

Characteristics of an Artificial Neural Network and its use in Business

¹Abhijit Dasgupta, ²Dr. Raja Roy Choudhury

¹Research Scholar, Dr. Ram Manohar Lohia Avadh University (UP)

²Director, N. L. Dalmia Institute of Management Studies and Research Mumbai

¹adg2004@gmail.com, ²director@nldalmia.edu.in

Abstract - Improvements in compute power over the years has helped mathematicians and computer scientist to mainstream complex mathematical models and prediction for business. One such method is using Artificial Neural Networks (ANN) which is helping business community to augment their abilities to predict business outcomes. Current researches in ANN has led to significant progress in understanding customer behaviour, financial risk, prevent fraud among many others. A significant role has also been played by cloud computing companies like Amazon, IBM, Microsoft together with open source initiatives of organizations like Apache Foundation, Google etc., has helped the cause of the erstwhile business manager who were otherwise were having limited tools to predict, say, sales for the next quarter.. In this paper, we describe the basics of an artificial neural networks and its usage in businesses.

Keywords: Artificial, Neural Networks, Perceptron's, Deep Learning, Business Application.

I. What is an ANN? (Artificial Neural Network)

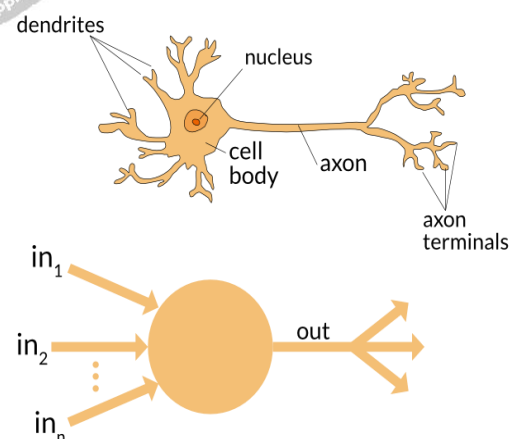
Artificial Neural Network is bio inspired, or more correctly, human brain. The human brain consists of nerve cells called neurons, linked together with other neurons via axons. Axons are used to transmit nerve impulses from one neuron to another. A neuron is connected to the axons of other neurons via dendrites, and the contact point between an axon and a dendrite is called a synapse. Neurologists have discovered that the human brain or for that matter an animal brain learns by changing the strength of the synaptic connect between neurons upon repeated stimulation by the same impulse. Similar to this human brain system, an ANN is composed of an inter-connected assembly of nodes and directed links. Perceptron is a simple model of Artificial Neural Network.

The perceptron (a simple form of ANN) consists of two types of nodes: input nodes and an output node, which is used to represent the model output.

II. Literature Survey of ANN

In 1943, Warren McCulloch and Walter Pitts developed a computational model on neural networks, and during late 1940s D.O. Hebb created a model on unsupervised learning. Later, Rosenblatt created the perceptron algorithm in 1958. However, the main impetus for further research on neural networks started when Werbos demonstrated back propagation algorithm to train multi-layer networks. Later in early 2000s, with the advent of distributed computing, graphic processing unit (GPU) neural networks were used to

solve visual recognition problems; this is known as deep learning. The original ideas are credited to Rumelhart and McClelland research published during mid-1980s. As of 2017, the state of feed-forward networks incorporates convolutional layers and max-pooling layers and these techniques have been effectively used in computer vision, speech and image recognition etc. pioneered by researchers at Carnegie Mellon University, Stanford University among many others, including companies such as Google, Palantir, Baidu etc.



Application of ANN

Artificial Neural Networks (ANN) have been applied to solve several problems, such as (among many others)

(a) Image recognition: Identifying an object (is that a pant? or a shirt? or a salwar-kameez?)

(b) Anomaly detection: Comparing a test data with usual pattern (for instance, is this customer a happy one or an angry one?)

(c) Approximating a target function: Commonly used in weather prediction, for example, will this storm turn into a tornado, and if yes, in how many days / hours?

(d) Speech recognition: Successfully used in call centres to identify gender, age of the caller and even the mood of the caller

ANNs have been used to identify faces (e.g., face recognition algorithm in Facebook), Speech, Language of a Text including its translation (Google translation system uses this effectively), Games (like Chess, Playing Poké mōngo) to complex subjects such as controlling Autonomous Vehicles and behaviour prediction as in if the driver of a vehicle is going to fall sleep.

III. TRAINING AN ANN

Simple ANN or a Perceptron consists of two types of nodes (a) Input Node and (b) Output Node. Input nodes are used to depict the input attributes whereas an Output node is used to depict the model output. The nodes in a neural network model is also known as neurons. In a Perceptron, each input is connected with a weight to the output node. These weights are used to signify the strength of the synaptic connection between neurons. Interestingly, neural network cannot be coded or configured to work in a traditional Input-Process-Output computer program. Like humans, a perceptron has to learn.

Below is a depiction of a simple Learning Algorithm of a Perceptron (steps are defined as a, b, c, etc.)

a) Let $D = \{(x_i, y_i) \mid i = 1, 2, 3, \dots, N\}$ be the training set example

b) Initialise the weight vector with random values $w^{(0)}$

c) Repeat

d) For each training example $(x_i, y_i) \in D$ do

e) Compute the predicted output $y_i^{(k)}$

f) For each weight w_j do

g) Update the weight, $w_j^{(k+1)} = w_j^{(k)} + \lambda (y_i - y_i^{(k)}) x_{ij}$

h) End for

i) End for

j) Until stopping condition is met

Typically there are four machine learning methods for training an ANN:

i) Supervised Learning

If there is a large enough data set with known output, it is a simple and useful way to train the perceptron. The process is depicted in the simple algorithm above. The algorithm provides for the iteration as follows: (1) process one data set (2) Compare the result against known output (3) Adjust the network and repeat. Gender determination from voice sample at a call centre is a good example of supervised learning method.

ii) Un-supervised Learning

The goal of unsupervised learning to unearth a model of the underlying structure of the observation (data) where there are no correct answers and no teacher. Unsupervised learning algorithms are left to their own to discover the information contained in the data set. Almost all data mining applications in marketing research, consumer behaviour, such as Cluster Analysis uses this method. Popular learning algorithms for un-supervised learning are (a) k-means for clustering problems (b) Apriori algorithm for association rule mining

iii) Semi-supervised Learning

Let's say there are a large pack of photographs animals (eg cat, dog, elephant etc.) and some of these are labelled whereas most are not. Most of the real life machine learning problems fall into this category. Semi-supervised learning algorithms are used, because, it might be expensive to label the entire dataset (in our example, label all the animals correctly using a human domain expert) and might also require time. In such a scenario, semi-supervised learning can be used to discover structure of the dataset and make prediction on new data, such computer vision application to identify objects (such as car, tree, road etc.).

iv) Reinforcement Learning

This is often called a "carrot and stick" method in machine learning. Over time, the network learns to prefer the correct action (or most logical action) and avoid the wrong one. For example, a robot takes a larger step and falls, then it takes a smaller step and doesn't fall. The robot learns to stay in balance (carrot) and not fall (stick). Over time the robot learns how to remain stable and not to fall. In supply chain distribution UPS is said to use a technique in USA whereby none of its truck takes a left turn (that would cut against a traffic) and always takes right irrespective of a situation where turning left would have required to cover a shorter distance. UPS used reinforcement learning techniques to figure out the lowest cost and lowest time to reach a destination by using always right turns, which helps it to avoid accidents, fuel, litigation costs and staff hours resulting in a total benefit estimated to be more than a billion dollar a year. Q-Learning is a popular Reinforcement learning method, which primarily focuses not on short term reward but on long term goals.

IV. Multi-layered Artificial Neural Networks: Deep Learning

A typical ANN is more complex than the simple Perceptron model. The additional complexities arise due to the following factors:

(a) The network may contain hidden layers, which are often called intermediary layers. In a feed forward ANN the nodes of one layer is connected to the nodes of another layer. (A

perceptron network has only one layer of input nodes that connect to one layer of output node).

(b) The network may use activation function, such as linear, logistic etc., rather than sign function. These activation functions allow the hidden and output nodes to produce values that are nonlinear

In a deep learning situation, the algorithm for ANN crunches far more data with multiple layering with both forward feed and back propagation. Simplistically, a Deep Learning algorithm is an ANN where it crunches more data than a normal perceptron based simple ANN.

V. Sentiment & Intent Analytics (Text Analytics) - Use Case of ANN

The use case for deep learning based text analytics revolves around its ability to parse through massive amounts of text data and either aggregate or analyze. Using deep learning, computers can perform tasks like e-discovery. Large investment houses like JP Morgan Chase, Goldman Sachs etc. are using deep learning based text analytics to identify insider trading and for statutory compliance. Hedge funds use text analytics to mine through massive document repositories for insights into future investment performance and market sentiment. Facebook, Instagram, LinkedIn, Twitter etc. uses text analytics to recommend relevant posts among other things. Advertisement agencies, digital marketers, use text analysis on social media for effective digital marketing research, such as to gauge brand reputation, customer sentiment and also to respond real-time to an identified issue, including customer support.

In each case, it isn't cost effective to hire the staff necessary to sift through all the documents. Being able to automate that task is not only a cost savings, but a competitive advantage. Text analytics is typically a hybrid project. Some of the code necessary to build deep learning text analytics capabilities are in open source libraries like Google's Parsy, IBM's Watson API, Spark ML Library, and several others. There will be additional work to extend, customize, train, and integrate these libraries. There's no text analytics solution that works out of the box at this time but the returns in productivity and improved capabilities make this worth the investment

VI. CHARACTERISTICS OF AN ANN

Following are the observed characteristics of an artificial neural network:

(a) Since an ANN has a very expressive hypothesis space, it is important to choose the approximate network topology (simplistic, layered, and multi-layered) for a given problem to avoid model over fitting

(b) ANN can handle redundant features because the weights are autonomous and automatically learned during the learning process. The weights for redundant features tend to be insignificant

(c) Training an ANN is time consuming particularly when the numbers of hidden layers are quite significantly large

(d) ANN are quite sensitive to the perspective of noise in the training data. One approach to address this issue is to use a validation set to determine and eliminate generalization error. Another approach is to adjust the weights by some percentages at each iteration.

VII. Use Cases of ANN in Business

a) Cyber Security: ANN has a large number of success stories in cyber security applications. The advantage of ANN based Deep Learning over other approaches since it provides 99% accuracy in detection. The high risk and cost associated with failing to detect a threat make the expense associated with deep learning worthwhile, and therefore companies like Cisco, AWS, Symantec, Microsoft etc. invests quite heavily here

b) Plant Maintenance: Predictive Plant Maintenance is one of the highest applied use cases for ANN. Using anomaly detection and survival analysis, ANN algorithms can predict when a machine (everything from an airplane engine, rocket propulsion system to a lathe machines in manufacturing facilities) will fail. That allows machine downtime to be planned with minimal impact to operations.

c) Healthcare: ANN algorithms can predict using data points from genetics, life style and clinical parameters and can predict with a very high degree of accuracy about a person's upcoming disease and hospitalization. It can even prescribe medicines and treatments like what IBM Watson does quite effectively

d) Recommender Systems: Very ubiquitous in online retail platforms (travel, shopping etc.) in India. It uses ANN and collaborative filtering based systems to suggest a product to a customer, based on previous purchases and or current interests.

e) Retail: Neural networks are excellent tool in the domain of sales forecasting, since these networks can factor in multiple parameters such as market demand, a disposable income, population size, product price, and the price of complementary products. Sales forecasting using ANN has traditionally outperformed simple statistical forecasting methods like Regression. An ANN using un-supervised learning methods can also recommend to a merchandiser "allocation levels" for a new product in a store, which otherwise would be a trial and error method for the human.

f) Banking & Finance: Neural networks is systematically used for banking & finance companies for trading and financial forecasting. ANNs have been applied successfully to solve complex pricing problems for derivatives (options & futures), securities and hedging, currency and stock.

g) Mobile Services / Telecom: ANN offers companies like Airtel, Jio, Vodafone etc a clear opportunity to ascertain a much more complete picture of their operations and their customers, as well as to further their innovation efforts. Companies like AT&T, Vodafone etc. are using a series of neural networks to analyze customer and call data in order to predict if, when, and why a customer is likely to leave for another competitor. Managers responsible for customer retention, promotions etc. are also slowly adopting ANN for business decisions.

h) Taxi Service: Pricing, and particularly surge pricing in apps based taxi services base on the ANNs that is trained with billions of data points. Uber, Lyft, Ola use ANN effectively to price their services for a specific customer at a specific time for a specific route.

VIII. CONCLUSION

Neural Networks and particularly artificial neural networks are slowly gaining traction in the mainstream business. However, the speed of adoption is directly linked to the level of technical maturity of the business owners who were hitherto were mostly driven by "guts" and past experience. The emergence of new generation trained business graduates is creating larger tractions, though not all projects are coming out successful, primarily due to their faulty planning and hyped objectives.

The key advantage of an ANN that they do not require complex programming knowledge and many machine learning algorithms are available out of box and free from companies / organizations like Apache (Spark has several ML libraries freely available), Google (Tensor Flow is a highly popular open sourced product on deep learning from Google). This open source initiative is further popularizing adoption of ANN as decisioning and decision support systems for many organizations worldwide. With initiatives line driver-less cars, smart cities, Internet of Things, Robotics etc., ANN is here to stay and consolidate its adoption in almost every possible human-business touch points in future.

REFERENCES

[1] Alexandridis, A. K., & Zaprani, A. D. (2013). Wavelet neural networks: A practical guide. *Neural Networks*, 42, 1-27.

[2] Baldi, P., Sadowski, P., & Lu, Z. (2017). Learning in the machine: The symmetries of the deep learning channel. *Neural Networks*, 95, 110-133.

[3] Bataineh, M., & Marler, T. (2017). Neural network for regression problems with reduced training sets. *Neural Networks*, 95, 1-9.

[4] Cireşan, D., Meier, U., Masci, J., & Schmidhuber, J. (2012). Multi-column deep neural network for traffic sign classification. *Neural Networks*, 32, 333-338.

[5] Du, K. L. (2010). Clustering: A neural network approach. *Neural networks*, 23(1), 89-107.

[6] Fayek, H. M., Lech, M., & Cavedon, L. (2017). Evaluating deep learning architectures for Speech Emotion Recognition. *Neural Networks*.

[7] Grant, W. S., Tanner, J., & Itti, L. (2017). Biologically plausible learning in neural networks with modulatory feedback. *Neural Networks*, 88, 32-48.

[8] Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5), 602-610.

[9] Grossberg, S. (2017). Towards solving the hard problem of consciousness: The varieties of brain resonances and the conscious experiences that they support. *Neural Networks*, 87, 38-95.

[10] Hodge, V. J., O'Keefe, S., & Austin, J. (2016). Hadoop neural network for parallel and distributed feature selection. *Neural Networks*, 78, 24-35.

[11] Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), 359-366.

[12] Huang, G., Huang, G. B., Song, S., & You, K. (2015). Trends in extreme learning machines: A review. *Neural Networks*, 61, 32-48.

[13] Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, 13(4), 411-430.

[14] Ijspeert, A. J. (2008). Central pattern generators for locomotion control in animals and robots: a review. *Neural networks*, 21(4), 642-653.

[15] Kim, J., Kim, J., Jang, G. J., & Lee, M. (2017). Fast learning method for convolutional neural networks using extreme learning machine and its application to lane detection. *Neural Networks*, 87, 109-121.

[16] Kohonen, T. (2013). Essentials of the self-organizing map. *Neural networks*, 37, 52-65.

[17] Madl, T., Chen, K., Montaldi, D., & Trapp, R. (2015). Computational cognitive models of spatial memory in navigation space: A review. *Neural Networks*, 65, 18-43.

[18] Meruelo, A. C., Simpson, D. M., Veres, S. M., & Newland, P. L. (2016). Improved system identification using artificial neural networks and analysis of individual differences in responses of an identified neuron. *Neural Networks*, 75, 56-65.

- [19] Parisi, G. I., Tani, J., Weber, C., & Wermter, S. (2017). Lifelong learning of human actions with deep neural network self-organization. *Neural Networks*, 96, 137-149.
- [20] Pedernana, M., & García, S. G. (2016). Smart sampling and incremental function learning for very large high dimensional data. *Neural Networks*, 78, 75-87.
- [21] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117.
- [22] Sharma, A., Liu, X., Yang, X., & Shi, D. (2017). A patch-based convolutional neural network for remote sensing image classification. *Neural Networks*, 95, 19-28.
- [23] Stallkamp, J., Schlipsing, M., Salmen, J., & Igel, C. (2012). Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural networks*, 32, 323-332.
- [24] Tan, P. N., Steinbach, M., & Kumar, V. (2013). Data mining cluster analysis: basic concepts and algorithms. *Introduction to data mining*.
- [25] Watanabe, C., Hiramatsu, K., & Kashino, K. (2017). Modular Representation of Layered Neural Networks. *arXiv preprint arXiv:1703.00168*.

