HARMON ON BPM, January 2021

DeepMind Is No Longer Playing Games

Science has made a lot of progress in the last 50 years, and nowhere has it made more progress than in biochemistry. Starting with the discovery of the structure of DNA in 1953, our understanding of biochemistry, biology, and medicine has grown by leaps and bounds.

Once DNA's overall structure had been resolved in the 50s, biochemists proceeded to work out how the genes caused things to happen. Without going into a lot of detail, suffice to say that the DNA is transcribed to RNA, which then generates specific proteins, each made up of some combination of 20 different amino acids that cause various chemical actions in living organisms. As time has passed the focus, in many cases, has shifted from what genes an organism has to what proteins an organism can generate, and, even more specifically, what actions are caused by each specific protein. As the focus shifted to proteins, a specific problem arose. Even with a complete knowledge of the amino acids that make up a specific protein, the way the amino acids fit together in the molecule makes an important difference. Consider a protein that has two open sites for a specific amino acid. The amino acid could bind at either site, and knowing which specific site it has bonded with, in a specific protein, makes all the difference. When you consider that some protein molecules are made up of thousands of amino acids and there are many hundreds of different bonding possibilities for various amino acids, you begin to see the nature of the problem.

Interestingly, the first academic expert system was Dendral, a software application developed in mid-Sixties that took data from a mass spectrometer and analyzed the organic chemical compounds in a sample. This application was developed by Joshua Lederberg, a Nobel winning chemist, and Edward Feigenbaum, the computer scientist who

would become the father of expert systems. Suffice to say AI folks have been focused on understanding the nature of biochemical molecules for some time.

DeepMind is an AI company owned by Alphabet, Google's parent company. DeepMind got a lot of attention during the past decade by developing an application to play GO – an oriental board game that is widely regarded as the most difficult game in which both players have a complete knowledge of all the moves. DeepMind's program, AlphaGO first beat the European GO champion, and then, vastly improved after playing thousands of additional games, beat the Korean international champion.

Recently, DeepMind developed a program to play an online game, StarCraft 2. This game hides information about opposing player's moves and allows simultaneous play, making it much more complex than Go. In a short time the DeepMind program was defeating all but a few of the very best StarCraft 2 players. DeepMind has a reputation for successfully applying AI (specifically Neural Nets) to game playing scenarios. Recently, however, the company has focused on medical and biochemical applications to provide some practical uses for its technology.

AlphaFold 2 is DeepMind's latest application. It looks at data about biochemical makeup of protein molecules and suggests how they are structured (folded). The first AlphaFold was created in 2018 and rapidly reached its limits. The current version is a new design that its developers believe has a lot of room to learn and become more sophisticated.

DeepMind selected the problem of identifying the three dimensional shape of a protein molecule as a challenge. In essence, a protein is a molecule made up of amino acids. Figuring out how the bonds work to form a protein molecule is very hard and very time consuming. Computer programs have been developed to determine what amino acids comprise any given protein, but figuring out exactly how each amino is situated in a three dimensional space, and how the amino acids bond together has proved very, very hard. Ever since biochemists first focused on the protein folding problem, they have depended on defining the three dimensional shape of a protein using x-ray crystallography. This requires months of experimentation and is a very long and tedious process.

There are thought to be around 10x170 legal positions in GO – a number greater than the observable atoms in the universe. This

makes winning Go a classic AI problem. A computer can't solve the problem by brute force. The best a computer can do is employ heuristics – rules of thumb – than can reduce the search space down to a size that can be easily handled. Human experts develop heuristics. For a while, the best we could do is get the heuristics from a human expert and enter them into an expert system. Now, using neural nets, and learning algorithms, AI systems can learn heuristics by themselves. AlphaGO learned GO play by playing thousands of Go games and finding and capturing rules of thumb, in each case, that identified which moves seemed to work in a given situation, and which didn't.

Protein folding is a bit more complicated. It is estimated that there are as many as 10x300 different shapes that a reasonably complex protein could assume. The challenge is to develop a set of heuristics that can reduce this impossibly complex problem to a manageable size.

In his Nobel Prize acceptance speech in 1972, the biochemist Christian Anfinsen argued that, in theory, a protein's amino acid sequence should be predicted if one knew the amino acids involved. Anfinsen's hypothesis launched a five decade quest to develop a computer application that could predict a protein's 3D structure, based simply on its known amino acids. In 1994, in order to judge how effective new software programs were in identifying how protein molecules were folded, John Moult, a biologist at the University of Maryland, set up a biannual competition, termed the Critical Assessment of Protein Structure Prediction (CASP) competition.

In essence, CASP presents computer challengers with raw information about several recently defined protein structures -- defined by months of human experimental effort -- and asks the computers to generate the three dimensional structures.

DeepMind made a first attempt at CASP in 2018 with AlphaFold 1, and a second attempt this year with its new AlphaFold 2. Scores are based on how accurate the computers are, with 100% on all proteins in the contest. This year AlphaFold 2 scored 92.4 – an accuracy that CASP's founder, John Moult, says is roughly comparable with the best results that can be obtained by X-ray crystallography.

Several reviewers point out that, at the moment, AlphaFold seems to be limited to static proteins, and isn't yet very good at determining structures resulting from the dynamic assembly of multiple protein structures. On the positive side, AlphaFold 2 is the first application to

do nearly as well as it has, and it was able to quickly predict the structures of several proteins used by SARS-CoV-2 virus, which helped in vaccine development efforts.

AlphaFold 2 is not quite so far ahead of the field as AlphaGO was. There are many other research groups that are working to apply machine learning techniques to the protein structure problem. Exactly what DeepMind has done to seize a clear lead isn't yet understood. DeepMind has been very good about publishing detailed scientific papers about its previous work, and apparently a paper is in the works to describe AlphaFold 2, but it has not yet been published.

Figure 1, from a press release by DeepMind, provides an overview of AlphaFold's current architecture.
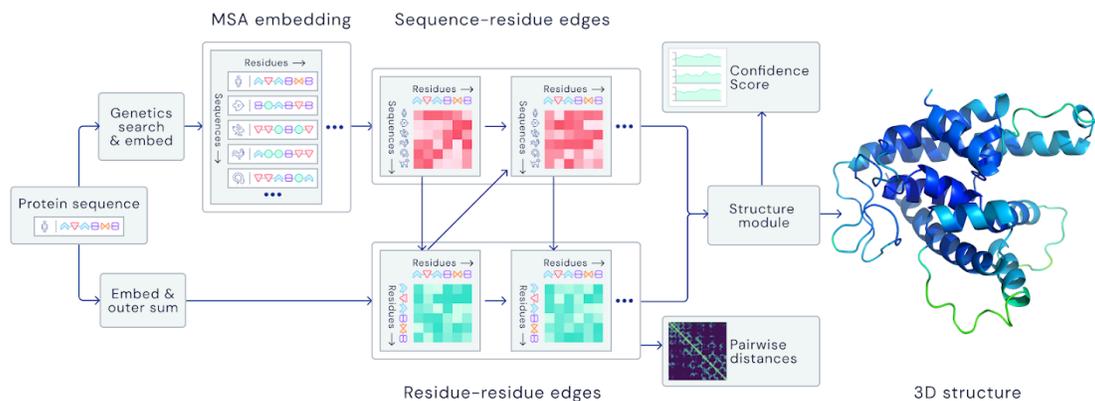

Figure 1. An overview of the main neural network model architecture. The model operates over evolutionarily related protein sequences as well as amino acid residue pairs, iteratively passing information between both representations to generate a structure. (After an article by DeepMind)

Recall what we know about neural networks and deep learning algorithums. The program learns by practicing with examples that can, in the end, tell it whether it succeeds or fails. There are about 170,000 proteins whose structure has already been identified by humans using X-ray crystallography. That's a small training set for a neural network, and presumably the DeepMind people have come up with some way to get more value out it. It will be interesting to see how DeepMind got around this problem.

Just as DeepMind did with AlphaGO, it has released a version of AlphaFold that is very impressive. In a year or so, having experienced

many more learning situations, AlphaFold is likely to be much better. Even as it is now, however, AlphaFold 2 represents a major achievement in the application of Neural Networks or Machine Learn to the solution of a world-class scientific challenge.

Unlike AlphaGO, where DeepMind trained it's application by playing it against human experts – and in effect, learning from humans – AlphaFold 2 learned by trying to predict structures and checking its results against correctly assembled proteins. Humans had defined the correctly assembled proteins, but not by applying theory. Humans had done it by tedious experimentation. AlphaFOld 2, however, developed and revised and corrected a set of heuristics to develop its answers – using heuristics humans have been unable to imagine. This suggests the kind of creativity that AlphaGO was able to demonstrate in a few cases where it developed powerful new sequences of play that human Go players had never tried. If DeepMind has found a way to develop powerful techniques to train new applications with much smaller data sets, the range of problems open to AI solution will be significantly enlarged.

Lots of business processes will be redesigned in the years ahead to accommodate the use of powerful new tools like DeedMind's AlphaFold. In a similar way, lots of companies will rethink problems they face and wonder if Machine Learning offers a way to significantly improve the speed the process currently requires.

Author

Executive Editor and Founder, Business Process Trends In addition to his role as Executive Editor and Founder of Business Process Trends, Paul Harmon is Chief Consultant and Founder of BPTrends Associates, a professional services company providing educational and consulting services to managers interested in understanding and implementing business process change. Paul is a noted consultant, author and analyst concerned with applying new technologies to real-world business problems. He is the author of *Business Process Change: A Manager's Guide to Improving, Redesigning, and Automating Processes* (2003). He has previously co-authored *Developing E-business Systems and Architectures* (2001), *Understanding UML* (1998), and *Intelligent Software Systems Development* (1993). Mr. Harmon has served as a senior consultant and head of Cutter Consortium's Distributed Architecture practice. Between 1985 and 2000 Mr. Harmon wrote Cutter newsletters, including *Expert Systems Strategies, CASE Strategies, and Component Development Strategies*.

Paul has worked on major process redesign projects with Bank of America, Wells Fargo, Security Pacific, Prudential, and Citibank, among others. He is a member of ISPI and a Certified Performance Technologist. Paul is a widely respected keynote speaker and has developed and delivered workshops and seminars on a wide variety of topics to conferences and major corporations through out the world. Paul lives in Las Vegas. Paul can be reached at pharmon@bptrends.comJ

References

Lindsay, Robert K., et al. *Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project.* McGraw-Hill, 198

Harmon, Paul. "Google's DeepMind and StarCraft 2" *Harmon on BPM* Column, www.bptrends.com, Nov. 4, 2019.

DeepMind AlphaF0ld Team. AlphaFold: a solution to a 50-year old grand challenge in biology. https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology Nov 30, 2020.

Markowitz, Dale. "AlphaFold 2 Explained: A Semi-Deep Dive" *Dale on AI*. daleonai.com/how-alphafold-works, Dec. 9, 2020.