

Parallel Approach for Diagnosis of Breast Cancer using Neural Network Technique

Dr. K. Usha Rani
Dept. of Computer Science
Sri Padmavathi Mahila isvavidyalayam (Women's University)
Tirupati , Andhra Pradesh

ABSTRACT

Classification is perhaps the most familiar and popular data mining technique. Inspired by biological neural networks, Artificial Neural Networks are developed to mimic the characteristics such as robustness and fault tolerance. To perform classification task of medical data, the neural network is trained. To speed up the training process parallel approach is adopted. In this paper a parallel approach by using neural network technique is proposed to help in the diagnosis of breast cancer. The neural network is trained with breast cancer data base by using feed forward neural network model and backpropagation learning algorithm with momentum and variable learning rate. The performance of the network is evaluated. The experimental result shows that by applying parallel approach in neural network model yields efficient result.

Keywords

Classification, Neural Networks, Parallelism, feed forward, backpropagation, Breast Cancer.

1. INTRODUCTION

Data mining is an essential step in the process of knowledge discovery in databases in which intelligent methods are applied in order to extract patterns. Parallelism offers a natural and promising approach to cope with the problem of efficient data mining in large databases. There has been considerable interest in parallel processing of data mining algorithms [1,2].

Classification is an important problem in the rapidly emerging field of data mining. It has been studied extensively by the machine learning community as a possible solution to the knowledge acquisition or knowledge extraction problem. The input to a classifier is a training set of records, each of which is tagged with a class label. A set of attribute values defines each record. Attributes with discrete domain are referred to as categorical, while those with ordered domains are referred to as numeric. The goal is to induce a model or description for each class in terms of the attributes. The model is then used to classify future records whose classes are unknown. Among the techniques developed for classification, popular ones include Bayesian classification, Neural Networks, Generic Algorithms and Decision Trees.

The decision tree approach is most useful in classification problems. With this technique, a tree is constructed to model the classification process. Once the tree is built, it is applied to each tuple in the database and results in a classification for that tuple. Disadvantages of decision trees are, they do not easily handle

continuous data. Handling missing data is difficult because correct branches in the tree could not be taken. Correlation among attributes in the database are ignored by the decision tree process.

Neural network is one of the most used data mining method to extract patterns in an intelligent and reliable way and has been greatly used to find models that describe data relationship [3,4]. Neural networks, Fuzzy sets and Genetic algorithm applications in data mining are discussed in the survey of data mining using soft computing [5]. In this paper a neural network technique is proposed to detect breast cancer. To fasten the process parallel computation also adopted.

2. NEURAL NETWORK FOR CLASSIFICATION

Another technique that is commonly applied for solving data mining problem is the Artificial Neural Networks (ANN). Originally inspired by biological models of mammalian brains, ANN have emerged as a powerful technique for data analysis. Neural Networks consists of compositions of single, non linear processing units that are organized in a densely inter connected graph. A set of parameters, called weights, are assigned to each of the edges of the graph [6]. These parameters are adapted through the local interactions of processing units in the network. By repeatedly adjusting these parameters, the neural network is able to construct a representation of a given data set. This adaptation process is known as training. Neural Network is able to solve highly complex problems due to the non linear processing capabilities of its neurons. In addition, the inherent modularity of the neural network structure makes it adaptable to a wide range of applications.

One of the main limitations of applying neural networks to analyze massive data mining databases is the excessive processing that is required. It is not uncommon for a data mining neural networks to take weeks or months to complete its task. This time constraint is infeasible for most real work applications. However, processing time can be substantially reduced by distributing the load of computation among multiple processors. Thus, parallelism presents a logical approach to managing the computation costs of data mining applications.

2.1 Neural Networks Parallelization Strategies

There are a variety of different parallelization strategies which have been considered for Artificial Neural Networks [7]. Due to the modularity of the Neural Network structure, there are several levels at which Neural Network processing can be divided into concurrently executable components. The following are the

some of the ways of parallelism that can be implemented in Neural Networks.

2.1.1 Exemplar Parallelism (EP)

This approach uses the existence of a large number of data examples as the source of parallelism. The work of Neural Networks training is reduced by distributing an equal size partition of the data set to each processor. Each processor trains an identical network on its local set of data examples.

2.1.2 Block parallelism

This approach partitions the network into blocks of adjacent neurons that are distributed among the processors.

2.1.3 Neuron Parallelism

For this approach, each individual neuron is treated as a concurrent process and is randomly distributed among the processors in a parallel machine.

The figures 1 to 3 illustrates each of the above Parallelism strategies [7].

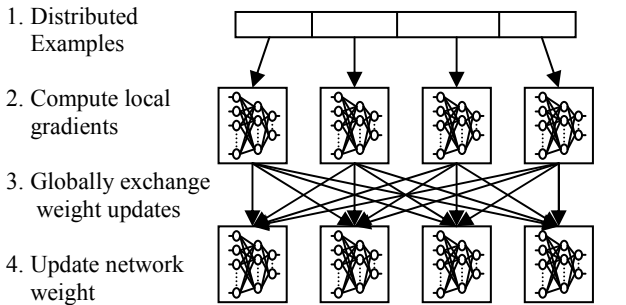


Figure 1: The Block diagram of Exemplar Parallelism

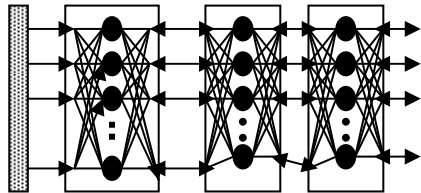


Figure 2: Block diagram of Block parallelism

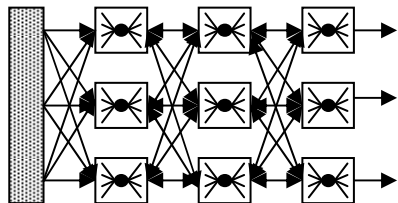


Figure 3: Block diagram of Neuron parallelism

There are two other levels of neural Parallelism that are in practice. The first one is training-session parallelism which entails the simultaneous training of independent Neural Networks on different processors. The other approach is weight parallelism in which the weights connected to every neuron in the network are distributed among several processors i.e, this

approach parallelizes the weighted sum computation for each neuron.

2.2 Advantages of Neural Networks for Classification

- Neural Networks are more robust than decision trees because of the weights
- The Neural Networks improves its performance by learning. This may continue even after the training set has been applied.
- The use of Neural Networks can be parallelized as specified above for better performance.
- There is a low error rate and thus a high degree of accuracy once the appropriate training has been performed.
- Neural Networks are more robust than decision trees in noisy environment.

2.3 Neural Network Models

There are three aspects involved in the construction of a Neural Networks.

1. Structure : The architecture and topology of Neural Networks.
2. Encoding : The method of changing weights (Training).
3. Recall : The method and capacity to retrieve information.

Various Neural Networks models exist and among these Feed Forward Neural Network is considered in this study for the construction of neural network. Because, this model, besides being popular and simple, is easy to implement and appropriate for classification applications.

2.3.1 Feed Forward Networks with Backpropagation

The feed forward backpropagation network is a very popular model in neural network. It does not have feedback connections, but the errors are back propagated during training. Backpropagation learning consists of two passes through the different layers of the network: a forward pass and backward pass.

In forward pass, input vector is applied to the sensory nodes of the network and its effect propagates through the network layer by layer. Finally, a set of outputs is produced as the actual response of the network. During the forward pass the synaptic weights of the network are all fixed.

During the backward pass, the synaptic weights are all adjusted in accordance with an error correction rule. The actual response of the network is subtracted from a desired (target) response to produce an error signal. This error signal is then backpropagated through the network, against the direction of synaptic connections[8].

Backpropagation algorithm can be improved by considering momentum and variable learning rate. Momentum allows a network to respond not only to the local gradient, but also to the recent trends in error surface. Acting like a low pass filter, momentum allows the network to ignore small features in the error surface. Without momentum, a network may get struck in a

shallow local minimum. In backpropagation with momentum, the weight change is in a direction that is a combination of the current and previous gradients. This is a modification of gradient descent whose advantages arise chiefly when some training data are very different from the majority of the data. Convergence is sometimes faster if a momentum term is added to the weight update formulas.

The performance of algorithm is very sensitive to the proper setting of the learning rate. If the learning rate is set too high, the algorithm may oscillate and become unstable. If the learning rate too small, the algorithm will take too long to converge. It is not practical to determine the optimal setting for the learning rate before training and in fact the optimal learning rate changes during the training process, as the algorithm moves across the performance surface. Performance of the backpropagation can be improved by allowing the learning rate to change during the training process. An adaptive learning rate will attempt to keep the learning step size as large as possible while keeping the learning process stable. The learning rate is made responsive to the complexity of the local error surface.

3. NEURAL NETWORKS IN MEDICAL FIELD

Keeping in view of the significant characteristics of NN and its advantages for the implementation of the classification problem, Neural Network technique is considered for the classification of data related to medical field in this study.

Owing to their wide range of applicability and their ability to learn complex and non linear relationships including noisy or less precise information Neural Networks are very well suited to solve problems in biomedical engineering. By their nature, Neural Networks are capable of high-speed parallel signal processing in real time. They have an advantage over conventional technologies because they can solve problems that are too complex-problems that do not have an algorithmic solution or for which an algorithmic solution is too complex. Neural Networks are trained by examples instead of rules and are automated. This is one of the major advantages of neural networks over traditional expert systems [9,10]. When NN is used in medical diagnosis they are not affected by factors such as human fatigue, emotional states and habituation. They are capable of rapid identification, analyses of conditions, and diagnosis in real time. With the spread of Neural Networks in almost all fields of science and engineering, it has found extensive application in biomedical engineering field also. The applications of neural networks in biomedical computing are numerous. Various applications of ANN techniques in medical field like medical expert system, cardiology, neurology, rheumatology, mammography and pulmonology were studied [11,12]. In this study medical data related to Breast Cancer is considered for classification purpose to identify the disease.

As the Neural Networks are inherently parallel in nature, this technique is considered in this study to implement parallelism for calculating the output at each node in different layers of the network. The basic unit of modularity in a network is neuron.

Data Representation Scheme:

The original data is present in the form of analog values with values ranging from 0-10. The given data sets are

Every neuron operates independently, processing the input receives, adjusting weights, and propagating its computed output thus a neuron is a natural level of parallelization for neural networks. Every neuron is treated as a parallel process.

For example a layer other than the input layer consists of m neurons and assume that processing time ' t units' to calculate the output at each neuron is similar. If the parallel concept is not adopted in neural network ' mt units' of time is needed to calculate the output. The needed time can be reduced by m times, if parallel concept is implemented at neuron level. If the network consists of many hidden layers, the processing time can be reduced at each layer in the network and thus the overall training time of the network can be reduced drastically. Hence, we adopted the above said parallel concept in this thesis to speedup the training process of the net to perform the classification task.

3.1 Experiment - Classification of Cancer Dataset

One of the leading causes of death of women is breast cancer. Mammography has been proved to be an effective diagnostic procedure for early detection of breast cancer. An important sign in its detection is the identification of micro calcification of mammograms, especially when they form clusters. In this experiment the medical data related to breast cancer is considered. This database was obtained from the university of Wisconsin hospital, Madison from Dr. William H. Wolberg. This is publicly available dataset in the Internet.

Descriptions of Database:

- Number of instances 699
- Number of attributes: 10 plus the class attribute
- Attributes 2 through 10 will be used to represent instances
- Each instance has one of 2 possible classes: benign or malignant
- Class distribution: Benign : 458 (65.5%)
Malignant : 241 (34.5%)

Attribute information:

Attribute	Domain
1. Sample code number	id number
2. Clump thickness	1-10
3. Uniformity of cell size	1-10
4. Uniformity of cell shape	1-10
5. Marginal adhesion	1-10
6. Single epithelial cell size	1-10
7. Bare nuclei	1-10
8. Bland chromatin	1-10
9. Normal nucleoli	1-10
10. Mitosis	1-10
11. Class	(2 for benign, 4 for malignant)

converted to their equivalent digital form. Scaling has the advantage of mapping the desired range of variables ranging between minimum and maximum range of network input. The

conversion of the given data sets into binary is done based on certain ranges, which are defined for each attribute. There are totally 10 attributes (1 class and 9 numeric features). The 9 numerical attributes are in the analog form scaled in the range between 0 and 1. First from the given range of inputs, the minimum and maximum value is picked up and this scaling is done by the following formula.

New value (after scaling) = (current value – Min value) / (Max - Min)

The new values obtained after truncating are converted into binary from by the following scaling. The values, which are in the range 0 to 5 are converted to 0 and 6 to 10 are converted to 1.

3.2 Training the Neural Network

In this experiment the neural network is trained with Breast Cancer database by using feed forward neural network model and backpropagation learning algorithm with momentum and variable learning rate. The cancer database consists 9 attributes. The input layer of the network consists of 9 neurons to represent each attribute as the cancer database consists of 9 attributes. The number of classes are 2, one Benign and another is Malignant.

So one neuron in the output layer is sufficient to represent these two classes. The description of the backpropagation algorithm is specified in the above is used to train the neural network during the training process. Several neural networks are constructed with and without hidden layers i.e, single and multi layer networks and trained with cancer dataset. Relationship between the number of epochs and the sum of squares of errors during training process for various networks can be observed from the Figures 4 and 5.

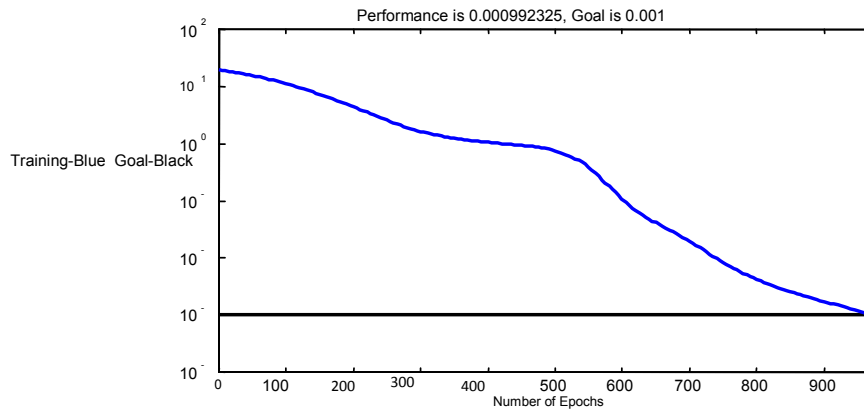


Figure 4: Training the Single Layer Network with Cancer Dataset

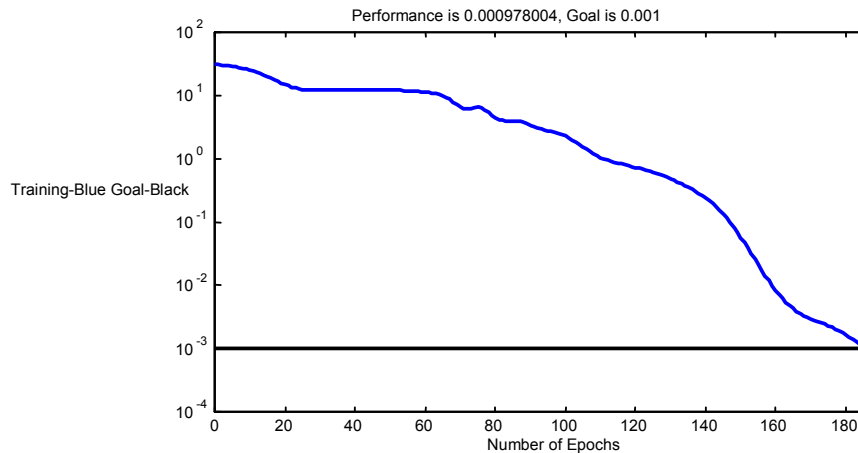


Figure 5: Training the Multi Layer Network with Cancer Dataset

3.3 Performance of the Network

The various phases in the classification problems solved by neural network techniques are construction, training and testing. Construction and training of the neural network are explained in the previous section. The classification of the test data and the performance of the network are discussed in this section. Various samples are collected as test data. The test data is given as the input to the trained network and the output of the net is calculated with the adjusted weights. Since we know the target output, the output of net is compared with this target output to study the learning ability of the network for classifying the cancer data. We observed that 92% test data are correctly classified and 8% are misclassified may be because of the analog conversion to digital conversion of dataset.

Table 1: Experimental Results of Cancer Dataset:

Training Samples	Test Samples	Classification Efficiency	
		Single Layer	Multi Layer
50	300	75%	80%
100	200	78.6%	82%
200	100	84.2%	89.4%
300	50	88.9%	92%

4. CONCLUSION

To classify the medical data set a neural network approach is adopted. Neural Networks are inherently parallel in nature. This technique is adopted to implement parallelism to calculate the output at each node in different layers for the classification of medical dataset such as Breast Cancer. The experiment is conducted with this dataset by considering the single and multi layer neural network models. Backpropagation algorithm with momentum and variable learning rate is used to train the networks. To analyze performance of the network various test data are given as input to the network. To speed up the learning process, parallelism is implemented at each neuron in all hidden and output layers. The results show that the multilayer neural network is trained quickly than single layer neural network and the classification efficiency is also high. The experimental results proved that neural networks technique provides satisfactory results for the classification task.

5. REFERENCES

- [1] A.A Freitas & S.H. Lavington. Mining Very Large Databases with Parallel Processing. Kulwer Academic Publishers, 1998. ISBN 0-7923-8048-7.
- [2] R. J. Bayardo. Efficiently mining long patterns from databases. In ACM SIGMOD Conf. Management of Data, June 1998.
- [3] John Shafer, Rakesh Agarwal, and Manish Mehta. SPRINT: A scalable parallel classifier for data mining. In Proc. Of the VLDB Conference, Bombay, India, Sep 1996.
- [4] Sunghwan Sohn and Cihan H. Dagli. Ensemble of Evolving Neural Networks in classification. Neural Processing Letters 19: 191-203, Kulwer Publishers, 2004
- [5] Sushmita Mitra. Datamining in Soft Computing Framework: A Survey. IEEE Transactions on Neural Networks, Vol 13, No. 1, Jan 2002.
- [6] R. Rojas. Neural Networks: a systematic introduction. Springer-Verlag, 1996
- [7] R. Owen Rigters. A framework for parallel data mining using neural networks. Technical report , Queen's University, Canada, 1997.
- [8] Simon Haykin. Neural Networks – A Comprehensive Foundation. Pearson Education, 2001.
- [9] K. Anil Jain, Jianchang Mao and K.M. Mohiuddin. Artificial Neural Networks: A Tutorial. IEEE Computers, 1996, pp.31-44.
- [10] George Cybenko. Neural Networks in Computational Science and Engineering. IEEE Computational Science and Engineering, 1996, pp.36-42.
- [11] Dr. A. Kandaswamy, Applications of Artificial Neural Networks in Bio Medical Engineering. The Institute of Electronics and Telecommunicatio Engineers, Proceedings of the Zonal Seminar on Neural Networks, Nov 20-21, 1997.
- [12] A. Kusiak, K.H. Kernstine, J.A. Kern, K>A. McLaughlin and T.L. Tseng, Data mining: Medical and Engineering Case Studies, Proceedings of the Industrial Engineering Research 2000 Conference, Cleveland, Ohio, May21-23, pp.1-7,2000