Modeling Real and Theoretical Disease Outbreak and Spread

With Artificial Intelligence

Cameron Cooke

Department of Psychology, University of California Santa Cruz

PSYCH 139P: Natural and Artificial Intelligence

Dr. Alan Kawamoto

February 24, 2020

Abstract

The large amount of recorded data and statistics that are present within epidemiology make disease modeling an excellent candidate for AI-assisted generative modeling. Although the inference of how a pathogen would proliferate is a complex and high-level cognitive behavior, there exists enough neuroscience research to provide a strong tie between the organically intelligent method of solving the problem and the theoretical artificial one. Within this paper, the history and development of artificial neural networks (ANNs) are discussed, as well as their details, benefits, and limitations. A potential plan for implementation of a disease outbreak modeling network is presented, based on vector-location encodings. Additionally, natural intelligence is consulted in the theorizing and design of said model. This paper concludes with a discussion of how AI differs from the manual method, as well as potential pitfalls and advantages.

Introduction

Although medical science strives valiantly to defeat infirmity and illness with ever-advancing technology and technique, disease and plague have been a constant companion to the human race throughout our history. With both large strides in pathogen-manipulation and the increasing connected-ness of today's world compared to the world of only a few hundred years ago, the potential for a contagion to decimate communities has never been higher. Implementation of a neural network that is engineered to model outbreaks across population centers given the characteristics of a locality and a real or theoretical pandemic could be massively helpful in preventing the deaths of thousands.

Comprehension of mathematical or statistical datasets is an extremely high-level process in human beings, and therefore is not easy to narrow down to specific brain areas. However, there is still a wealth of neuroscience research to pull from. For example, Arsalidou et. al (2017) provided an excellent meta-analysis of FMRI data that showed a significant statistical relationship between mathematical problem solving and the parietal and frontal cortices, as well as the insula and claustrum. Oftentimes, it seems that in order to solve a problem efficiently, Bubic et. al (2010) assert that people are required to construct complex internal models of the problem. In fact, this mode of high-level thought, when relating to prediction, is considered by many leading neuroscientists to be the primary function of the human brain (Bubic, 2010). However, it is very difficult to pin down mathematical reasoning and statistical extrapolation to one brain area. According to Butterworth (2011), there is a link between statistical and logistic tasks and the intraparietal sulcus, or IPS. However, there also appears to be a developmental trend from the right hemisphere to a bilateral representation, which may be related to a developmental linkage of the numerical processes to language (Butterworth, 2011). When relating the prediction of the metastasization of a

communicable disease on a map to the internal navigation and extrapolation of 3d spaces in the brain, Shikauchi and Ishii (2016) show that the superior prefrontal gyrus and temporal pole are key to predicting details of imagined spaces. Additionally, their results suggest that the human brain uses scene anticipation, mediated especially by parietal and medial prefrontal cortical areas, as a robust and effective navigation processing (Shikauchi & Ishi, 2016).

Simulation

Artificial Neural Networks, or ANNs, have long been a subject of great interest and popular support amongst not only the researchers contributing to their development, but to the public at large. Partly due to their name and partly due to their seemingly miraculous capabilities, ANNs have become both a staple of modern computer science and an enduring specter in the subconscious of the populace. It isn't hard to find evidence of the impact that just the fantasy of advanced neural networks has had on public culture. Films such as *Terminator* (1984) and *2001: A Space Odyssey* (1968) illustrate how the idea of a truly intelligent artificial intelligence has captured the public imagination. Unfortunately (or fortunately), even the most advanced ANNs are quite far from the general intelligence commonly depicted in mass media.

For such an inspirational and widely-vaunted technology, ANNs have a relatively short history. The first "neural network" was created in the early 1940s by a pair of neuroscientists named Warren McCulloch and Walter Pitts as a computational model based on high threshold logic. (Kulo, 2012). Originally little more than a curiosity, neural networks gained significant traction after D.O. Hebb developed what was later called the Hebbian Algorithm, a method for changing the weights of "neural" nodes over time as a response to stimulus. (Zsebet, 2013). Under this model, ANNs started being able to improve their accuracy over time. Interest and research into AI and machine learning

exploded over the next two decades, with major inventions such as the perceptron gaining large amounts of traction with both the computing and cognitive science communities. (Rosenblatt, 1959). The perceptron was in effect the world's first automatic pattern recognizer, with the ability to distinguish between and linearly classify different images represented in a 20 by 20 pixel field. Shortly after this in 1969, Minsky and Papert forcibly stagnated the burgeoning field of AI and ML by publishing their famous *Perceptrons*. In this book, Minsky points out two large problems with the rising tide of optimism in artificial intelligence that existed during the mid-twentieth century. Firstly, no current ANN configuration could learn even the XOR logic representation, a seemingly very simple concept. Additionally, even the most powerful supercomputers during that time were orders of magnitude slower than they would need to be to even attempt the AI goals that were ascribed to them by both hopeful scientists and the public. This revolutionary publication essentially shut down funding in AI for many years in an event known as the first "AI Winter". The exaggerated hopes of many research scientists were shown to be little more than fantasies, and most public and private funding going into the field of artificial intelligence was cut.

ANNs are at their core biologically inspired. Modeled after animal brains, ANNs are a group of artificial "neurons" that are able to process input and are connected to each other. Each of these connections pass data through themselves to other nodes, providing both output and input. Given an appropriate architecture and learning algorithm, ANNs are able to 'learn' to perform surprisingly complex tasks such as captioning images, modeling theoretical situations, or even generating realistic but non-existent faces. All ANNs, amongst other things, require neurons, connections, a defined state of activation, and a series of rules to govern the changing behavior of the neurons and connections. These rules generally consist of an activation rule, propagation rule, and a learning rule.

Respectively, these rules determine the function that combines the input from incoming connections with the present state of activation in a given neuron, the function that passes information about the current fitness of the network back through the nodes, and the function that determines how the neurons change their weights and connections in order to change and learn over time. In addition to this, most ANNs require training in the form of large example data sets. For example, if a neural network is created for the purpose of distinguishing dogs from cats, it would require a large amount of images of dogs or cats. Once the neural network made a prediction based on an image, it would then be revealed whether the image was in fact a dog or a cat. If the neural network was correct in its prediction, the behavior that led it to that prediction would be reinforced. If the neural network was incorrect, it would modify its heuristics to account for this new information. This takes place via the idea of backpropagation, or the result of the prediction filtering backwards through the nodes of the neural network and changing all or some of the individual connections.

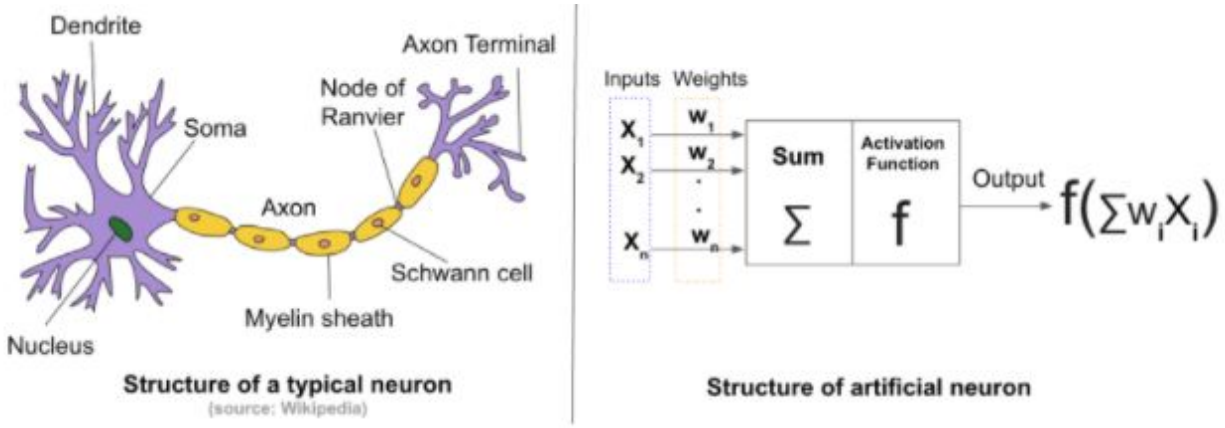

Figure 1. Comparing and contrasting structure of human neuron and artificial simulated neuron. Adapted from 'Prediction, Cognition, and the Brain' by Bubick et. al, 2010.

An innovation that has proved revolutionary since its inception in the field of neural networks and artificial intelligence in general is deep learning. Although the first recorded use of a deep learning architecture was in 1967 by Alexey Ivakhnenko, deep learning remained difficult to implement for many years. The core idea of deep learning is to introduce multiple layers of neurons between the input layer and output layer, passing the output of each layer as the input to the next. This relatively simple idea has allowed for much of the innovation and creativity in AI in recent years. By introducing multiple 'hidden' layers in between the input and output layer, neural networks can extract different features from the input and pass that to the output, greatly increasing their complexity and analytical ability.

Generally, large amounts of data are integral to training any network or group of networks. Luckily, the spread of pestilence is one area where a massive amount of data has been collected. The World Health Organization (WHO) has been working tirelessly for 71 years to inoculate and heal the world's citizens, and has collected a truly gigantic amount of data in their Global Health Observatory Data Repository (WHO, 2020). The National Center for Health Statistics (NCHS) has also conducted large amounts of research on pandemics, and has made their data publically available (NCHS, 2020). Using at least these sources, and almost certainly more, a sufficient amount of sanitized data is easily obtainable for the training of a hybrid CNN/LSTM/DBN. Specifically, the WHO retains extensive time-series data relating to the spreading of at least nine different pandemics (WHO, 2020), which could easily be converted into the form discussed later in this essay.

The execution of a network aimed at predicting disease statistics, given the authors current knowledge base, would proceed as follows. A CNN (Convolutional Neural Network), LSTM (Long Short-Term Memory) network, and Deep Belief Network hybridized together would be the most

suitable for the task. LSTMs are specialized for data that changes over time, as the nodes in the network map to themselves over time. Traditionally used in order to combat the vanishing gradient problem, an LSTM is an essential component in a project with data that varies over time. In contrast, Deep Belief Networks excel at taking large quantities of data and either generating content based on it or classifying and categorizing the data. Finally, Convolutional Neural Networks, or CNNs, specialize in analyzing visual stimuli. As none of these architectures would be sufficient on their own for the purposes of this paper, a hybrid LSTM/DBN/CNN would be constructed in order to accurately predict spread of disease.

Essentially, this model would take in 3 different forms of input. The first would be a vector-based grid-coordinate encoding of a geographical location, with each bit in the location vector representing a specific quality of that location. Specifically, each square in the grid-overlay that is mapped onto the location would have a x-bit vector associated with it that changes over time in response to environmental stimuli. The first bit would signify the presence of a significant geographical boundary, such as a river, ocean, or mountain. The second bit would signify a notable political boundary, while the third bit would represent the presence of a noteworthy transportation network. In terms of mildness or severity, the climate would be established by the paired presence of the next two bits and so forth. This bit-based vector model is entirely theoretical and extremely mutable in response to what the CNN/LSTM/DBN would need to predict, but it scales down well into a format that the network could understand. Depending on the desired resolution of the output, a given square could map onto anywhere from half of a square mile to 10 square miles. Over time, the model would have to write to itself not only in the form of changing the LSTM cell-state, but also changing the map itself. This brings us to the next form of input, which is a standard map of the

location in question, overlaid by a layer representing population density. This layer is constituted of a simple RGB filter, corresponding to each square and therefore vector. The values would range from light green, denoting little to no population density, to dark red or black, denoting extremely high population density. This map would be updated over time to show how population density changes over a given period of time. The CNN component to the model would analyze this data using overlapping visual fields as depicted in figure 2. Overlapping visual fields are an extremely efficient way to represent spatial data (Rumelhart & McClelland, 1986). This component of the input is integral to the success of the model, as varying population density is extremely important when predicting how people will be in contact and therefore spread the pathogen. For the last form of input, the network would also have to be given several key characteristics of the disease, such as its R number (Reproduction Number), lethality, duration, and method of transmission.

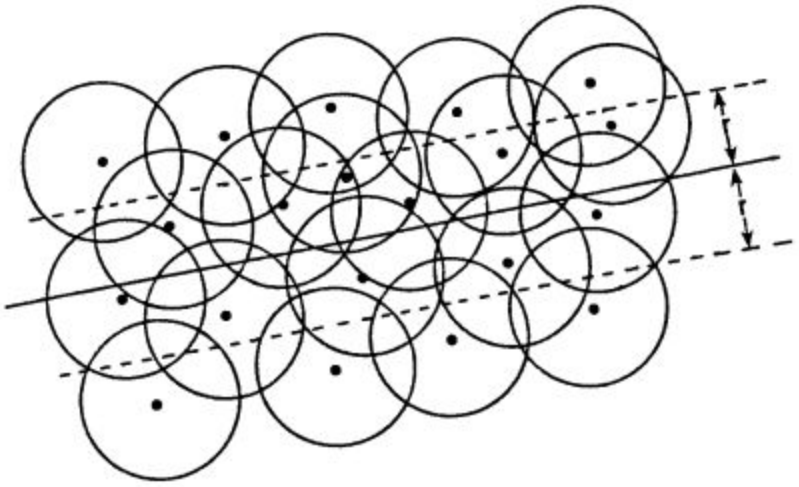

Figure 2. An example of coarse coding, used to illustrate the effectiveness and theory behind overlapping visual fields. Adapted from Rumelhart & McClelland, 1986.

As output, the model would produce both an altered version of every vector-based location encoding that it is fed and a few overall statistics. The augmented vectors would be supplemented by several new bits concatenated on to them in the form of average deaths, number of people requiring medical attention, and lost productivity in terms of monetary value. These vectors would then be reassembled utilizing data from the CNN into a map that graphically displays the spread of the disease. The general statistics would report total estimated deaths over the entire area, estimated length of time to develop a vaccine, total number of infected, and level of medical intervention warranted on the CDC disease scale (1-5). The interval that the network would output updated representations of these location vectors would be specified by the operator.

Deep Belief Networks (DBNs), although not as popular as CNNs or more widely known ANNs, are very powerful and general networks that are modifiable for almost any data set. In contrast to most deep learning networks, DBNs do not have connections between the nodes in their hidden layers. Instead, DBNs connect the nodes in each hidden layer immediately to the subsequent layer. Each layer in the DBN is what is called a Restricted Boltzmann Machine, which are relatively simple ANNs that learn a probability distribution over their inputs. Geoffrey Hinton, a famous figure in artificial intelligence and a person who some consider to be the father of deep learning, developed a fast, greedy method of training DBNs one layer at a time (Vector Institute, 2019). When a given RBM layer has finished learning, its feature activations are summarily passed to the next layer as training data. Due to their many layers, DBNs are especially competent at learning hierarchical forms of data, a form that many ANNs struggle with (Bubic, 2013). DBNs also excel at learning statistical or demographic data despite their generality.

Long Short-Term Memory Networks (LSTMs), by contrast of DBNs, are usually specialized for temporal information. In traditional recurrent neural networks, the process of gradient descent often becomes an issue, via the exploding or vanishing gradient problem. In the exploding gradient problem, a gradient is multiplied by a number bigger than 1 over and over, resulting in an impossibly high slope. In the vanishing gradient problem, the reverse occurs, with the gradient becoming exponentially smaller over time. The central idea behind LSTMs is that of persistence. Temporal context is often integral to understanding a data set, and LSTMs deal with that by implementing a loop wherein neural nodes write to themselves over time. LSTMs utilize a 'cell state', or a persistent memory that data is added to or removed from throughout activation of the neural network. Each cell possesses various gates, each performing a different function to the persistent memory of the network. Depending on how the network is trained and the data that the LSTM is given, these gates will either 'forget' (remove information), calculate the input data, or calculate the output data.

The most essential part of how neural networks actually improve themselves over time is referred to as the backpropagation algorithm. Backpropagation is the workhorse that supports algorithms such as LMS by efficiently changing the values of the weights between nodes one layer at a time. This process, called gradient descent, essentially views the error as a 3rd or higher dimensional space, and utilizes the finding of the gradient to find local or global minimums within that space. Once the gradient is calculated and the direction in which the minimum lies is found, the weight adjusts itself as a function of that distance. For example, if the error is very low, the weight changes only very slightly, but if the error is high then the weight makes a large change in the appropriate direction. LMS was a revolutionary algorithm because it considered the overall error of the network, and finely tuned the network to the exact specifications of the data.

Another integral component of any neural network is the activation function that it employs. An activation function is a function that compresses the various input data that a neuron receives into a single value, often between 0 and 1. Another way to think about it is that an activation function determines the level of 'activation' (output) a neuron produces as a function of the inputs that it receives. Depending on the type of output required, an activation function can produce a wildly varying range of numbers. Until recently, the most common activation function was the sigmoid function, which approximates input as a probability distribution between 0 and 1. Each activation function has its own unique benefits and drawbacks, so it's important to specify the optimal activation function for a given neural network. A recent milestone in the deep learning revolution overthrew the sigmoid function as the most commonly used activation function, the ReLU function. The ReLU function was biologically inspired from visual neurons, resulting in sparser activation of hidden units, better gradient propagation, and more efficient computation amongst other improvements. (Erzsebet, 2013). For the purposes of this paper, LMS with backpropagation would most likely be used, along with either a sigmoid or ReLU activation function.

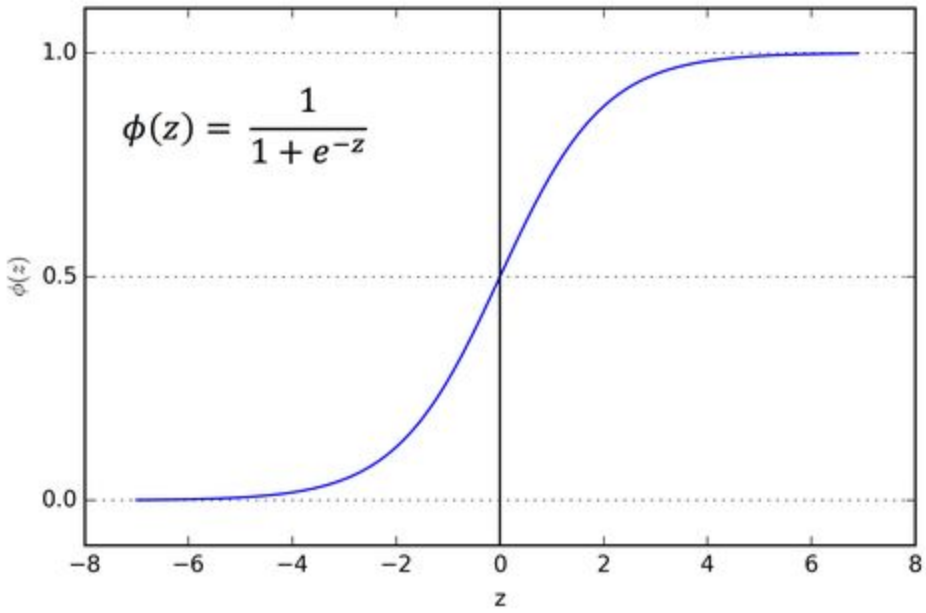

Figure 3. The Sigmoid activation function. Adapted from Nwanka et. al, 2017.

A classic issue with analyzing data sets with neural nets is overfitting. Overfitting is the concept that, unless trained carefully, a neural network will tend to learn the training data very well but not be able to generalize to real-world phenomena. If trained incorrectly or for too long, a neural network will fail to extract the salient features from a data set that it needs to in order to actually perform, instead just conforming exactly to the data that it is given. This is combated in a few ways. Stochastically dropping out a few neurons and all of their associated weights after each training pass guarantees good results, as it models a random distribution. Alternatively, you could utilize the testing data as another data set to train on in parallel with the training data, only stopping when error starts to increase after every generation of training (Grookey, 2017). Decreasing your networks tendency to overfit is even possible by simply decrementing every weight by a small amount every generation, thereby regularizing the connections.

Discussion

To predict the outbreak and spread of diseases manually, masses of data and many people with expert educations are required. Generally, a team of epidemiologists and data analysts will work together after a disease is discovered to have infected a great deal of people in order to pin down its origins and hopefully discover more about it. This manual inference is quite difficult to both perform and quantify, as it depends on many specialized people working together. A team led by Drake et. al (2019) recently developed a statistical model for predicting disease outbreak based on time elapsed since outbreak and characteristics of diseases, but did not account for what happens when the disease actually starts spreading. Both the WHO and the CDC incorporate advanced statistical modeling into their disease prediction strategy, but the field is in its infant stage (CDC, 2019). The benefits of an accurate disease modeling network are obvious and plentiful. Appropriate resource allocation, faster medical response, and suitable quarantining are all advantages that are very difficult to procure otherwise. The primary difficulty however, is creating a generalizable ANN that models disease outbreak with such a small data set. In every outbreak there is inherently a large amount of data, but it is difficult both to gather and compile, and data-compilation is often of low priority during a pandemic. Nevertheless, if an efficient model was implemented, it would save both lives and money almost immediately.

References

Bubic, A., Cramon, Y. V., & Schubotz, R. (2010, March 22). Prediction, Cognition and the Brain. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2904053/

Butterworth, B., & Walsh, V. (2011). Neural basis of mathematical cognition. *Current Biology*, *21*(16). doi: 10.1016/j.cub.2011.07.005

Drake, John, Tobias S. Brett, Shiyang Chen, Bogdan I. Epureanu, Matthew J. Ferrari, Éric Marty, Paige B. Miller, Eamon B. O'Dea, Suzanne M. O'Regan, Andrew W. Park, Pejman Rohani. The statistics of epidemic transitions. *PLOS Computational Biology*, 2019; 15 (5): e1006917 DOI: 10.1371/journal.pcbi.1006917

Grookey, R. (2017, December 19). Man vs machine: comparing artificial and biological neural networks – Sophos News. Retrieved from https://news.sophos.com/en-us/2017/09/21/man-vs-machine-comparing-artificial-and-biological-neural-networks/

Marie Arsalidou, Matthew Pawliw-Levac, Mahsa Sadegh, Juan Pascual, Leone. (2017). Brain areas associated with numbers and calculations in children: Meta-analyses of fMRI studies. (n.d.). https://www.sciencedirect.com/science/article/pii/S1878929317300105

National Center for Health Statistics (NCHS). (n.d.). Retrieved February 2, 2020, from https://data.cdc.gov/browse?category=NCHS&sortBy=last_modified

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2017, August 3). Activation Functions: Comparison of Trends in Practice and Research for Deep Learning. Retrieved from https://arxiv.org/pdf/1811.03378.pdf

Rosenblatt, F. (n.d.). Perceptron Algorithm, 1959; Rosenblatt. *Cornell Aeronautical Library*. doi:10.1007/springerreference_57806

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing. Explorations in the Microstructure of Cognition: Volume 1: Foundations*. Cambridge, MA: The MIT Press.

Shikauchi, Yumi & Ishii, Shin. (2016). Robust encoding of scene anticipation during human spatial navigation. Scientific Reports. 6. 37599. 10.1038/srep37599.

Vector Institute. (2019, April 22). Geoffrey Hinton. Retrieved from https://vectorinstitute.ai/team/geoffrey-hinton/

World Health Organization. (n.d.). Weekly epidemiological record: cholera articles. https://www.who.int/cholera/statistics/en/

Zsebet, E. R. (2013, December 15). Artificial Neural Networks: Hebbian Learning. Retrieved from https://www.ece.rice.edu/~erzsebet/ANNcourse/handouts502/course-cf-3.pdf