, where the observations correctly reflect reality and are tagged with a ‘goodness’ indicator. Crappy or mis-labeled data does not count!

Here are several points to note…

- More data always wins over less data! The company that controls the most data eventually wins over all other companies!
- At low data volumes, any predictive algorithm is as good as any other. Therefore, DL has no advantage; hence, use conventional machine learning algorithms (like random forest) since they are easier, quicker, and offer better interpretability.
- At high data volumes, the opportunity for better performance increase dramatically with deep neural networks.
- Performance is often ill defined. Prediction accuracy (correct predictions / total predictions) of test data is frequently used, but this can be misleading for many use cases. Dig much deeper here!

© Bolder Technology, Inc., 2019 – Attribution 4.0 International (CC BY 4.0)

The insight is… all is dependent on the data, which is the 'Achilles heel' of DL. Curate the data as a valuable asset.

Data is More than Numbers

During the labs, a subtle shift in thinking about data occurred for me. With a background in business intelligence, I have always conceived of data as rows (instances) and columns (features) in a big table of numbers. Dealing with images and music jars that mindset.

I distinctly remember Ng's Octave lab that processed images of single digits (from MNIST). The grayscale number of each pixel was thrown into a long linear array (of features), voiding any indication of x-y position for the pixels. I assumed that this exercise would not generate any useful results. *How could it?* However, this simple one-layer neural network achieved 82% accuracy. *Surprise!*

In the latest Ng's CNN course, we learned how to apply the convolutional transform, which does preserve positional information. The prediction accuracy was boosted into the high 90's, surpassing human-level performance. *Amazing!*

The point is that DL requires one to view data from two perspectives concurrently:

1. Numbers structured as matrices, ready for efficient computation on GPUs. Here the shape (vector of dimension sizes) is critical to allow the matrix calculations and, more importantly, to preserve the meaning of the matrices. Every time I messed up the sequence of shapes, it was because I forgot the meaning of the matrix, especially its mapping back to the math.
2. Pieces of reality, each with their unique qualities. Nature resists being digitized, whether as an image, sound, vibration, and the like. For me, this was apparent in the RNN course, dealing with sequential patterns within music. Often, I had to remind myself that these numbers were actually a jazz pianist tapping out a melody, digitized into the MIDI coding. Now, I could properly deal with perspective #1.

The insight is that data has this dual personality of both numbers and reality, whose full co-existence is difficult for anyone to sustain.

Bordering on Alchemy

DL is a proto-science struggling to find its theory. *Think of chemistry in the 1600's.* Because DL lacks a firm theory, researchers are trying crazy ideas for model structure, cost functions, data transforms, and so on. And, some of these crazy ideas have produced amazing results! This is exciting!

But, it is also bewildering to the new student of DL. So, when a student asks, "why use inception networks", the answer is often "because it generates a greater test accuracy". A truthful answer should be "because we are trying everything possible since we are not sure of what to do".

As discussed in the 'hero' interview with Geoffrey Hinton, the proto-science struggle is most apparent with Stochastic Gradient Descent (SGD) combined with back propagation. It works, but how does one understand and optimize its processes within a nontrivial network? Hinton stated that he is "deeply suspicious" of SGD, which perhaps should be "thrown away and started again". He is emphasizing, instead, research into unsupervised learning that may mean "getting rid of back-propagation". [03]

The insight is that DL is an experimental proto-science, lacking a firm theory to guide it. Embrace it! Contribute to its maturity as appropriate. Meanwhile, realize that one has great responsibility to be firmly grounded in constantly testing the accuracy of algorithmic performance based on good data.

Think about appearing in a court room to explain why your DL algorithm that controlled the autonomous vehicle caused a fatality. Hence, be obsessive and even paranoid about the validity of your model! [04]

Meta-Learning as the New Secret Sauce

It seemed to me that DL is moving rapidly beyond SGD into murky areas that some labeled as 'meta-learning'—becoming aware of and controlling the learning (model training) processes. [05]

As glimpses of meta-learning, I was especially fascinated with Ng's lectures and labs for:

- **Face Recognition**, reusing pre-trained models to 'transfer' its weights to a new application.
- **Neural Style Transfer**, teasing the cost function to balance content with style activations.
- **Jazz Solo**, tricking a pre-trained model to generate a likely sequence of input data.
- **Debiasing Word Vectors**, detecting and correcting sexual bias with analogies.
- **Language Translation**, enhancing by managing the attention placed nodes.

My fascination has motivated me to learn about various meta-learning approaches, such as:

- **AlphaGo Zero** demonstrating that good simulations of the problem domain (like the game Go) can be surrogates for generating labeled data, using Generative Adversarial Networks (GAN).
- **Capsule Networks** (CapsNet) introduced by Hinton to correct flaws in image classifier.
- **CoDeepNEAT** optimization of DNN typology, components, and hyperparameters via genetic evolution algorithms.

The insight is that one should eagerly explore the new secret sauce for DL—meta-learning.

Potential for Smart Scalable Learning

As apparent in these DL courses for me, MOOC infrastructure has matured over recent years. As I viewed the videos, took the quizzes, and worked on the labs, I generated lots of click data, which could be labeled with quiz scores and lab completions. Given the activity on the forums, I was one of hundreds (perhaps thousands) of other students. With continuing cycles of course offerings, the accumulation of this click data will be very useful for guiding the learning process. [06]

Could we design and train a DNN model to guide effective learning customized for a new student? Think of this learning process as an adventure game of journeying to a destination. Along the way are many puzzles to be solved. And, many paths are possible. Get to the destination, and you get the certification.

Not just for the topic of Deep Learning, but this approach could be applicable for thousands of other topics. Further, if customized learning for one student could be provided, then this customized learning could be scaled to thousands of new students, eventually maturing into global learning services.

The insight is… Do we now have the technology to provide smart scalable learning resources for the entire world? *The 'teaching' Wikipedia? Turbo-charged Kahn Academy?*

My New Super-Power

In concluding, my mind returns to the social and ethical implications of DL. In his final video, Andrew Ng congratulated his students on completing this specialization and urge them to further their careers and pursue their dreams.

© Bolder Technology, Inc., 2019 – Attribution 4.0 International (CC BY 4.0)

Going further, Ng compared the learning of Deep Learning to one who has acquired the super powers to enable computers to see, to synthesize art & music, to translate languages, and to diagnose radiology images. He then urges:

Do whatever you think is the best work you can do for humanity.
The world today has its challenges. But with the power of deep learning,
I think we can make it a much better place.
Now that you have this super-power, I hope you will use it to go out there
and make life better for yourself, but also for other people. [07]

His point is…

Great ability necessitates great responsibility.

Deep Learning bestows upon us a great power and **also** a great responsibility to be used wisely and for the good of humanity.

As a positive substantive example of promoting this responsibility, Francois Chollet recently argued that Deep Learning tools should not be used to manipulate people. Instead, Deep Learning should give people the control over those tools to pursue their own goals and passions. [08]

Some experts have compared Deep Learning to other major inventions of mankind, like nuclear energy. If valid, this comparison is a very sobering thought.

References

[01] Coursera, “Deep Learning Specialization”. link All five courses are now cycling monthly. For an overview of its content, see the excellent 28-page slideshare notes by Tess Ferrandez, March 8, 2018. Also, Ryan Shrott’s article on “11 Lessons Learned” is a good sampling of concepts taught in these courses.

[02] Knight, “Andrew Ng’s Next Trick: Training a Million AI Experts”, MIT Technology Review, August 8. 2017. link Interview focusing on Ng’s online classes and his goals of educating new AI experts. An excellent 29:18 video of Andrew Ng talking about the State of Artificial Intelligence by the MIT Technology Review. link One comment stated “Ng is the perfect canary in the coal mine. As long as he isn’t freaking out, we should be relatively safe.”

[03] LeVine, “Artificial intelligence pioneer says we need to start over”, Axios, September 15, 2017. link

[04] Moore, “Deep Misconceptions About Deep Learning”, Towards Data Science, January 26, 2018. link Practical best practices for developing DL applications. *I am paranoid! My model is purposefully deceiving me! It is out to get me!*

© Bolder Technology, Inc., 2019 – Attribution 4.0 International (CC BY 4.0)

[05] The definition for meta-learning was crafted from this Wikipedia entry.
Less-dense articles on DL meta-learning are appearing, such as:
* Perez, “Taxonomy of Methods for Deep Meta Learning”, Intuition Machine, March 5, 2017. link
* Rodriguez, “What’s New in Deep Learning Research: Understanding Meta-Learning”, Towards Data Science, March 15, 2018. link
* Miikkulainen etal, “Evolving Deep Neural Networks”, arXiv, March 1, 2017. link CoDeepNEAT optimizes DNN typology, components, and hyperparameters via genetic evolution algorithm.

[06] In Note #02 above, Andrew Ng commented that Coursera has examined the MOOC interaction data for “Automating Education Itself”.

[07] Ng, “Conclusion and Thank You” video, Coursera Course “Sequence Models”, Week 3.

[08] Chollet, “What worries me about AI”, March 28, 2018. link

---

If you benefit from this articles, please support my Patreon to create and mentor small peer groups of managers to explore key management issues of analytical systems enabled by Deep Learning. If this program might be of interest to a colleague, please share a link or tweet. Thanks, Richard