

A REVIEW OF DATA MINING WITH AI

Shobana B & Dr.Savithri V

Abstract— This study predicts the best supervised learning method of knowledge discovery in databases patterns. Data mining algorithms are used to extract the information and patterns derived by the knowledge discovery database process. Classification are done using predictive model of data mining techniques of back propagation algorithm and Radial basis function. The values of the features are evaluated by artificial Intelligence algorithm.

Index Terms— Pattern recognition, Back propagation network, Data mining, Knowledge discovery in database, artificial neural network.

Abstract— This study predicts the best supervised learning method of knowledge discovery in databases patterns. Data mining algorithms are used to extract the information and patterns derived by the knowledge discovery database process. Classification are done using predictive model of data mining techniques of back propagation algorithm and Radial basis function. The values of the features are evaluated by artificial Intelligence algorithm.

Keywords— Pattern recognition, Back propagation network, Data mining, Knowledge discovery in database, artificial neural network.

1. INTRODUCTION

Amatur, et al., (1992), examined the segmentation of magnetic resonance images by optimizing neural networks. This study has demonstrated the applicability of Hopfield net for the tissue classification in MRI. Levinski, et al., (2009), describes the approach for correcting the segmentation errors in 3D modeling space, implementation, principles of the proposed 3D modeling space tool and illustrates its application. Paragios, et al., (2003), introduces a knowledge based constraints, able to change the topology, capture local deformations, surface to follow global shape consistency while preserving the ability to capture using implicit function. Suri, et al., (2002), an attempt to explore geometric methods, their implementation and integration of regularizers to improve robustness of independent propagating curves/surfaces. Yuksel, et al., (2006), reveals the 100% classification accuracy of carotid artery Doppler signals using complex-values artificial neural network. Wendelhag, et al., (1991, 1997) results shows variations secondary to subjective parameters when manual measurement methods are employed.

A thorough computerized system is necessary to evaluate the pattern recognition using data mining techniques. Our proposed method acts as a tool to predict the same patterns in data effectively and efficiently with less time and less memory allocation.

2. MATERIAL AND METHODS

In the present study, 200 samples of datas of medical images with same patterns are collected. Data mining techniques for classification using ANN, Back propagation algorithm(BPA) and Radial Basis Function(RBF) were examined to extract the selection of data, preprocessing the data, transformation of data into common format and data mining with ann techniques are used to generate desired results. The results of these two classifiers were compared. The same dataset features were compared with BPA and RBF This methodology provides a reliable tool to generate desired result to the users. The results demonstrate that it has the potential to perform qualitatively better than applying existing methods in pattern recognition using data mining techniques.

3. Data Mining

Knowledge discovery in database patterns refers to the collection of data from many different data sources. Erroneous data may be corrected or removed, whereas missing data must be supplied or predicted using SPSS tool. Data from different sources must be converted into common format. Applied data mining algorithm to the transformed data to generate desired results.

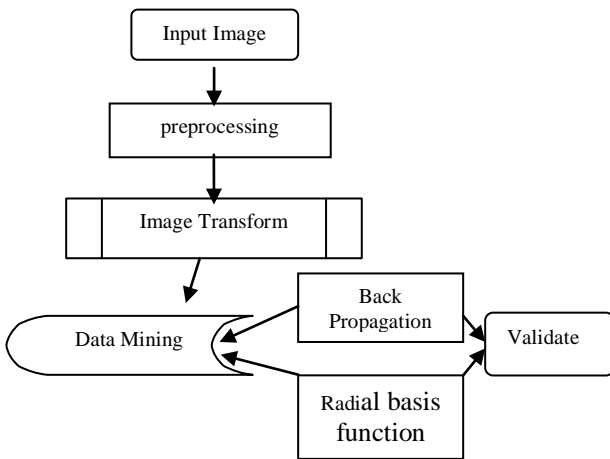


Fig.1 Overview of proposed method

3.1 Back propagation Network

A back propagation neural network is a multi-layer feed-forward neural network consisting of an input layer, a hidden layer and an output layer. The neurons presents in the hidden and output layers have biases, which are connections from the units whose activation is always one. The training of the back propagation network is done in three stages.

- 1) Input training pattern.
- 2) Error calculation.
- 3) Updation of weights.

3.1.1 BACK PROPAGATION ALGORITHM

The back propagation learning procedure has become the single most popular method to train networks. It has been used to train networks in problem domains. The algorithm is developed considering supervised method for measuring minimum error value and steepest-descent method to examine a global minimum.

BPA helps to classify a boundary based on nine characteristics of sample ultrasound segmented images. The dataset consist of 200 samples dataset input is an 9x200 matrix, whose characteristics are age, size, height, sex, length, internal diameter, wall thickness and circularity.

200 samples of dataset are randomly used for validation and testing. 70% for training,(140 samples), 15% for validation(30 samples) and 15% for testing(30 samples) are considered in BPN

Beginning with an initial (possibly random) weight assignment for a three layer feed forward network proceed as follows

Step 1: Present i^p and form outputs o_i of all units in network.

Step 2: Use equation $\Delta^p W_{ji} = \epsilon (t_j^p - o_i^p) f_j'$ $(net_j^p) O_i^p$ to update W_{ji} for the output layer.

Step 3: Use equation $\delta_k^p = f_k' (net_k^p)$ to update W_{ji} for the hidden layer(s).

Step 4: Stop if updates are insignificant or the error is below a preselected threshold; otherwise return to step 1.

3.2 RADIAL BASIS FUNCTION NETWORK

RBF is a function whose values increases or decreased with the distance from a central point. The Gaussian activation function is an RBF function with a central point of 0. Three layers of RBF NN are 1) The input layer used to simply input the data 2) In hidden layer, Gaussian activation function is applied. The hidden nodes learn to respond only to the subset of the input . 3) The output layer displays one predicted output as per rules processing. The dataset consist of 200 samples are given as input. Stop the process until the error is less than the threshold value.

4. RESULTS AND DISCUSSION

A brief study of two different methods is discussed. Features are extracted from segmented images using pattern recognition of neural network. Back propagation algorithm and Radial basis function network is applied to train and to classify the inputs according to the outputs. To validate the output confusion matrix is implemented for training, testing and validation done to refer true positive rate versus false positive rate.

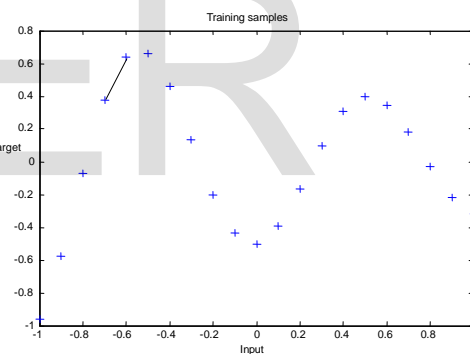


Fig.2 Training in BPA

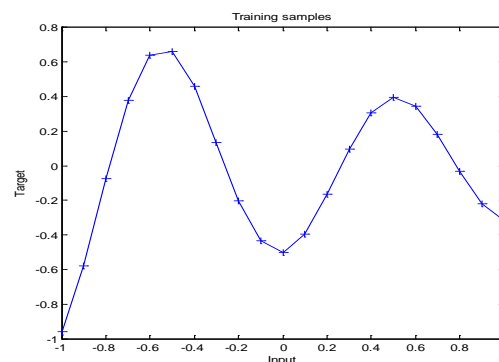


Fig.3 Training in RBF

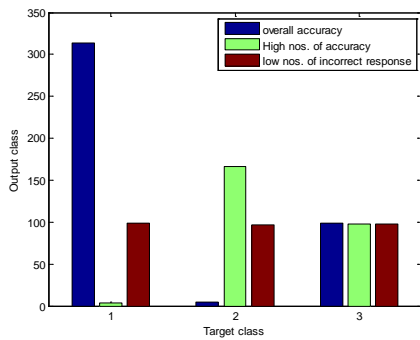


Fig. 4 High accuracy in training confusion matrix

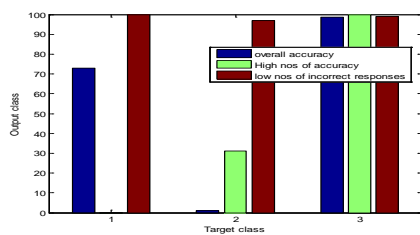


Fig. 5 High accuracy in validation confusion matrix

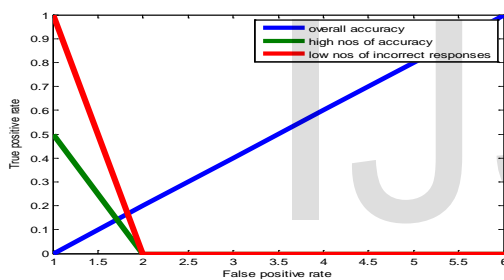


Fig. 6 ROC for training true positive rate vs false positive rate

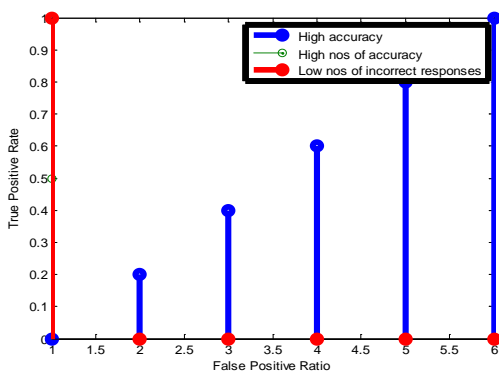


Fig. 7 ROC for training true positive ratio vs false positive ratio

To validate the desired output training, testing and validation is done by using confusion matrix network. Confusion matrix is applied for training, validation, testing ROC (Region of characters) using true positive rate vs false positive rate. Figure 3.3 and Figure 3.4, shows the 99.51%

accuracy and no number of incorrect responses helps to predict the Back propagation algorithm is best compared to radial basis function network.

5. CONCLUSION

A real time measurement made and classified by back propagation network which produces more accurate results than Radial Basis function method.

It is believed that this will provide a faster solution and effective way for classification of data patterns in data mining.

Thus concluded that prediction of knowledge based database patterns can be detected effectively and accurately with back propagation algorithm and the same can be used as the second observer apart from practitioner's opinion.

In future, the proposed method is also suitable for medical applications to detect closer contours. This method can also be suitable to detecting vessel and spine boundary in angiography and radiography.

References

- [1] Amatur, S.C., Piraino, D., and Takefuji, Y., 1992, Optimization neural networks for the segmentation of magnetic resonance images, IEEE Transactions in Medical imaging, Vol.II, Issue 2, pp.215-220.
- [2] Bing Nan Li, Chee Kong Chuti, Stephen Chang, and Ong, S.H., 2011, Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation, Computers in Biology and Medicine, Vol.41, Issue 1, pp.1-10.
- [3] Cai, W., Chen, S., and Zhang, D., 2007, Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation, Pattern Recognition, Vol.40, pp.825-838.
- [4] Da-Chuan Cehng, Christian Billich, Shing-Hong Liu, Horst Brunner, Yi-Chen Qiu, Yu-Lin Shen, Hans J Brambs, Arno Schmidt-Trucksass and Uwe HW Schutz, 2011, Automatic detection of the carotid artery boundary on cross-sectional MR image sequences using a circle model guided dynamic- programming, Biomedical engineering Vol.10, pp.1-17.
- [5] Levinski, K., Sourin, A., and Zagorodnov, V., 2009, Interactive surface-guided segmentation of brain MRI data, Computers in Biology and Medicine, Vol.39, Issue 12, pp.1153-1160.
- [6] Markos, G., Tsipouras, Themis, P., Exarchos, Dimitrios, I., Fotiadis, Anna, P., Kotsia, Konstantinos, V., Vakalis, Katerina, K., Naka and Lampros K., Michalis, 2008, Automated Diagnosis of coronary artery disease based on data mining and fuzzy modeling, IEEE transactions on Information technology in biomedicine, Vol.12, Issue 4, pp.447-456.
- [7] Paragios, N., 2003, A level set approach for shape-driven segmentation and tracking of left ventricle, IEEE Transactions on Medical Imaging, Vol.22, pp.773-776.
- [8] Santhiyakumari, N., and Madheswaran, M., 2010, Intelligent medical decision system for identifying ultrasound carotid artery images with vascular disease, International journal of Computer Application, Vol.1, Issue 13, pp.32-39.
- [9] Suri, J.S., 2001, Two-dimensional fast magnetic resonance brain segmentation, IEEE Engineering in Medicine and Biology, Vol.20, pp.84-95.
- [10] Wendelhag, I., Gustavsson, T., Suurkula, M., Berglund, G., and Wikstrand, J., 1991, Ultrasound measurement of wall thickness in the carotid artery: fundamental principles and description of a computerized analysing system, Clin Physiol., Vol.11, pp.565-577.
- [11] Wendelhag, I., Liang, Q., Gustavsson, T., and Wikstrand, J., 1997, A new automated computerized analyzing system simplifies readings and

- reduces the variability in ultrasound measurement of intima-media thickness, *Stroke*, Vol.28, pp.2195-2200.
- [12] Yuksel Ozbay and Murat Ceylan, 2006, Effects of window types on classification of carotid artery Doppler signals in the early phase of atherosclerosis using complex-valued artificial neural network, *Ultrasound in Medicine and Biology*, Vol.37, Issue 3, pp.287-295.
- [13] Li, B.N., Chui, C.K., Ong, S.H., and Chang, 2009, Integrating FCM and level sets for liver tumor segmentation, *Proceedings of the 13th International Conference on Biomedical Engineering, (ICBME 2008), IFMBE Proceedings 23*, pp.202–205.
- [14] Lei, W.K., Li, B.N., Dong, M.C., and Vai, M.I., 2007, AFC-ECG: an adaptive fuzzy ECG classifier, in: *Proceedings of the 11th World Congress on Soft Computing in Industrial Applications (WSC11), Advances in Soft Computing*, Vol.39, pp.189–199.

IJSER