Egyptian Informatics Journal (2014) **15**, 129–147

Cairo University

Egyptian Informatics Journal

www.elsevier.com/locate/eij
www.sciencedirect.com

ORIGINAL ARTICLE

A novel Neuro-fuzzy classification technique for data mining

Soumadip Ghosh [a,*], **Sushanta Biswas** [b], **Debasree Sarkar** [b], **Partha Pratim Sarkar** [b]

[a] *Academy of Technology, Aedconagar, Hooghly, 712121 WB, India*
[b] *Department of Engineering and Technological Studies (DETS), University of Kalyani, 741235 WB, India*

Received 14 February 2014; revised 15 July 2014; accepted 25 August 2014
Available online 17 September 2014

KEYWORDS

Classification;
Data mining;
Neural network;
Neuro-fuzzy

Abstract In our study, we proposed a novel Neuro-fuzzy classification technique for data mining. The inputs to the Neuro-fuzzy classification system were fuzzified by applying generalized bell-shaped membership function. The proposed method utilized a fuzzification matrix in which the input patterns were associated with a degree of membership to different classes. Based on the value of degree of membership a pattern would be attributed to a specific category or class. We applied our method to ten benchmark data sets from the UCI machine learning repository for classification. Our objective was to analyze the proposed method and, therefore compare its performance with two powerful supervised classification algorithms Radial Basis Function Neural Network (RBFNN) and Adaptive Neuro-fuzzy Inference System (ANFIS). We assessed the performance of these classification methods in terms of different performance measures such as accuracy, root-mean-square error, kappa statistic, true positive rate, false positive rate, precision, recall, and f-measure. In every aspect the proposed method proved to be superior to RBFNN and ANFIS algorithms.

1. Introduction

Data mining has attracted many researchers and analysts in the information industry and in research organizations as a whole in the last decades, due to the availability of large amounts of data and the immediate need for transforming such data into meaningful information and knowledge. The useful knowledge gathered can be applied in many areas such as market survey, customer retention, production control, evolutionary analysis and science exploration [1,2].

Classification is an important data mining technique which involves extracting interesting patterns representing knowledge

* Corresponding author.
E-mail addresses: soumadip@rediffmail.com (S. Ghosh), biswas.su@gmail.com (S. Biswas), dsarkar70@gmail.com (D. Sarkar), parthabe91@yahoo.co.in (P.P. Sarkar).
Peer review under responsibility of Faculty of Computers and Information, Cairo University.

Production and hosting by Elsevier

1110-8665 © 2014 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University.
Open access under CC BY-NC-ND license. http://dx.doi.org/10.1016/j.eij.2014.08.001

from large real-world databases. Such analysis can provide a deep insight into the better understanding of different large-scale databases. The study related to effective knowledge development is also very popular in any research as because the decision-making process mainly depends upon the effectiveness of the classification method being utilized.

Basically data classification [3,4] is the method of discovering a model or classifier that describes and differentiates data classes so that the model could predict the class of entities with unknown class label value. It is a two-step procedure, in the first step; a classifier is constructed denoting a predefined set of concepts or data classes. This is the training phase, where a classification algorithm constructs the classifier by learning from a training data set and their associated class label attributes. In the next step the model is used for classification. In order to estimate the performance of the classifier a test set independent of the training tuples is used. Several preprocessing steps such as data cleaning, data selection and data transformation are also applied to the data set before the classification procedure takes place.

Artificial neural network (ANN) or simply neural network (NN) [5–7] is a popular data modeling tool that can perform intelligent tasks similar to the human brain. NN is well-known for high precision and high learning ability even when a very little information is available. One of the reliable methods of data classification from the neural network domain is the Multilayer Perceptron Backpropagation network (MLPBPN) [8,9] algorithm. The Radial Basis Function Neural Network (RBFNN) [10,11] is another powerful neural network model that utilizes radial basis functions as the activation functions. The output of such a neural network model is the linear combination of radial basis functions of inputs and neuron parameters. RBFNNs have many utilities, comprising classification, function approximation, system control, and prediction of time series.

Due to the presence of imprecise input information, ambiguity or vagueness in input data, overlapping boundaries among classes, and indefiniteness in defining features some uncertainties can still arise at any stage of a data classification system. The fuzzy set theory [12–14] as a generalization of the classical set theory is very flexible in handling different aspects of uncertainties or incompleteness about real life situations. In a fuzzy system the features are associated with a degree of membership to different classes. Both NNs and fuzzy systems are very adaptable in estimating the input–output relationships. Neural networks deal with numeric and quantitative data while fuzzy systems can handle symbolic and qualitative data. Neuro-fuzzy hybridization leads to a crossbreed intelligent system widely known as Neuro-fuzzy system (NFS) [15,16] that exploits the best qualities of these two approaches efficiently. The hybrid system unites the human alike logical reasoning of fuzzy systems with the learning and connectedness structure of neural networks by means of fuzzy set theory based approach.

There is another Neuro-fuzzy classification based model which comprises a set of interpretable IF-THEN rules. They consider two conflicting requirements in fuzzy modeling: interpretability against accuracy. In reality, one of the two properties persists. Therefore the rule based Neuro-fuzzy modeling research area is divided into two branches: the linguistic fuzzy modeling that focuses on interpretability, primarily the Mamdani model; and the exact fuzzy modeling that focuses on accuracy, mainly the Sugeno model or Takagi–Sugeno–Kang (TSK) model. The rule based Neuro-fuzzy classification approach normally applies the concept of adaptive neural network. An adaptive network is a network of nodes (processing elements) and directed links (weights) that is functionally equivalent to a Fuzzy Inference System and is referred to as Adaptive Neuro-fuzzy Inference System or ANFIS [15]. It normally employs the Sugeno fuzzy model to produce IF-THEN learning rules. The nodes of an adaptive network are associated with certain parameters which might have an impact on the final output. ANFIS generally utilizes a hybrid learning algorithm which is the combination of gradient descent and least square method to adapt the parameters in adaptive network. To put it simply, ANFIS is the combination of MLPBPN and Sugeno fuzzy model. A fuzzy rule in the Sugeno model has the following form

$$\text{IF } x \text{ is } P \text{ and } y \text{ is } Q \text{ THEN } z = f(x, y) \tag{1}$$

where P and Q are the fuzzy sets in the antecedent part of the given IF-THEN learning rule and $z = f(x,\ y)$ is a crisp function in the consequent part of the rule.

In our present study, we proposed a novel Neuro-fuzzy classification technique for data mining that uses a combination of MLPBPN and fuzzy set theory approach. We applied our method to ten benchmark data sets from the UCI machine learning repository for analysis and, therefore compare its performance with RBFNN and ANFIS based classification models.

This research study is arranged as follows: Section 2 includes the related works done in this area. Section 3 describes our proposed Neuro-fuzzy classification method while Section 4 explains the detailed procedure. Section 5 discusses the performance analysis and results; and Section 6 is reserved for the conclusion.

2. Related works

Neuro-fuzzy classification is a field of research that has caused a great deal of attention in the recent decades. In the Neuro-fuzzy paradigm, several attempts have been made [17–20] which led to a strong foundation for research. In that context we have presented some of the works done in this area by other researchers.

A study performed by Kuncheva [21] described how to utilize fuzzy pattern recognition concept in solving real life problems. According to her work, fuzzy pattern recognition can be related to fuzzy clustering or with fuzzy IF-THEN systems used as classifiers and it is close to any pattern classification paradigm that involves fuzzy sets. She also indicated that fuzzy systems combined with neural networks should exploit the merits of these two approaches efficiently.

Pedrycz [22] specified that artificial neural network model combined with fuzzy set theory based approaches should possess the merits of both and it should permit one to arrive at a more knowledgeable decision making systems. The research study established that Neuro-fuzzy hybridization leads to a crossbreed intelligent system that can handle real life situations reasonably well.

A research work performed by Castellano et al. [23] identified that the use of a Neuro-fuzzy system and an evolutionary fuzzy system hybridizes the approximate reasoning mechanism