

## Overview of the Research Status on Artificial Neural Networks

WANG Xin-gang

Administration Center for the DTH Service in China, SAPPRFT, No. 2, FuXingMengWai Street,  
XiChen District, Beijing, China

wangxingang@huhutv.net

**Keywords:** Artificial Intelligence; Neural Network; Machine Learning

**Abstract:** With the development of artificial intelligence, artificial neural network algorithm models that simulate the working pattern of human brain neurons have attracted more and more attention. The ability of artificial neural network on feature extraction and classification, has been widely used in image recognition, speech recognition, natural language processing, trend prediction and other fields that effectively assists or replaces the work which needs to be done by human image thinking. At the same time, artificial neural network is a machine learning system, and it will get better results with the progress of training. Based on the development history and research status of artificial neural networks, this paper analyzes and compares several neural network technology models and algorithms, and puts forward the analysis ideas and opinions.

### Introduction

With the artificial intelligence is recognized by the academia, more and more machine learning technologies have emerged, among which artificial neural network is one of the best. Developing artificial neural networks is inspired by the biological nervous system. The biological neuronal cells are connected by synaptic connections and become complex neural network systems, when a creature is externally stimulated, signals are transmitted through neurons, and the creatures' consciousness is activated to make corresponding feedback. Artificial neural network simulates this characteristic, and it connects a large number of neuron nodes, transfers information through appropriate activation functions, and thus information process is performed anthropomorphically.

The human brain consists of three thinking modes: abstract thinking, image thinking and inspiration thinking. Among them, image thinking is the combination of subjective cognition and acquired knowledge to produce inspiration or ideas, which is the exact development direction of artificial intelligence, and make it possible for machines to learn knowledge and produce subjective cognition that can replace human work. Artificial neural network is an effective tool to simulate human image thinking which can help or even replace human beings to work in many tasks that need subjective identification or prediction.

Entering the 21st century, artificial neural networks have entered the stage of high-speed development, and have spawned excellent models and algorithms, such as deep neural networks and convolutional neural networks, and have been applied in many realistic scenarios. At the same time, the evaluation indexes of artificial neural network can be divided into two directions of theory and application. In theoretical aspects, the effect of a algorithm can be evaluated from the accuracy of recognition or prediction. In addition, artificial intelligence is in the development period, and construction of a complete theoretical system is also a key to the development of artificial neural network theory. In application aspects, with the neural network model becoming increasingly complex, the amount of data processing becomes more and more huge, how to reduce the running overhead and improve the learning speed of neural networks is also the focus of researchers. This paper studied the basic theory of neural network first, and analyzed the main research results of artificial neural network from two aspects of theory and practical application.

## **The basic theory of artificial neural network**

Artificial neural network is a multi-layer network model which is connected by neuron nodes. For any neuron in a non-input layer, its output value is determined by the output value of the connected neuron in the lower layer and the corresponding activation function. Neurons in the input layer are used to receive the vector of input parameters, and output to the neurons in next layer. Hidden layers cannot be observed from outside, and neurons are complexly connected, which are used for data processing and feature extraction. There can be one or more neurons in the output layer to output the classification results or estimated values.

A single neuron node establishes a mapping relation from input data to output data. Supposing a neuron uses  $x_1, x_2, x_3$  as the input and  $+1$  as the input values (the function of  $+1$  is to introduce bias value  $b$  to set threshold for the activation function), then the neuron uses  $hw, b(x) = f(WT x) = f(\sum_{i=1}^3 W_i x_i + b)$  as the output,  $f$  as the activation function,  $W$  as the weight parameter of the neural network. The selection of the activation function mainly simulates the behavior of biological neurons, when facing complex nonlinear classification tasks and so on, nonlinear functions are used, such as Sigmoid function, Rectified linear rectifier unit (ReLU) function etc., to active the input data.

The basic structure of artificial neural network includes data input layer, several hidden processing layers and data output layer. The most important parameter in the neural network is the weight  $W$ , and there is a weight value between each of two connected neurons. Except for the input layer, any neuron in the neural network accepts the product of the output data from the connected neuron of the low layer and their weight, after all the the products are summed, and then the output value of this neuron can be calculated by the activation function. The training of neural network mainly refers to constantly adjusting the weight parameters in the network to be optimal.

Artificial neural network is a machine learning system, which needs training and learning through big data to achieve practical application effect. At present, the most commonly used neural network training algorithm is the Back Propagation (also known as the BP algorithm)[3]. The execution steps of the BP algorithm are: firstly, take data from the training data set as the input of the neural network, calculate the output values of the neural network by forward propagation, and define the mean square cost function, so that the error between the actual output and the expected output value of the training data set is minimized, and then adjust the weight parameters reversely layer by layer through the Gradient Descent Algorithm[4], thus the first training process of the neural network is completed. For the neural network which is trained through multiple iterations, if the error between the network output value and the expected value fluctuates in a small stable value range, the neural network can be considered convergent, and can be used in practice.

## **The Development of Artificial Neural Network**

From the idea of artificial neural network to the present high-speed development, the development of artificial neural network has mainly gone through three stages, the first representation of neural network prototype is the perceptron model, neurons are linearly connected, but there are almost no practical application values. After introducing nonlinear connection, the back propagation algorithm is invented, which laid the foundation for the training of neural network. After entering the 21st century, deep learning has become the mainstream of the neural network ideas, and the application of neural network is greatly improved.

### **Perceptron model**

In the field of machine learning, perceptron is a supervised learning algorithm used for binary classifiers (it can determine whether the input represented by a numeric vector belongs to a particular class of function)[5]. It is a linear classifier, which uses a classification algorithm based on the linear prediction function combining the weight set and the feature vector to predict. Perceptron algorithm can be traced back to 1950s, and it is one of the first implementation of artificial neural networks in customized hardware.

Before the birth of perceptron model, experts in neuroscience and cybernetics and etc. began to think about the possibility of intelligent computing tools, logician Walter Pitts and cybernetics expert Warren McCulloch designed a neural network computing prototype based on mathematics and threshold logic algorithm, called MP model[6]. This linear model classify the input by judging the positive and negative values of the activation function, which is the earliest linear classification model. Since then, the research of neural network has been extended from the study of biological nervous system to the study of artificial neural network model.

Inspired by the MP model, psychologist Frank Rosenblatt invented the perceptron model[7] in 1957. The perceptron is designed for character image recognition. In the background of neural networks, the perceptron is an artificial neuron that uses the Heaviside step function as the activation function. The perceptron algorithm is also called a single-layer perceptron, and as a linear classifier, a single-layer perceptron is the simplest feedforward neural network, and it uses the idea of hidden neurons, which is referred to many times by later algorithms. Although the perceptron was initially promising, it was soon proven unable to be trained to identify many types of patterns and could not solve the problem of XOR circuits, which also led to the study of neural networks stagnating for many years.

### **Back propagation algorithm**

Back propagation is a method used in artificial neural network to calculate each neuron's error contribution after a batch of data processing. It is used to adjust the weight parameters of each neuron and complete the learning process. Technically, this algorithm calculates the gradient of the loss function and uses the gradient descent optimization algorithm, so it is also called the error back propagation. In the training process of the neural network, each input value and its expected output value are required, and the whole algorithm repeats the two-stage loop: forward propagation and weight update. When the input vector is presented to the neural network, it propagates forward through the network layer by layer until it goes out from the output layer. The output value of the neural network is then compared with the expected output value using the cost function, and calculate the error value for each neuron in the output layer. The error value is then passed backward from the output layer through the neural network until each neuron has an associated error value that reflects its contribution to the original output. Back propagation algorithm uses these error values to compute the gradient of the loss function and uses the gradient descent algorithm to optimize the cost function, in order to try to minimize the cost function. Finally, after adjusting the weights and threshold parameters between the neurons in the network layer by layer, one training process of the back propagation algorithm is completed.

The idea of the back propagation algorithm first appeared in the 1970s. Paul J. Werbos first wrote a paper to describe the training process of artificial neural networks by back propagation error, which is the prototype of the back propagation algorithm. In 1982, John Hopfield proposed the Hopfield neural network model, which is a content-addressable memory system with binary threshold nodes, and it is guaranteed to converge to local minimum values. In addition, Hopfield network also provides a pattern for understanding human memory. Subsequently in 1986, Rumelhart and Hinton etc. developed a back propagation algorithm[10], and described a new cognitive theory called connectionism. With the increasing amount level of data processing, the method of using distributed parallel computing for the training of neural networks gradually become popular.

During the same period of time when the back propagation algorithm developed, Paul Smolensky invented restricted Boltzmann machine (RBM) in 1986. Restricted Boltzmann machine is a kind of random artificial neural network that can be used to learn the probability distribution on its input set. As the name implies, RBM is a variant of the Boltzmann machine, the restriction indicates that neurons have to form a dichotomous graph: a pair of nodes from two unit groups can have symmetrical connections between them, and there are no connections among nodes within the same group. This restriction allows more efficient training algorithms than the general Boltzmann machine, especially for the gradient-based contrast divergence algorithm. RBM has been widely used in dimensionality

reduction, classification, feature learning and theme modeling tasks, and today it still has a wide range of uses, which can also be applied to the deep learning network to be introduced in the next section.

### **Deep Learning**

Deep learning (also known as deep structured learning or hierarchical learning) is part of a broader range of machine learning methods based on learning data representation, rather than a task-specific algorithm. Deep learning can be supervised, partially supervised or unsupervised. Deep learning architectures, such as deep neural network, deep belief network and cyclic neural network etc., have been widely used in image recognition, target detection, speech recognition, natural language processing and other fields. As a representative, deep neural network is an artificial neural network with multiple hidden layers between the input layer and the output layer, which can simulate complex nonlinear relationship. The architecture of deep neural network produces a constituent model, in which objects are represented as basic hierarchical combinations, and extra layers can combine features from lower layers, which implies that it is capable of modeling complex data with fewer units than similar small layer networks. A deep neural network is usually a feedforward network, in which data flows from the input layer to the output layer without looping.

In the 21st century, with the new discoveries in the field of neuroscience, there is a new cortex in human brains that allows neuronal information to be transmitted through complex network systems, and the cortex can learn to describe the signal through constant signal stimulation. This deep-level learning mechanism of the cerebral cortex promoted the development of deep learning[11]. In 2006, Geoffrey Hinton et al, a well-known artificial intelligence scientist, designed a model called Deep Belief Networks(DBNs)[12], using a greedy algorithmic training method. In the early training phase, unsupervised learning algorithm is used; when the number of training reaches a certain requirement, supervised learning algorithm is used instead in order to achieve the best results. This training method is extended to many models, which greatly enhanced the generalization of the neural network model. Deep convolutional neural network is a kind of deep and forward-looking artificial neural network, which shows high performance in image processing and other fields. Convolutional Neural Network (CNN) is a variant of a multi-layer perceptron designed to require minimal preprocessing, also called a shift invariant or spatial invariant artificial neural network, based on its weight sharing and translation invariant characteristics. CNN also widely uses back propagation algorithm as its training method, and relying on its weight sharing and sparse connection characteristics, it can greatly reduce the computational overhead, which has a high practical value. In 1998, Yann LeCun et al. designed a five-layer convolutional neural network LeNet-5 to solve the problem of recognition of handwritten digits, which is a groundbreaking work that used convolution to extract features from images and reduced preprocessing and training costs. In 2012, Alex Krizhevsky and Hinton et al. published AlexNet[14]. Compared to LeNet, AlexNet enhanced the network layers, optimized the activation function between neurons, and introduced the dropout[15] technology and GPU parallel computing technology, which makes the training of neural networks become reality in the era of big data. In recent years, more convolution neural network models, such as Inception Module[16], ResNet[17] etc., have been proposed, and the performance record of image recognition classification using artificial neural networks have been refreshed constantly.

Recurrent Neural Network (RNN) is another mainstream deep neural network model, and the connections between its neurons form a directed loop, which allows it to present dynamic temporal behavior. Recurrent neural networks are characterized by the ability to use their internal memory to process arbitrary input sequences, which cannot be achieved by the feedforward neural network, and makes it suitable for natural language processing, speech recognition and other tasks [18]. In 1997, Sepp Hochreiter and Jürgen Schmidhuber proposed a long short-term memory (LSTM) recurrent neural network[19], which can remember the value of any interval, and the stored values are not modified with the progress of learning. The LSTM network is well-suited to classify, process and predict time sequences of given time lags between important events of unknown size and duration. The relative insensitivity to the interval length makes the LSTM better than the alternative RNN, the hidden

Markov model and other sequence learning methods in many applications. It has become an effective improved recurrent neural network architecture.

Deep learning has become the core direction of the research of artificial neural network today, and Google, Microsoft, Facebook and other high-tech companies invest a lot of human and financial resources in the research and development of deep learning every year. Relying on the development of big data, using deep learning for data processing and analysis has become the most effective way.

## **Conclusion**

This paper mainly introduced the basic theory and development history of artificial neural network, and analyzed the current mainstream neural network models and the future development direction. Today, artificial intelligence has become the main research direction of computer science, and artificial neural network has been widely used in the directions of image processing, speech recognition, natural language semantic processing, data analysis and forecasting etc. Relying on the development of big data, using parallel computing technology to train neural networks make it possible to process massive data. The deep learning technology developed in recent years has become the research direction of future artificial intelligence. Millions of neurons are built in deep neural networks, which increases the ability of data fitting and analysis by several orders of magnitude, and makes it possible to accomplish high intelligence work. By the way, Google's AlphaGo system is a successful case. In summary, the research and development of artificial neural network has a high theoretical significance and practical value for realizing a high artificial intelligence.

## **References**

- [1] Dj.M. Maric, P.F. Meier and S.K. Estreicher: Mater. Sci. Forum Vol. 83-87 (1992), p. 119
- [2] Andrew Ng, Jiquan Ngiam, Chuan Yu Foo, Yifan Mai, Caroline Suen . UFLDL Tutorial. Stanford University, 2013.
- [3] Glorot Xavier, A. Bordes and Y. Bengio. Deep Sparse Rectifier Neural Networks. International Conference on Artificial Intelligence and Statistics. 2011.
- [4] Shuguang Liu, Chongxun Zheng, Mingyuan Liu. Back Propagation algorithm in feedforward neural networks and its improvement: progress and prospect [J]. Computer Science.23(1). 1996:76-79.
- [5] Madgwick S O, Harrison A J, Vaidyanathan A. Estimation of IMU and MARG orientation using a gradient descent algorithm.[C]// IEEE International Conference on Rehabilitation Robotics. IEEE, 2011:1-7.
- [6] Freund Y, Schapire R E. Large margin classification using the perceptron algorithm[J]. Machine Learning, 1999, 37(3):277-296.
- [7] W. McCulloch and W. Pitts. A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics. vol. 10. 1943:115-133.
- [8] F. Rosenblatt. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. vol. 65. 1958:386-408.
- [9] Ke Xu. Research on application of convolutional neural network in image recognition[D]. Zhejiang University:2012.
- [10]P. Werbos. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.[Dissertation]. Harvard University:1974.

- [11] D. E. Rumelhart, G. E. Hinton and R. J. Williams. *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*. MIT Press. 1986.
- [12] T. Lee and D. Mumford. Hierarchical Bayesian Inference in the Visual Cortex. *Journal of the Optical Society of America A*. vol. 20. 2003:1434-1448.
- [13] G. Hinton, S. Osindero. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*. vol. 18. 2006:1527-1554.
- [14] Yann Lecun, Leon Bottou, Yoshua Bengio. Gradient Based Learning Applied to Document Recognition. *Proceedings of IEEE*. 86(11). 1998:2278-2324.
- [15] Alex Krizhevsky, Ilya Sutskever and G. Hinton. Imagenet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems* 25. 2012. 1106-1114.
- [16] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. *Computer Science*, 2012, 3(4): 212-223.
- [17] C. Szegedy. W. Liu, Y. Jia, et al. Going Deeper with Convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015:1-9.
- [18] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015:770-778.
- [19] Sak H, Senior A, Beaufays F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling[J]. *Computer Science*, 2014:338-342.
- [20] Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8):1735-1780.