

Outlook of ANN in Agricultural Data Mining

S.S. Baskar, L. Arockiam and L. Jeyasimman
Department of Computer Science, St. Joseph's College, Trichirappalli, India

Abstract: Organizations are flooded with large quantities of data with the increase in time duration. These data are available in the data base used to extract the information. To extract the information, Artificial Neural Network (ANN) is one of the leading methods in the data mining. Though ANN seems complicated to work, it is very easy to get high accuracy in noisy data sets. This study presents an overview of artificial neural network techniques with respect to data mining.

Key words: Artificial neural network, neurons input layer, output layer, hidden layer

INTRODUCTION

Artificial intelligence and Artificial Neural Networks (ANNs) were considered a promising approach for finding the good learning algorithm to solve practical application problems (Prechelt, 1996) in the early days. Though ANNs is not considered to be the top 10 data mining techniques (Wu *et al.*, 2008), some researchers are using these techniques in the data mining. In general, ANN are used for modeling functions without known mathematical expressions. ANN is mostly used to modeling the functions where statistical approach such as regression and interpolation techniques for modeling. Artificial Neural Networks (ANN) have been developed as generalizations of mathematical models of biological nervous systems. A first wave of interest in neural networks (also known as connectionist models or parallel distributed processing) emerged after the introduction of simplified neurons by McCulloch and Pitts (1943).

Methods of data mining: Few innovative methods for extracting knowledge from data sets are statistical methods. Apart from statistical methods, few novel methods such as artificial neural network are used. This method is contemporary and used in data mining.

CHARACTERISTICS OF NEURAL NETWORKS

Neural network is composed of a large number of very simple processing elements called neurons. Each neuron is connected to other neurons by means of interconnections or links with an associated weight. Memories are stored or represented in a neural network in the pattern of interconnection strengths among the neurons. Information is processed by changing the strengths of interconnections and/or changing the state

of each neuron. A neural network is trained rather than programmed. A neural network acts as an associative memory. It stores information by associating it with other information in the memory. For example, a thesaurus is an associative memory. It can generalize that is it can detect similarities between new patterns and previously stored patterns. A neural network can learn the characteristics of a general category of objects on a series of specific examples from that category. It is robust. The performance of a neural network does not degrade appreciably if some of its neurons or interconnections are lost (distributed memory). Neural networks may be able to recall information based on incomplete or noisy or partially incorrect inputs. A neural network can be self-organizing. Some neural networks can be made to generalize from data patterns used in training without being provided with specific instructions on exactly what to learn.

STRENGTHS OF NEURAL NETWORKS

Neural networks offer a number of advantages such as it require less formal statistical training. It is ability to implicitly detect complex nonlinear relationships between dependent and independent variables. It has also ability to detect all possible interactions between predictor variables and the availability of multiple training algorithms. One of the major advantages of neural nets is their ability to generalize it means that trained set could classify data from the same class as the learning data that it has never seen before. It has the advantage called self-organization. It means that learning without a teacher. Neural network has the ability to recall information based on incomplete or noisy or partially incorrect inputs. Neural network performs well in data-intensive applications. It also performs well where standard technology is inadequate and data is intrinsically noisy and error-prone.

Neural network plays good role where the qualitative or complex quantitative reasoning is required. Algorithms based neural network have been used in knowledge engineering and human decision making (Tan *et al.*, 1996; Towell and Shawlik, 1994).

APPLICATIONS OF NEURAL NETWORKS

Artificial neural networks have been applied for areas such as science and engineering and management studies. Most commonly used areas are fraud detection, telecommunication, medicine, marketing, bankruptcy prediction insurance and agriculture:

- Economic modeling
- Detection
- Mortgage application
- Oil refinery production
 - Assessments
 - Forecasting
- Sales lead assessments
- Foreign exchange analysis
- Disease diagnosis
- Market and customer behavior analysis
- Manufacturing quality control
- Optimal resource allocation
- Sports forecasting
- Financial investment analysis
- Process fault detection
- Optical character recognition
- Bond rating
- Optimization
- Credit card fraud

LIMITATIONS OF NEURAL NETWORK

The major drawback of neural network is difficult to explain. In this case, blackbox approach is followed to solve the problems. There are no established development guidelines. Neural network is not appropriate for all types of problems.

ARCHITECTURE NEURAL NETWORK

In general, following neural network architecture are usually used in data mining:

- Multilayered feedforward neural networks
- Kohonen's self-organizing maps

In a neural network as in Fig. 1, neurons are grouped into layers. The neurons in each layer are the same type. There are different types of layers:

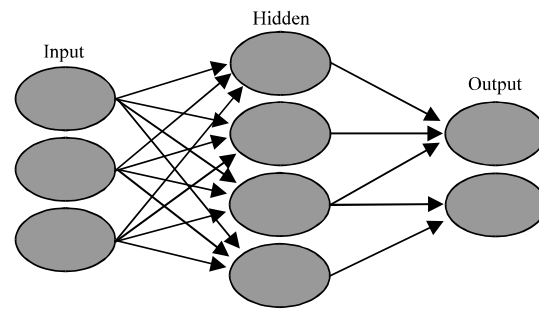


Fig. 1: Network

- The input layer consists of neurons that receive input from the external environment
- The output layer consists of neurons that communicate to the user or external environment
- The hidden layer consists of neurons that only communicate with other layers of the network

Input to neurons: A neuron receives inputs from other neurons. It does not matter whether the input to neuron comes from the neurons on the same layer or on another layer. The message that a neuron receives from another neuron is modified by the strength of the connection between the two. The net input to a neuron is the sum of all the messages it receives from all neurons it is connected to. There may also be external input to a neuron. The net input to a neuron creates an action potential. When the action potential reaches a given level, the neuron fires and sends out a message to the other neurons.

Output from neurons: Neuron sends output or action potential to the other neurons. The action potential or the net input to the neuron should go through a filter or transformation. This filter is called activation function or transfer function. There are number of different activation functions:

- Step function
- Signum function
- Sigmoidal function
- Hyperbolic tangent function
- Linear function
- Threshold linear function

CONNECTIONS OF NEURONS

The neurons are connected in the layers. They are categorized in this study:

Inter-layer connections: Inter layer connection is one of the mechanism where the neurons are connected with in the neuron. They are further categorized in the study.

Fully connected: Fully connected neurons are the connection of neurons in one layer is connected with another layer.

Partially connected: It explains that each neuron in the first layer does not have to be connected to all neurons on the second layer.

Feed-forward: It means that where neurons on the first layer send their outputs to the neurons of the second layer but they do not receive any input back from the neurons in the second layer.

Feed-backward: Here it explains that where the output signals from the neurons on a layer are directly fed back the neurons in the same of preceding layer.

Bi-directional: There is a set of connections going from neurons of the first layer to those of the second layer. There is also another set of connections carrying outputs of the second layer into the neurons of the first layer.

Intra-layer connections: Intra layer connections of neurons are neurons on a layer with other neurons of the same layer.

Recurrent: The neurons within a layer are fully or partially connected to one another. When neurons on a layer receive input from another layer, they communicate their outputs with one another a number of times before they are allowed to send their outputs to the neurons in another layer.

On-center/off-surround: A neuron within a layer has excitatory connections to itself or its immediate neighbors and has inhibitory connections to other neurons. Sometimes it is called self-organizing.

STATES OF A NETWORK

The states of neural network can be categorized as training state and operation state.

Training state: In the training state, what the network uses the input data to change its weights to learn the domain knowledge, the system is said to be in training mode. This is the mode in which the network learns new knowledge by modifying its weights. The network's

weights are gradually changed in an interactive process. The system is repeatedly presented with the case data from a training set and is allowed to change its weights according to training method.

Operation state: The state in which when the system is being used as a decision tool, it is in the operation mode. The interconnection weights do not change when the network is in the operation mode.

LEARNING STATE NEURAL NETWORKS

The network has already achieved learning prior to the presentation on new data and information. The network's training mode precedes the operation mode. The process neural network uses to compute the interconnection weights among neurons. For the neural networks to learn, it should be given a learning method to change the weights to the ideal values. Learning methods is the most important distinguishing factor in various neural networks. There are two distinct types of learning.

Supervised learning: In supervised learning, the system developer tells the network what the correct answer is and the network determine the weights in such a way that once given the input, it would determine the output. The idea in supervised learning is that the network is repeatedly given facts about the various cases, along with the expected outputs. The network uses the learning method to adjust the weights in order to produce the outputs close to what is expected.

Unsupervised learning: In the unsupervised learning, the network receives only the inputs and no information about on the expected output. The system learns to produce the pattern of what it has been exposed to. These networks are sometimes referred to as self organizing networks.

APPLICATION OF ANN IN AGRICULTURE

These techniques are used in irrespective of the research areas. It is also used in the agriculture domain also for finding new model. Few of this neural network application in the agriculture are listed here. Further detailed techniques can be researched in the respective paper in the references. Classifications of fertile and infertile eggs by machine vision (Das and Eans, 1992) are made using this neural network technique. Similarly prediction of flowering and maturity dates of soybean, forecasting water resources variables (Maier and Dandy, 2000), detection of pig coughs in farms by recorded

sounds (Chedad *et al.*, 2001), detection of watercores in apples by X-ray images (Shahin *et al.*, 2001), wine classification by taste sensors made from ultra thin films (Riul *et al.*, 2004) and Modeling of sediment transport (Bhattacharya *et al.*, 2007) are few of the neural network techniques implemented for modeling.

CONCLUSION

Data mining is a novel area of data analysis research to find the novelty on the data and infer the knowledge out of the data. The specialty of neural network and its characteristics are robustness, self organizing adaptive, high degree of fault tolerance and parallel processing. Data mining using neural network model suits for efficient data mining. From this study of neural network techniques, we can come for conclusion; neural network will receive more attention on data mining.

REFERENCES

- Bhattacharya, B., R.K. Price and D.P Solomatine, 2007. A machine learning approach to modeling sediment transport, ASCE. J. Hydraulic Eng., 133: 440-450.
- Chedad, A., D. Moshou, J.M. Aerts, A. Van Hirtum, H. Ramon and D. Berckmans, 2001. Recognition system for pig cough based on probabilistic neural networks. J. Agric. Eng. Res., 79: 449-457.
- Das, K.C. and M.D. Eans, 1992. Detecting fertility of hatching eggs using machine vision. II: Neural network classifiers. Trans. Am. Soc. Agric. Eng., 35: 2035-2041.
- Maier, H.R. and G.C. Dandy, 2000. Neural networks for the prediction and forecasting of water resources variables: A review of modeling issues and applications. Environ. Modeling Software, 15: 101-124.
- McCulloch, W.S. and W. Pitts, 1943. A logical calculus of ideas immanent in neural activity. Bull. Math. Biophys., 5: 115-133.
- Prechelt, L., 1996. A quantitative study of experimental evaluations of neural network learning algorithms. Curr. Res. Prac. Neur. Networks, 9: 457-462.
- Riul, Jr. A., H. de Souse, R.R. Malmegrim, D.S. Jr. dos Santos and A.C.P.L.F. Carvalho *et al.*, 2004. Wine classification by taste sensors made from ultra-thin films and using neural networks. Sensors Actuators B Chem., 98: 77-82.
- Shahin, M.A., E. Tollner and R. McClendon, 2001. Artificial intelligence classifiers for sorting apples based on watercore. J. Agric. Eng. Res., 79: 265-274.
- Tan, C.L., T.S. Quah and H.H. The, 1996. An artificial neural network that models human decision making. Computer, 29: 64-70.
- Towell, G. and J.W. Shawlik, 1994. Knowledge based artificial neural networks. Artificial Intellig., 70: 119-165.
- Wu, X., V. Kumar, J.R. Quinlan, J. Ghosh, D.J. Hand and D. Steinberg *et al.*, 2008. Top 10 algorithms in data mining. Knowledge Inform. Syst., 14: 1-37.