Attribute Selection Technique for the Training of Artificial Neural Network Predictive Model

Zakia Hussain, Shoukat Ullah, Aman Ullah
Institute of Information Technology, UST Bannu, N-W.F.P, Pakistan
zakiahussain_bn@yahoo.com, shoukatullah@teacher.com, am_ullah@hotmail.com

Abstract

In last several years, databases have created lots of innovations in various areas of life. Organizations are collecting high quality data on a large scale. The huge amount of data can be a gold mine for business organizations and management. It is therefore increasingly important to analyze the data. However, timely and accurately processing of data in traditional methods is a difficult task. This paper introduces an enhanced algorithm that helps Artificial Neural Network (ANN) to predict faster with accurate results. Artificial Neural Networks involve long learning or training time, complex structure for larger input dimensionality and poor interpretability. The enhanced algorithm selects only those relevant attributes that take part in the learning process of ANN. Ultimately, by selection of relevant attributes, the complexity and training time of the network are also reduced.

Keywords – Artificial Neural Network, Backpropagation, Data Mining, Data Prediction

1. Introduction

Data mining is defined as the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [1].

Data mining is an essential step in the process of knowledge discovery in databases (KDD). KDD is an iterative process and containing the following steps [2]:

i. Data Cleaning
ii. Data Integration
iii. Data Selection
iv. Data Transformation
v. Data Mining
vi. Pattern Evaluation
vii. Knowledge Representation

Data mining can be classified either as descriptive or predictive mining, which performs the tasks of classification, prediction, and regression, or clustering and association respectively.

Data classification and prediction are two-step processes to construct a model and then to use the model for classifying future or unknown objects [3].

This study discusses attribute selection and data prediction techniques. By attribute selection technique we select only those attributes which contribute in the learning process of ANN. The training is carried out by feedforward Multilayer Perceptrons (MLPs) network, which is a widely used model in ANN using the backpropagation algorithm [4].

The field of Neural Network (NN) was originally kindled by psychologists and neurobiologists who sought to develop and test computational analogues of neurons which are building blocks of human brain and nervous system [5]. An ANN consists of an input layer, one or more hidden layers, and an output layer (see Figure 1), where each connection has a weight associated with it. During the learning phase, the network learns by adjusting the weights so as to be able to predict the correct class of the input tuples/samples [3].

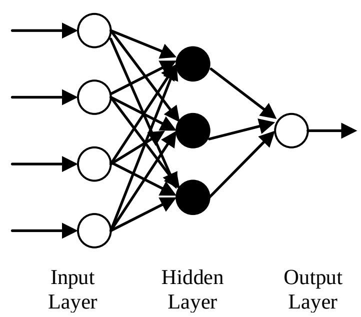

Figure 1. Feedforward MLPs network

The rest of the paper is organized as follow:

Section 2 provides review of related literature, section 3 encompasses an enhanced backpropagation algorithm, section 4 includes experiments on the real dataset using Neuralworks Predict Version 3.22, and section 5 covers conclusion and future work.

2. Related Literature

2.1. An Artificial Neural Network

"... a neural network (NN) is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes". [6]

A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects [7]:

- Knowledge is acquired by the network through a learning process.
- Interneuron connection strengths known as synaptic weights are used to store the knowledge.

Artificial neural systems, or neural networks, are physical cellular systems which can acquire, store, and utilize experiential knowledge. [8]

2.2. Types of ANN

There are many types of ANNs that has been employed and these include MLPs, radial basis function (RBF) and Bayesian NNs [11, 12]. New pattern recognition methods called Support Vector Machines (SVMs) and Gaussian Mixture Models (GMMs) have been proposed and found to be particularly suited to classification problems [13]. The most well known types of ANN are following:

A. Feedforward NN

In a feedforward NN, the connections between units do not form cycles. It usually produces a response to an input quickly. The Feedforward Multilayer Perceptrons [14, 15, 9] network is a widely used model in Artificial Neural Network using the backpropagation algorithm [16] for real world data. However, it is often avoided due to the large size of network and the training that would be too slow to be tolerable.

B. Feedback NN

In a feedback or recurrent NN, there are cycles in the connections. In some feedback NNs, each time an input is presented, the NN must iterate for a potentially long time before it produces a response. Feedback NNs are usually more difficult to train than feedforward NNs.

2.3. Training Set, Validation Set, and Test Set

In NN methodology, the sample (usually a subset of the population [16] is often subdivided into "training", "validation", and "test" sets. The distinctions among the "validation" and "test" sets are given below [17]:

Training set:

A set of examples used for learning that is to fit the parameters [i.e., weights] of the classifier.

Validation set:

A set of examples used to tune the parameters [i.e., architecture, not weights] of a classifier, for example to choose the number of hidden units in a neural network.

Test set:

A set of examples used only to assess the performance [generalization] of a fully-specified classifier.

2.4. Backpropagation Algorithm

"Standard backpropagation" is a training algorithm that was popularized by [16] which remains the most widely used supervised training method for neural networks. It can be used for both batch training (in which the weights are updated after processing the entire training set) and incremental training (in which the weights are updated after processing each case). For batch training, standard backpropagation usually converges (eventually) to a local minimum, if one exists. For incremental training, standard backpropagation does not converge to a stationary point of the error surface.

3. The Enhanced Backpropagation Algorithm

```
ALGORITHM: AST_BACKPROPAGATION
//Attribute selection technique and backpropagation
//algorithm
Input:   Dataset
Output: Trained neural network
(1) for each attribute A {
(2)          apply attribute selection method
             and obtain new groups of attributes
(3) All weights and biases are initialized
(4) Repeat Steps 5 to 9 while terminating condition is not satisfied
{
(5) for each training tuple R in Data set {
(6) Calculate Net input and output of
             each unit in hidden and output layers
(7) Calculate Error of each unit in hidden and output layers
(8) if Error exists Then
(9) Calculate updated weights and biases}}
```

Figure 2. The enhanced backpropagation algorithm

The excerpts of an enhanced algorithm are highlighted below:

- In steps 1 to 2, attribute selection technique [3] is carried out to calculate the Information gain of the attribute.
  Gain (A) = Info (D) – $Info_A$ (D) ---------- 1
- In step 3 the weights and biases are initialize small random numbers, ranging from -1.0 to 1.0 or -0.5 to 0.5.
- In step 4 a loop structure is employed to repeat steps 5 to 9.
- In step 5 a loop structure is again employed, where as in body of loops steps net input I_j and output O_j are calculated for each unit of hidden and output layers, an error of units in hidden and output layers are computed, and at the end weights and biases are updated [3].

$$I_j = \sum w_{ij} O_j + b_j \qquad \text{-------------} \quad 2$$

$$O_j = \frac{1}{1 + e^{-} I_j} \qquad \text{-------------} \quad 3$$

$$E_j = O_j\,(1 - O_j)\,(T_j - O_j) \qquad \text{-------------} \quad 4$$

$$E_j = O_j\,(1 - O_j) \sum_k E_k\, w_{jk} \qquad \text{----------} \quad 5$$

$$w_{ij} = w_{ij} + (l)\, E_j\, O_i \qquad \text{-------------} \quad 6$$

$$b_j = b_j + (l)\, E_j \qquad \text{-------------} \quad 7$$

Table 1. Information gain, mean deviation and percentages of different attributes

Attribtes	Info(A)	M.D	%age
A1	0.338	0.031	91.59891599
A2	0.338	0.031	91.59891599
A3	0.338	0.031	91.59891599
A4	0.338	0.031	91.59891599
A5	0.338	0.031	91.59891599
A6	0.338	0.031	91.59891599
A7	0.616	0.247	33.06233062
A8	0.276	0.093	74.79674797
A9	0.276	0.093	74.79674797
A10	0.338	0.031	91.59891599
A11	0.553	0.184	50.13550136
A12	0.338	0.031	91.59891599
Average			**0.369**

4. Experiments

Experiments have been performed on real dataset of admission data in MS-Computer Science [18] that contains one hundred and eleven (111) instances and twelve (12) attributes with a class label. Information gain of each attribute is calculated by the formula given in Eq. 1, then Mean Deviation (M.D) and percentages of M.D are computed (see Table 1). According to percentages, different groups of attributes are selected and at the end these groups of attributes are trained by backpropagation algorithm using Neuralworks Predict Version 3.22 software (see Figures 3 and 4). The results are summarized in Table 2.

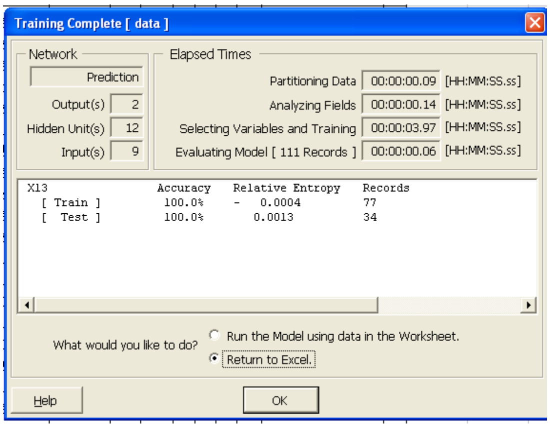

Figure 3. Training results of Group No. 4

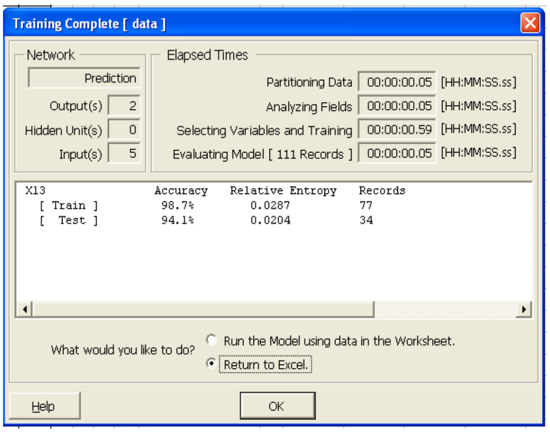

Figure 4. Training results of Group No. 3

Table 2. Training time and test accuracy of different groups

Group No.	%age Classes	Attributes	Training Time	Test Accuracy
G1	0-25	Nil	Nil	No
G2	0-50	A7, A11	0.47	No
G3	0-75	A7, A8, A9, A11	0.59	Yes
G4	0-100	All Attributes	3.97	Yes

4.1. Running a Model with New Instances

The reason for building any model is so that output values can be accurately predicted when a new data record is processed by the model. The results of the new instances are shown in Table 3.

Table 3. The results of new instances

A7	A8	A9	A11	A13
CS	15	1st	44	No
CS	16	1st	71	Yes
IT	16	2nd	66	No
CS	16	1st	58	Yes
CS	16	1st	76	Yes

5. Conclusion and Future Work

In this research paper, we have introduced an enhanced backpropagation algorithm which was able to train the real dataset in relatively short amounts of time as compared with traditional MLPs algorithm. The enhanced algorithm decreased 58% training time as compared to traditional MLPs.

The preliminary results of the experiments are promising, but not enough. There is a lot more work to do. More instances of the population should be introduced to the training set, the effects of different attribute selection methods should be examined, analysis of different activation functions at hidden and output layer should be compared, improvements of the combined methods like statistical analysis and neural networks should be examined, and performance of system should be verified in large databases or data warehouses as well as performance can be checked in the real-time and in the real environment.

References

[1]. Fayyad U., Piatetsky-Shapiro G., and Smyth P., *Data Mining to Knowledge Discovery: An overview*, CA: AAAI Press, 1996.

[2] Chaudhuri S., *Data Mining and Database Systems : Where is the Intersection?*, IEEE Bulletin of the Technical Committee on Data Engineering, Vol.21 No.1, pp. 4-8, March 1998.

[3] Han J., and Kamber M., *Data Mining: Concepts and Techniques*, Elsevier India Pvt Limited, Second Edition, 2007.

[4] Pieter A., and Dolf Z., *Data Mining*, Dorling Kindersley Publishing Inc (India).

[5] Richard S., *Clinical Neuroanatomy*, 6th Edition, Lippincot William and Walkins 530 Walnut Street Philadelphia, PA 19106.

[6] DARPA Neural Network Study, AFCEA International Press, p. 60, 1998.

[7] Haykin S., *Neural Networks: A Comprehensive Foundation*, NY: Macmillan, 1999.

[8] Zurada J.M., *Introduction to Artificial Neural Systems*, Boston: PWS Publishing Company, p. xv, 1992.

[9] Fausett L., *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*, Prentice-Hall, New Jersey, USA, 1994.

[11] Marwala T., *Fault Identification Using Neural Networks and Vibration Data*, University of Cambridge Ph.D. Thesis, Cambridge, UK, 2001.

[12] Bishop C.M., *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK, 1995.

[13] Joachims J., *Making large-scale SVM learning practical*. Scholkopf B., Burges C. J. C., and Smola A.J., *Advances in Kernel Methods-Support Vector Learning*, 169-184, MIT Press, Cambridge, MA: ,1999.

[14] Bishop, C.M., *Neural Networks for Pattern Recognition*, Oxford: Oxford University Press, 1995.

[15] Reed, R.D., and Marks, R.J. II., *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*, Cambridge, MA: The MIT Press, ISBN 0-262-18190-8, 1999.

[16] Rumelhart, D.E., Hinton, G.E., and Williams, R.J., *Learning Internal Representations By Error Propagation*, in Rumelhart, D.E., and McClelland, J. L., eds. (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Volume 1, 318-362, Cambridge, MA: The MIT Press, 1986.

[17] Ripley, B.D., *Pattern Recognition and Neural Networks*, Cambridge: Cambridge University Press, 1996.

[18] Institute of Information Technology, University of Science & Technology Bannu, N-W.F.P, 2006.