Scholedge
the scholarly publishing
SCHOLEDGE INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY & ALLIED STUDIES
VOL. 1, ISSUE 2(NOVEMBER 2014) ISSN-2394-336X

DATA MINING TECHNIQUE WITH CROSSBREEDING NEURAL NETWORK: LVQ AND HOPFIELD

Amitab Maurya[1],Navnish Goel[2], Tarun Kumar[3], Sachin Goel[4], Rohit Rana[5]

[1,3,4,5]Assistant Professor, Vidya College of Engineering, Meerut, UP India.
[2]Assistant Professor, SD Engineering and Technology Muzaffarnagar, UP India.

ABSTRACT

Neural network is a collection of "neurons" which send "signals" to produce an output. Human brain contains approximately 100 billion neurons. One neuron may be connected with more than 10,000 other neurons. The main part of our body is Brain which sends the signals in form of the neurons and again achieves the output. In this paper we want to try a combined approach with Linear Vector Quantization and Hopfield neural network to implement the new data mining techniques. There are so many data mining techniques which is used by different types of neural network such as multilevel neural network, Back propagation neural network.

Keyword: LVQ, Hopfield NN, Neurons, feed forward, Back Propagation, CFBP and MATLAB.

Introduction

The intelligence of a neural network emerges from the collective behavior of
neurons, each of which performs only limited operation. Even though each individual neuron works slowly, they can still quickly find a solution by working in parallel. This fact can explain why humans can recognize a visual scene faster than a digital computer, while an individual brain cell responds much more slowly than a digital cell in a VLSI circuit. Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the connections between elements largely determine the network function. You can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements. The strength of the interconnections between neurons is implemented by means of the synaptic weights used to store the knowledge.

Typically, neural networks are adjusted, or trained, so that a particular input leads to a specific target output. The figure 1, illustrates such a situation. There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically, many such input/target pairs are needed to train a network.

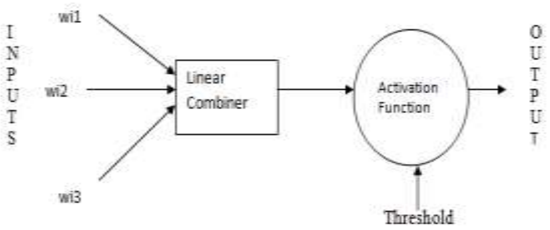

Fig 1: Basic neural network

According to the interconnection scheme, a network can be either feed-forward or recurrent and its connection either symmetrical or asymmetrical. Their definitions are given below.

- **Feed-forward networks:** All connections point in one direction (from the input toward the output layer). In fig 2. It has the following characteristics:

1. Perceptions are arranged in layers, with the first layer taking in inputs and the last layer producing outputs. The middle layers have no connection with the external world, and hence are called hidden layers.
2. Each perception in one layer is connected to every perceptron on the next layer. Hence information is constantly "fed forward" from one layer to the next, and this explains why these networks are called feed-forward networks.
3. There is no connection among perceptions in the same layer.

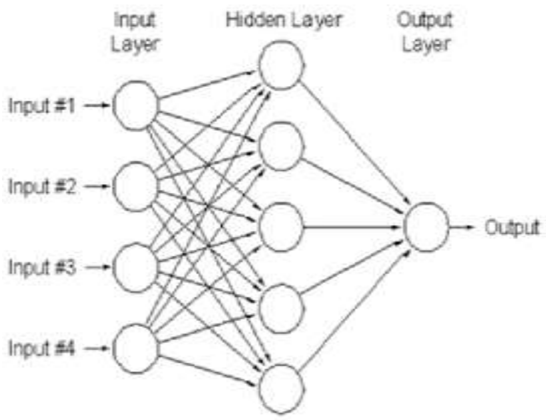

Fig 2 Feed-forward network

- **Fully Recurrent Networks:** All units are fully connected to all other units and every unit is both an input and an output. Some connections are present from a layer to the previous layer, there is no hierarchical arrangement and the connections can be bi-directional. Recurrent networks are also useful in that they allow to process sequential information. Processing in recurrent network depends on the state of the network at the last step.

In these networks there are feedback loops are present. These networks can learn from their mistakes and are of highly adaptive in nature. These kinds of networks train slowly and work well with noisy inputs.

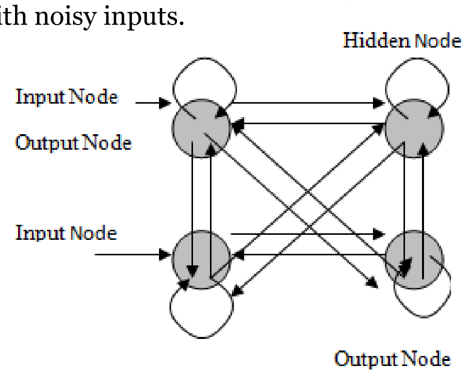

Fig 3: Fully Connected Asymmetric Network

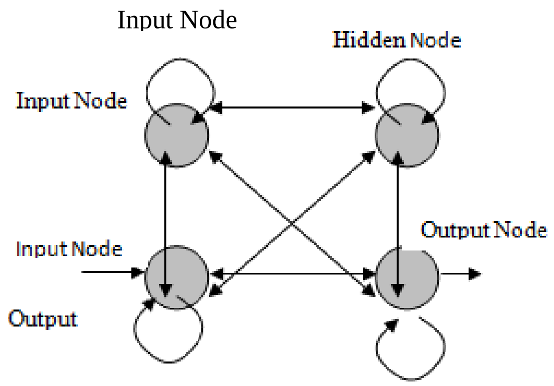

Fig 4: Fully Connected Symmetric Network

- **Symmetrical connections:** If there is a connection pointing from node i to node j, then there is also a connection from node j to node i, and the weight associated with the two connection are equal, or notation ally, Wji= W_{ij}. Fig 4.
- **Asymmetrical connection:**the connection from one node to another may carry a different weight than the connection from the second node to the first.Connection weight can be real number or integers. Fixed by design. When training is completed, all of them should be fixed. Fig 3.

What is Data Mining?

Data mining, or knowledge discovery, is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining tools predict behaviors and future trends, allowing businesses to make proactive, knowledge-driven decisions. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Data mining derives its name from the similarities between searching for valuable information in a large database and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find where the value resides.

What Can Data Mining Do?

Although data mining is still in its infancy, companies in a wide range of industries - including retail, finance, healthcare, manufacturing transportation, and aerospace - are already using data mining tools and techniques to take advantage of historical data. By using pattern recognition technologies and statistical and mathematical techniques to sift through warehoused information, data mining helps analysts recognize significant facts, relationships, trends, pattern,exceptions and anomalies that might otherwise go unnoticed.

Learning Vector Quantization

Learning Vector Quantization (LVQ) is a supervised version of vector quantization, similar to Selforganising Maps (SOM) based on work of LINDE et al., GRAY and KOHONEN (for a comprehensive overview). It can be applied to pattern recognition, multi-class classification and data compression tasks, e.g. speech recognition, image processing or customer classification. As supervised method, LVQ uses known target output classifications for each input pattern of the form. LVQ algorithms do not approximate density functions of class samples like Vector Quantization or Probabilistic Neural Networks do, but directly define class boundaries based on prototypes, a nearest-neighbor rule and a winner-takes-it-all paradigm. The main idea is to cover the input space of samples with 'codebook vectors' (CVs), each representing a region labeled with a class. A CV can be seen as a prototype of a class member, localized in the centre of a class or decision region ('Voronoï cell') in the input space. As a result, the space is partitioned by a 'Voronoï net' of hyper planes perpendicular to the linking line of two CVs (mid-planes of the lines forming the 'Delaunay net'; see Fig. 4). A class can be represented by an arbitrarily number of CVs, but one CV represents one class only.

Hopfield Neural network:

One of the milestones for the current renaissance in the field of neural networks was the associative model proposed by Hopfield at the beginning of the 1980s. Hopfield's approach illustrates the way theoretical physicists like to think about ensembles of computing units. No synchronization is required, each unit behaving as a kind of elementary system in complex interaction with the rest of the ensemble. An energy function must be introduced to harness the

Theoretical complexities posed by such an approach. The next two sections deal with the structure of Hopfield networks. We then proceed to show that the model converges to a stable state and that two kinds of learning rules can be used to find appropriate network weights

Literature Review

Itedal Sabri Hashim Bahia (2013) demonstrate the importance and possible value of housing predictive power which provides independent real estate market forecasts on home prices by using data mining tasks. A (FFBP) network model and (CFBP) network model are one of these tasks used in this research to compare results of them. We estimate the median value of owner occupied homes in Boston suburbs given 13 neighborhood attributes. An estimator can be found by fitting the inputs and targets. This data set has 506 samples. "Ousting inputs" is a 13 × 506 matrix. The "housing targets" is a 1 × 506 matrix of median values of owner-occupied homes in $1000's. The result in this paper concludes that which one of the two networks appears to be a better indicator of the output data to target data network structure than maximizing predict. The CFBP network which is the best result from the Output network for all samples are found from the equation output = 0.95 * Target + 1.2. The regression value is approximately 1, (R = 0.964). That means the Output network is matching to the target data set (Median value of owner-occupied homes in $1000's), and the percent correctly predict in the simulation sample is 96% [1]

Ghosh, S. Proposed (2011) weather Data Mining is a form of Data mining concerned with finding hidden patterns inside largely available meteorological data, so that the information retrieved can be transformed into usable knowledge. A variety of data mining tools and techniques are available in the industry, but they have been used in a very limited way for meteorological data. In this paper, a neural network-based algorithm for predicting the atmosphere for a future time and a given location is presented. We have used Back Propagation Neural (BPN) Network for initial modeling. The results obtained by BPN model are fed to a Hopfield Network. The performance of our proposed ANN-based method (BPN and Hopfield Network based combined approach) tested on 3 years weather data set comprising 15000 records containing attributes like temperature, humidity and wind speed. The prediction error is found to be very less and the learning converges very sharply. The main focus of this paper is based on predictive data mining by which we can extract interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of meteorological data.

Objectives of the Study

The objective of this synopsis is to present capability of MATLAB as a data mining tool when neural network will be used for training the data with different training function with the help of nntool in MATLAB. Data mining is recognized as an important field where one has the possibility to become accustomed both with analysis techniques and methods and with a state of mind. By means of data mining it is possible to develop critical skills that are essential in today's information technology. I want to develop a MATLAB based technique by crossbreeding the two neural networks (LVQ + Hopfield Neural Network) and try to optimize the data which is handled by data mining system.

Research Methodology

Hence, we aim to provide, not only an analysis of selected data mining tools available within MATLAB and a synthesis of these tools, but more importantly, a means to analyze and synthesize further data mining tools, thus providing an increasingly holistic view of the data mining capabilities of MATLAB. Essentially then, we wish to discover the extent to which each of a number of MATLAB data mining tools is capable of carrying out the different stages of the data mining process. We wish to synthesize these tools in order to bring greater clarity to the potential of MATLAB in the data mining arena and to provide recommendations for further extension to these tools in light of this analysis and synthesis. And, as we do this, to clearly define the methodology used in carrying out this work, in order that it might be used in future work in this area.

In summary, our aim is to create a means for obtaining a holistic view of the data mining capabilities of MATLAB. We will accomplish this by setting forth the methodology of this process and by demonstrating this methodology by investigating and synthesizing several data mining tools available for MATLAB.

This work ensures that data mining in MATLAB becomes an increasingly straightforward task, as the appropriate tools for a given analysis become apparent. As a logical extension of the synthesis provided, recommendations are given with regard the creation of a data mining toolbox for MATLAB. The opportunities for extension to this work are numerous, not only in terms of extending the tools themselves but also of data mining in MATLAB as a whole. The data mining tools around which will revolve are: the Neural Network Toolbox, a proprietary tool available from The Math Works, distributors of MATLAB. The Fuzzy Clustering and Data Analysis Toolbox and the Association Rule Miner and Deduction Analysis tool, which are both open source; and lastly an implementation of the decision tree algorithm.

Conclusion

Our central aim in this work was to not only provide an analysis and synthesis of data mining tools in MATLAB, but also a methodology which can be used in continuing this work into the future and possibly extending it to the creation of a data mining toolbox. We have been successful in creating such a methodology and we have validated these findings by evaluating and synthesizing MATLAB data mining tools.

Reference

[1] Itedal Sabri Hashim Bahia, "A data mining model by using ANN for prediction estate market: comparative study", International Journal of Intelligence Science, 2013, 1, 162-169.

[2] Soumadip Ghosh, Sushanta Biswas, "Weather Data Mining using ANN", 978-1-4244-9477-41, 2011, IEEE.

[3] Han J., Kambar M., Morgan Kaufmann "data mining : Concept and Technique", 2000.

[4] Dr. S. SanthoshBaboo and I. Kadar Shereef. "An efficient Weather forcasting System using ANN", international journal of environmental science and development, vol 1, no 4, oct 2010.

[5] S.N. Sivanadan, S.N. deepa, "Introduction to Neural Networks Using Matlab 6.0", TMH, 5th editions,2010.

[6]ItedalSabriHashimBahia , "Using Artificial Neural Network Modeling in Forecasting Revenue: Case Study in National Insurance Company/Iraq", International Journal of Intelligence Science, 2013, 3, 136-143 doi:10.4236/ijis.2013.33015 Published Online July 2013.

[7] R. Rojas: Neural Networks, Springer-Verlag, Berlin, 1996.

[8]http://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/