Mathematical Model of ANN

Dr. Sushil Kumar Saini
Associate Professor of Mathematics
Dronacharya Government College Gurugram

Abstract: Pattern recognition, prediction, system identification, and control may all benefit from the usage of neural networks. Artificial or formal neural networks (ANNs), parallel distributed processing (PDP), neuromorphic or connectionist models, and other terms for these models and methods have been developed to replicate how the human brain processes information and acquires knowledge. Many people use the word "network" to describe anything from computer networks to communications to organisations and marketplaces. In the beginning, the idea of an artificial neural network (ANN) was conceived as an optimistic vision of artificial intelligence (AI) synthesis by mimicking the human brain. To incorporate neural-inspired notions in artificial intelligence contexts, ANNs provide an alternative to symbol programming.

Keywords: Artificial or formal neural networks (ANNs), parallel distributed processing (PDP), artificial intelligence (AI) synthesis

I. Introduction

Intelligent or pattern-recognition jobs are notoriously difficult to automate, despite the fact that people seem to handle them with ease. Humans, for example, seem to need relatively little effort to distinguish distinct things and make sense of the huge quantity of visual information in their environment. Understanding how humans do these jobs and recreating these processes to the degree that physical restrictions allow is a logical next step for computational systems that try comparable tasks. Consequently, the study and modelling of Neural Networks is required to address this. A human's nervous system has a huge number of linked neurons known as the neural network (nerve cells). Nervous and network are terms that describe a network's graph-like structure. To put it another way, an artificial neural network is any kind of computer system that borrows its inspiration from biological neural networks. The terms "Neural Nets," "parallel distributed processing systems," and "connectionist systems" are all used to describe artificial neural networks. Some basic calculations must be performed on the nodes of a labelled, directed graph structure in order for the system to be referred to as "beautiful". There are "Nodes" (vertices), and then there are "Connections" (e.g., edges/links/arcs) that connect these two nodes together. Nodes in a neural network conduct basic calculation, while connections carry signals from one node to the next, with the "Link Strength" or "Weight" value reflecting the degree to which a signal is amplified or lessened by a connection. Human skill and knowledge have been replaced by this system. A lot of the language used in artificial neural networks is drawn from neurology since they are designed after the brain.

II. Literature review

Pavlenko et al. (2018), The purpose of this article is to propose a scientific and methodical approach to employing artificial neural networks (ANN) to solve practical mechanical engineering challenges. ANN and current numerical analysis tools (such as the finite element method) and analytical research methods based on mathematical modelling of the dynamic state of mechanical systems are used in this approach. For a variety of multidisciplinary issues, such as the dynamics of rotary machines and the hydro-aeroelastic interaction of gas-liquid mixtures with deformable structural components, conceptual methods for the application of the abovementioned methodology are presented. It is possible to train and enhance the ANN design, and to tackle nonlinear issues of parameter identification for mathematical models using data from physical experiments and numerical simulations, as opposed to conventional regression analysis. This method allows for the refinement of parameters in linear and nonlinear mathematical models representing complex mechanical and hydro-mechanical interactions despite the fact that an exact solution to the equations defining the process cannot be determined and the starting data is incomplete.

Elkatatny et al. (2018), The dynamic geomechanically characteristics, such as Poisson's ratio, Young's modulus, and Lamé parameters, may be estimated using compressional (P-wave) and shear (S-wave) velocities. Static characteristics of the formation rocks and in-situ stresses are estimated using these parameters. When it comes to older wellbores, sonic logs may not be accessible. When the sonic logs are accessible, it is possible that missing parts in the well logs may alter the findings of the study. Well log data alone cannot be utilised to reliably estimate P- and S-wave travel periods, therefore this is a novel approach for which the authors are grateful. P-wave velocity is used in most known correlations to estimate the S-wave velocity. Wireline log data (bulk density, gamma ray and neutron porosity) will be used to construct accurate and simple empirical models to forecast sonic transit durations (P-wave and S-wave). This well has been using slandered wireline log data, which is a standard practise in the majority of wells. They used support vector machine (SVM), neural network (ANN), as well as an adaptive neuro fuzzy interference system to evaluate their prediction performance and determine which was the best. A simple generalised empirical correlation may be constructed from the weights and biases of the optimised ANN model, which can be used to run AI models without the usage of expensive commercial software.

Kilickap et al. (2017), In this work, cutting factors such as cutting speed, feed rate, and depth of cut were tested to see how they affected cutting force, surface roughness, and tool wear while milling Ti-6242S alloy using 10 mm diameter cemented carbide (WC) end mills. ANN and Response Surface Methodology (RSM) were used to define the experimental data (RSM). The Levenberg-Marquardt (LM) and weights of an ANN were used to train the network. Box Behnken design was used to develop RSM's mathematical models. A close match was discovered between the results of the ANN and the RSM and experimental data. At high cutting speeds and low feed rates and depths of cut, the lowest cutting force and surface roughness were achieved. At modest cutting speeds, feed rates, and depths of cut, little tool wear was achieved.

Kuo (2016), Using convolutional neural networks (CNNs), this research aims to answer two key issues concerning their structure: a nonlinear activation function is necessary at the output of all intermediate layers because of this. Which system has a greater benefit, the one that has a single layer or a two-layer cascade system? In order to resolve these problems, a mathematical model known as "Rectified-Correlations on a Sphere" (RECOS) is presented. The converging filter weights define a collection of anchor vectors in the RECOS model after the CNN training procedure. The most common patterns are represented by anchor vectors (or the spectral components). The RECOS model is used to illustrate why correction is necessary. Finally, the performance of a two-layer RECOS system is compared to that of a single-layer system. Illustrations are provided by way of LeNet-5 and MNIST. Finally, using the AlexNet as an example, the RECOS paradigm is extended to include multilayer systems.

Shanmuganathan (2016), It has been a great success for Knowledge Engineers in computer science to include heuristics into computational algorithmic modelling using brain models that have already been established by scientists in neuroscience, medicine, and high-performance computing. To develop therapies for currently classified as incurable brain and nervous system disorders including Alzheimer's and epilepsy, it is important to learn more about human brain/nerve cell architecture, structure, and how the human brain operates. Researchers in the medical field have made some big advances in the previous several decades, but they still do not fully grasp how people think, learn, and remember; nor do they comprehend the connections between cognition and behaviour. ANN structures, components, associated terminologies, and hybrids are discussed in this context, based on contemporary human brain research endeavours, in this chapter.

Mohamadou et al. (2020), Mathematical modelling and AI have been used to study COVID-19's dynamics and early detection in the last several months (AI). We hope this work will serve as a complete review of the methodology employed in these investigations and a database of open-source datasets related to COVID-19. " COVID-19 was the subject of 61 peer-reviewed journal papers, reports, info sheets, and websites examined for this study. The majority of mathematical modelling was based on the SEIR and SIR models, whereas the majority of AI implementations were Convolutional Neural Networks (CNN) using X-ray and CT images, according to the study. Case reports, medical pictures, treatment tactics, healthcare workers' demographics, and epidemic movement are only some of the statistics that are readily accessible. This epidemic can be combated with the use of mathematical modelling and artificial intelligence. The COVID-19 has also been the subject of a number of open-source datasets. However, there is still a great deal of work to be done in terms of expanding the databases. In light of the COVID-19, further healthcare-related AI and modelling applications should be investigated.

Vo et al. (2019), The growing relevance of natural gas-based hydrogen generation has made the steam methane reformer (SMR) increasingly appealing. Models for the reactor, wall, and furnace of an SMR were produced as part of this research. The generated SMR model was verified using reference data such as temperature, pressure, mole fraction, and average heat flow within a tiny error (less than 4 percent). It was proven that the model's predictions were accurate when the catalyst's specifications and operating circumstances were altered. In order to create the SMR performance data, the proposed model was employed with four primary operating variables: the intake flow rate, temperature, S/C ratio, and the inlet flow rate of the furnace side. Singular value decomposition was used to evaluate the resulting dataset in order to minimise its dimensionality. There were 81 datasets used to train an artificial neural network (ANN) that used feed-forward back propagation to map the link between operational variables and anticipated outputs. It anticipated the outputs (temperature, velocity, pressure and mole percent of components) with more than 98% accuracy. Additionally, the computing time was slashed from 1200 s to only 2 s (using a dynamic simulation) (ANN). An online operation and optimization of a reformer with high precision may also be used to the design of a hydrogen production system at cheap computing cost using the methods proposed in this study.

Ceylan (2008), Poplar and pine wood were dried in a heat pump drier that ran 24 hours a day in this study. In the drying room, the weight change of all timbers was monitored, and drying was halted when the appropriate weight was reached. There was an initial moisture content of 1.28 kg water/kg dry matter for poplar timbers, which was decreased to 0.15 kg water/kg dry matter moisture content in 70 h, and 0.60 kg water/kg dry matter moisture content for pine timbers in 50 h. Drying air temperature, relative humidity, and stack weight were all recorded and stored on a computer during drying and examined subsequently. Semi-theoretical models and empirical data were used to examine the moisture ratios in the Stratigraphic computer application. Standard error of estimate (SEE) and R2 values were reached.

III. Mathematical Model

With the use of modern statistics and biological principles, an artificial neural network (ANN) can address issues in areas like pattern recognition and game play. The fundamental concept of ANNs is based on neuron mimics coupled in various ways.

Perceptrons: The simplest neural network, developed by Frank Rosenblatt in 1958, has n inputs, one neuron, and one output, where n is the number of characteristics in our dataset. Forward propagation is the name given to the method of sending data through a neural network, and the three stages below explain how forward propagation works in a perceptron.

Step 1: Multiply the input value xi by the weights wi and add the resulting multiplied values for each input. A neuron's output is influenced by the strength of the connections between its neurons, which are represented by weights. W1 is more influential than W2 because of its larger weight, and this is true even when both are equal in weight.

$$\sum = (x_1 \times w_1) + (x_2 \times w_2) + \cdots + (x_n \times w_n)$$

The row vectors of the inputs and weights are x = [x_1, x_2, … , x_n] and w =[w_1, w_2, … , w_n] respectively and their dot product is given by

$$x.w = (x_1 \times w_1) + (x_2 \times w_2) + \cdots + (x_n \times w_n)$$

Hence, the summation is equal to the dot product of the vectors x and w

$$\sum = x.w$$

Step 2: Let's call this z, the result of adding bias b to the multiplied numbers. To obtain the desired output values, a bias—also known as an offset—must be applied to the activation function as a whole.

$$z = x.w + b$$

Step 3: Transform z into a non-linear activation function with the help of the given value. Neuronal output would be a linear function without activation functions, which are utilised to induce non-linearity. The learning pace of the neural network is also significantly affected by these factors. The activation function of perceptron’s is a binary step function. Our activation function will, however, be the sigmoid function, also known as the logistic function.

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

We receive the projected value following forward prorogation, which represents the sigmoid activation function.

3.1 Learning Algorithm

Backpropagation and optimization make up the learning algorithm.

Backpropagation: For the purpose of determining the gradient of the loss function, the term backpropagation (short for backward propagation of mistakes) might be used. Learning algorithm is a more general phrase that is typically used interchangeably with "learning process." Perceptron backpropagation is described in the following steps:

Step 1: A loss function is used to estimate how far we are from our target answer. For regression issues, the loss function is often mean squared error, and for classification problems, cross entropy. A regression issue using the loss function of mean squared error, which is the squared difference between predicted (yi) and actual values I is shown here.

$$MSE_i = (y_i - \hat{y}_i)^2$$

Loss function is calculated for the entire training dataset and their average is called the Cost function C.

$$C = MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Step 2: We need to know how the cost function varies in respect to weights and bias in order to identify the optimal weights and bias for our Perceptron. Gradients (rate of change) are used to show how the change in one quantity compares to the change in the other. Using the weights and biases, we need to calculate the cost function's gradient.
Let's use partial derivation to figure out the gradient of the cost function C with regard to the weight wi. Using the chain rule instead of the cost function, which is dependent on weight wi, we can get a more accurate estimate of the cost.

$$\frac{\partial C}{\partial w_i} = \frac{\partial C}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z} \times \frac{\partial z}{\partial w_i}$$

Now we need to find the following three gradients

$$\frac{\partial C}{\partial \hat{y}} = ? \qquad \frac{\partial \hat{y}}{\partial z} = ? \qquad \frac{\partial z}{\partial w_1} = ?$$

Let's start with the gradient of the cost function (C) with respect to the predicted value (ŷ)

$$\frac{\partial C}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 2 \times \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$$

Let $y = [y_1, y_2, \ldots y_n]$ and $\hat{y} = [\hat{y}_1, \hat{y}_2, \ldots \hat{y}_n]$ be the row vectors of actual and predicted values. Hence the above equation is simplified as

$$\frac{\partial C}{\partial \hat{y}} = \frac{2}{n} \times sum(y - \hat{y})$$

Now let's find the gradient of the predicted value with respect to the z. This will be a bit lengthy.

$$\begin{aligned}\frac{\partial \hat{y}}{\partial z} &= \frac{\partial}{\partial z}\sigma(z) \\ &= \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right) \\ &= \frac{e^{-z}}{(1+e^{-z})^2} \\ &= \frac{1}{(1+e^{-z})} \times \frac{e^{-z}}{(1+e^{-z})} \\ &= \frac{1}{(1+e^{-z})} \times \left(1 - \frac{1}{(1+e^{-z})}\right) \\ &= \sigma(z) \times (1 - \sigma(z))\end{aligned}$$

The gradient of z with respect to the weight w_i is

$$\begin{aligned}\frac{\partial z}{\partial w_i} &= \frac{\partial}{\partial w_i}(z) \\ &= \frac{\partial}{\partial w_i}\sum_{i=1}^{n}(x_i.w_i + b) \\ &= x_i\end{aligned}$$

Therefore, we get,

$$\frac{\partial C}{\partial w_i} = \frac{2}{n} \times sum(y - \hat{y}) \times \sigma(z) \times (1 - \sigma(z)) \times x_i$$

3.2 Estimate the Biasing

Bias is theoretically considered to have an input of constant value 1. Hence,

$$\frac{\partial C}{\partial b} = \frac{2}{n} \times sum(y - \hat{y}) \times \sigma(z) \times (1 - \sigma(z))$$

The selection of the optimal weights and biases for the perceptron is an example of an optimization. This approach uses gradient descent to update the weights and biases, based on the gradient of the cost function with respect to each weight or bias. Let's use this algorithm. An important hyperparameter for managing the amount by which weights and bias are altered is learning rate (). Backpropagation and gradient descent are performed until convergence, with the weights and biases adjusted as shown.

$$\begin{aligned}w_i &= w_i - \left(\alpha \times \frac{\partial C}{\partial w_i}\right) \\ b &= b - \left(\alpha \times \frac{\partial C}{\partial b}\right)\end{aligned}$$

IV. Conclusion

Pattern recognition, prediction, system identification, and control may all benefit from the employment of neural networks. Artificial or artificial neural networks (ANNs), parallel distributed processing (PDP), neuromorphic or connectionist models, and other terminology for similar models and methodologies have been created to recreate how the human brain processes information and gains knowledge. Many people use the term "network" to represent everything from computer networks to communications to organisations and markets. In the outset, the notion of an artificial neural network (ANN) was developed as an optimistic vision of artificial intelligence (AI) synthesis by mimicking the human brain. To include neural-inspired ideas in artificial intelligence applications, ANNs give an alternative to symbol programming. Choosing the best weights and biases for the perceptron is an example of optimization in action, and it can be seen here. Gradient descent is used to update the weights and biases in this strategy. The weights and biases are updated depending on the gradient of the cost function with respect to each weight or bias. This paper explored the algorithm to understand the ANN methodology. The learning rate is a critical hyperparameter for controlling the amount by which weights and bias are updated throughout the optimization process. Backpropagation and gradient descent are carried out until convergence is achieved, with the weights and biases being changed as illustrated in this paper.

References

1. Pavlenko, I., Trojanowska, J., Ivanov, V., & Liaposhchenko, O. (2018, June). Scientific and methodological approach for the identification of mathematical models of mechanical systems by using artificial neural networks. In *International Conference on Innovation, Engineering and Entrepreneurship* (pp. 299-306). Springer, Cham.
2. Elkatatny, S., Tariq, Z., Mahmoud, M., Mohamed, I., & Abdulraheem, A. (2018). Development of new mathematical model for compressional and shear sonic times from wireline log data using artificial intelligence neural networks (white box). *Arabian Journal for Science and Engineering*, *43*(11), 6375-6389.
3. Kilickap, E., Yardimeden, A., & Çelik, Y. H. (2017). Mathematical modelling and optimization of cutting force, tool wear and surface roughness by using artificial neural network and response surface methodology in milling of Ti-6242S. *Applied Sciences*, *7*(10), 1064.
4. Kuo, C. C. J. (2016). Understanding convolutional neural networks with a mathematical model. *Journal of Visual Communication and Image Representation*, *41*, 406-413.
5. Shanmuganathan, S. (2016). Artificial neural network modelling: An introduction. In *Artificial neural network modelling* (pp. 1-14). Springer, Cham.
6. Mohamadou, Y., Halidou, A., & Kapen, P. T. (2020). A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Applied Intelligence*, *50*(11), 3913-3925.
7. Vo, N. D., Oh, D. H., Hong, S. H., Oh, M., & Lee, C. H. (2019). Combined approach using mathematical modelling and artificial neural network for chemical industries: Steam methane reformer. *Applied Energy*, *255*, 113809.
8. Ceylan, I. (2008). Determination of drying characteristics of timber by using artificial neural networks and mathematical models. *Drying Technology*, *26*(12), 1469-1476.