Volume 7, No. 7, Nov-Dec 2016
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

# Extraction of Classification Rules from Enhanced Fuzzy Min-Max Neural Network

Jaishri M. Waghmare
Assistant Professor,
Department of Computer Science and Engineering, SGGS Institute of Engineering and Technology, Nanded, India

Dr. U. V. Kulkarni
Professor,
Department of Computer Science and Engineering, SGGS Institute of Engineering and Technology, Nanded, India

Abstract—Though neural networks have capability in solving complicated problems, their deficiency of being ‘black box models’ has prevented them from being acknowledged as a typical practice for applications such as real-time applications, restorative medical and business for predictions. Enhanced fuzzy min max neural network (EFMN) has a number of enhancements to the original fuzzy min max neural network (FMN). Hence it is more accurate than original FMN and other related classification algorithms. Like other artificial neural networks (ANNs), EFMN is also like a black box and the knowledge is expressed in terms of min-max values of the hyperboxes and associated class labels. So the justification of classification results given by EFMN is required to be obtained to make it more adaptive to the real world applications. This paper proposes a model to extract classification rules in the form of if then from trained EFMN using partial decision trees (PART) algorithm. These rules justify the classification decision given by EFMN without loss of the accuracy. For this, EFMN is trained for the appropriate value of θ. The min-max values of all the hyperboxes and their respective class labels are given as input to PART algorithm which extracts rules by repeatedly generating partial decision trees from the input instances. These rules are readable and represents the trained network. These rules give accurate justification of the classification decision. The applicability of the proposed method is tested on widely used Fisher’s Iris dataset.

Index Terms—Rule extraction, fuzzy min max neural net- work(s) and partial decision trees

## I INTRODUCTION

L. A. Zadeh has proposed fuzzy sets in [11]. By definition of crisp classic sets, an element is either its member or not. On the other hand, fuzzy sets allow a member to be partially in the set as their boundaries are vague and ambiguous. Hence they are more compatible for genuine real-world systems. Hence many machine learning and pattern recognition problems use them as solution. The concept of ANN has been inspired by biological neural networks. ANN consists of large number of processing units called nodes or neurons. They are capable of doing parallel processing. Nodes are interconnected by connecting links and each of the connection links has weight which is modifiable. ANNs learn from the inputs provided [2]. Neurofuzzy frameworks are made by consolidating fuzzy logic and ANNs. Computational efficiency of ANNs and complex class boundaries representation capability of fuzzy logic make FMN an uncommon type of a neurofuzzy framework. The efficiency of neurofuzzy systems is high as compared to the other machine learning techniques. FMN network involves a dynamic network structure. It can learn new patterns online. Hence as like ANNs, it does not need retraining. As per [9] there are several notable properties associated with FMN learn- ing, which are as follows: online adaptation that is incremental learning, class boundary can be nonlinearly separable, mini- mum misclassification by separation of overlapping classes, short training time, capability of soft and hard decisions, mechanism for verification and validation of performance and number of configuration constant parameters should be as few as possible.

In spite of the notable properties of FMN, there are cer- tain shortcomings pertaining to the FMN learning algorithm. Specifically, the present hyperbox expansion criterion can affect FMN adversely. It can lead to overlapping between different hyperboxes of different classes, which could be avoided. Other than that, the present hyperbox overlap check rule cannot find all of the overlapping regions, and hence the contraction step is not performed for those overlapping hyperboxes. In order to overcome the mentioned shortcomings of FMN, [6] proposed three heuristic rules.

The drawback of all the FMNs is that they need to have explanation capability. The justification of the classification results given by EFMN is not readily available. Hence they are considered as black boxes[1]. So for safety critical prognostic and diagnostic tasks, domain specialists many times hesitate to utilize these models as solution [7]. It encourages the need of classification rules extraction from EFMN that provides the explanation of the classification prediction. [5] and [8] are methods for rule extraction that extracts fuzzy rules in form of if then from trained FMNs. In order to reduce the number of rule, for each of the fuzzy set hyperbox, a confidence factor is calculated and a threshold value which is user defined is used to remove the hyperboxes which has low confidence factors and then rules are extracted from the model. [7] proposed a two-stage system. Modified FMN based pattern classification is the first stage and genetic algorithm based rule extraction is the second stage. They have extracted fuzzy if-then rules from the modified FMN classification model. The rules are optimized using genetic algorithm with don't care approach.

In this paper, EFMN is used to train input patterns with different values of θ and for each value of θ classification accuracy is checked. The knowledge acquired in the EFMN, for highest classification accuracy value of θ, is used as input to PART algorithm. PART constructs partial decision tree in order to extract rule. Such extracted rules clarify the classification prediction of EFMN for input test data. The proposed model is able to extract high quality rules which measured in terms rule accuracy, rule fidelity, rule consistency, rule comprehensibility as described in [10].

The paper is organized as follows. In section II, subsection A presents description of FMN, subsection B gives details of EFMN and subsection C describes rule extraction using PART algorithm. Section III describes the proposed approach. Experimental setup and results are given in Section IV. Finally, conclusion and future directions are presented in Section V.

## II. BACKGROUND

### A. Fuzzy min-max neural network

FMN is a machine learning technique that was introduced in 1992 by Simpsons[9]. This method uses hyperboxes to represent patterns in the pattern space. A hyperbox is represented using min and max points. Each hyperbox belongs to a class. The membership functions of the fuzzy set hyperbox measure the degree to which the input pattern falls outside of the hyperbox. For the pattern which falls on or inside the hyperbox, its membership value is one. The learning algorithm of FMN allows for overlapping of hyperboxes of the same class. But if overlap occurs between hyperboxes of different classes then it is eliminated. These hyperboxes are created and adjusted (if overlap occurs) during training phase. In the testing phase, the testing samples are presented one by one as input to all the hyperboxes and each hyperbox calculates its membership value which is used to predict a class value. In this method, θ parameter gives the bound on maximum expandable size of hyperboxes whose range is $0 \leq \theta \leq 1$. Upon the arrival of the input sample, it is checked to find out whether there is a hyperbox that belongs to the same class as that of the input pattern and whether this pattern falls inside it. If such hyperbox is found, then no further processing is required for the current input sample and training algorithm resumes with the next input sample. If such hyperbox do not exist then the below given three steps are executed.

1) Hyperbox expansion: In this step, when a sample is presented, a hyperbox belonging to same class as that of the input pattern is found and it is checked for being capable of expansion to cover the present sample, provided expansion constraint is not violated. If no such hyperbox is found, a new hyperbox is created with min and max points equal to the corresponding points of the input sample.
2) Overlap test: In this step, the overlapping area of the hy- perbox which was expanded in previous step is checked against all hyperboxes that have class labels other than the class label of the current pattern. To determine if this expansion has created any overlap, a dimension by dimension comparison of the two hyperboxes is performed. To remove the overlap, the dimension which has the minimum overlap is selected for contraction.
3) Contraction: If there is no overlapping between the two hyperboxes, this step remains unexecuted either; otherwise, the hyperboxes involved in the overlap are contracted appropriately.

### B. Enhanced fuzzy min-max neural network

In [6] many modifications to the learning steps of FMN are suggested. As result of these modifications proposed in EFMN, classification performance is improved. These proposed modifications are as follows:

1) Hyperbox expansion rule: The present expansion crite- rion of FMN can cause overlapping among the hyper- boxes of different classes. The new constraint proposed avoids such expansions. Specifically, it checks each dimension separately whether it exceeds θ. If yes, then the expansion is not performed.
2) Hyperbox overlap test rule: The overlap test cases pre- sented in FMN are insufficient to identify all overlap- ping cases. To tackle this problem, additional cases are introduced to detect other possible overlapping areas.
3) Hyperbox contraction rule: Appropriate contraction cri- terion is given in EFMN for each of the overlapping cases which ensures minimum disturbance to the hyper- boxes involved in the overlap.

### C. Rule Extraction using Partial Decision Trees

The rule extraction methods can be assembled in three categories based on the level of details of the fundamental ANN: pedagogical, decompositional and eclectic[10]. The pedagogical group of rule extraction methods is of the highest granularity which treats the ANN as a black box and gives relationships between the inputs and the outputs globally. Decompositional methods looks at the ANN at its smallest level of detail that is, each hyperbox and each output unit is analyzed and rules are extracted from these units. The parts of rules are then summed to present global relationships between inputs and outputs. But using these two groups, all rule extraction techniques cannot be mapped properly. Therefore a new type 'eclectic' was added to handle particularly hybrid techniques which inspect the individual units of ANN and extracts global rules for classification. The goal of the classification model presented by us is to use the trained EFMN and extract eclectic type if then rules. For this, the knowledge acquired in the EFMN is used as input to PART tree and the global rules giving the relationships between input attribute and output class information of the EFMN network are extracted.

Rule extraction for classification usually has two stages. Firstly, rules are induced. In second stage, they are refined using a complex global optimization method. This is normally accomplished in one of the following two ways:

- By generating a decision tree and then mapping the decision tree to a rule set or refining the rule set based on its boundaries of the coverage achieved by each rule,
- By employing the divide-and-conquer technique.

As with decision trees this method usually involves a rule optimization step. Frank and Witten in [4] have proposed an algorithm called PART which combined these two approaches in attempt to solve the problems that can arise with both these techniques. This algorithm employs divide and conquer technique, it constructs a rule and remove its cover. This is performed continuously till all the input samples are covered. The rule formation step is different as compared to standard divide-and-conquer methods as the PART algorithm constructs a partial pruned decision tree for a subset of input samples and the leaf of the tree with largest coverage is translated into a rule and then the tree is removed. The overpruning problem is avoided using pruned decision tree.

## III. PROPOSED APPROACH FOR RULE EXTRACTION

EFMN network indeed gives improved results as compared to FMN because of the improvements in terms of its expansion criterion and addition of unrecognized overlap and contraction cases. It's significant drawback that it is a 'black box model'. Hence domain specialists do not tend to use it as solution for safety critical tasks. It encourages the need of framework for rule extraction from EFMN that gives the justification of the results. These 'black box models' can be converted into a 'white box' by translating their internal knowledge into a set of comprehensible and meaningful rules [7].

In proposed approach, we have considered trained EFMN. Min max values of the hyperboxes and their respective class values are used as input to the PART algorithm. PART uses separate and conquer technique, to build the rule using input instances. The output is a pruned ruleset in the form of If then. The experimental setup is discussed in the next section.

## IV. EXPERIM ENTAL SETUP AND RESULT DISCUSSIONS

The data set selected for experiments is the Fisher iris data [3]. This data set is selected because it contains a good mix of patterns that are linearly and non-linearly separable and this is perhaps the best known database to be found in the pattern recognition literature and is referenced frequently to this day. Hence the tremendous number of results available from a wide range of classification techniques which can provide a measure of relative performance. The iris data consisted of 150 four-dimensional feature vectors that represent plant attributes, namely sepal length, sepal width, petal length and petal width (all in cm), which separates them in three separate classes (Iris Setosa, Iris Versicolour and Iris Virginica), 50 instances of each class. Following two experiments are conducted. First, observes the effectiveness of EFMN over FMN and second compares the rules extracted from FMN and EFMN.

### A. Effectiveness of EFMN over FMN

In this experiment, θ is varied from 0 to 1 in the step of 0.1. For each value of θ, FMN and EFMN are executed for randomized input data. Such 10 runs are performed. Table I shows minimum percentage accuracy (Min PA), maximum percentage accuracy (Max PA) and average percentage accuracy (Avg PA) for each value of θ for FMN and EFMN. 100% Iris input data set is considered for training and testing. Table II gives the number of hyperboxes created in the middle layer of FMN and EFMN for each value of θ. The maximum, minimum and average number of hyperboxes are calculated from 10 random runs of the input data.

TABLE I: TABLE SHOWING MINIMUM, MAXIMUM AND AVERAGE PERCENTAGE ACCURACY FOR IRIS DATA SET

| Algorithm | FMN | | | EFMN | | |
|---|---|---|---|---|---|---|
| Theta | Min PA | Max PA | Avg PA | Min PA | Max PA | Avg PA |
| 0 | 100 | 100 | 100 | 100 | 100 | 100 |
| 0.1 | 94 | 99.33 | 97.8 | 95.33 | 99.33 | 97.73 |
| 0.2 | 88 | 91.33 | 89.47 | 90.67 | 98 | 95.47 |
| 0.3 | 88 | 93.33 | 90.73 | 88 | 95.33 | 91.13 |
| 0.4 | 87.33 | 94.67 | 89.53 | 87.33 | 92 | 90 |
| 0.5 | 88.67 | 94 | 90.47 | 87.33 | 95.33 | 89.27 |
| 0.6 | 87.33 | 92.67 | 89.4 | 86 | 96.67 | 90.67 |
| 0.7 | 86.67 | 94.67 | 90.13 | 87.33 | 93.33 | 89.2 |
| 0.8 | 87.33 | 94.67 | 90.27 | 88.67 | 93.33 | 90.6 |
| 0.9 | 88 | 93.33 | 90 | 85.33 | 96 | 90.93 |
| 1 | 87.33 | 94.66 | 89.6 | 87.33 | 92.67 | 88.93 |

TABLE II: TABLE SHOWING MINIMUM, MAXIMUM AND AVERAGE PERCENTAGE HYPERBOXES FOR IRIS DATA SET

| Algorithm | FMN | | | EFMN | | |
|---|---|---|---|---|---|---|
| Theta | Min # HBs | Max # HBs | Avg # HBs | Min # HBs | Max # HBs | Avg # HBs |
| 0 | 147 | 147 | 147 | 147 | 147 | 147 |
| 0.1 | 16 | 43 | 27.9 | 26 | 38 | 32.3 |
| 0.2 | 3 | 4 | 3.6 | 6 | 15 | 9.8 |
| 0.3 | 3 | 3 | 3 | 4 | 4 | 4 |
| 0.4 | 3 | 3 | 3 | 3 | 3 | 3 |
| 0.5 | 3 | 3 | 3 | 3 | 3 | 3 |
| 0.6 | 3 | 3 | 3 | 3 | 3 | 3 |
| 0.7 | 3 | 3 | 3 | 3 | 3 | 3 |
| 0.8 | 3 | 3 | 3 | 3 | 3 | 3 |
| 0.9 | 3 | 3 | 3 | 3 | 3 | 3 |
| 1 | 3 | 3 | 3 | 3 | 3 | 3 |

### B. Rule generation

The rule generation part is carried out by using PART algorithm. Min max values of hyperboxes and respective class labels of the trained FMN and EFMN are passed to PART algorithm and pruned rule sets are generated. The θ value used was 0.1. The rules generated for FMN and EFMN are shown in Table III and IV respectively. Table V gives the comparison of the results of the two algorithms when Iris dataset is tested against the generated rules, which clearly proves that EFMN is effective than FMN.

## V. CONCLUSION AND FUTURE WORK

We have presented the PART based rule extraction method from trained EFMN in this paper. As per the results, EFMN network is more accurate as compared to FMN. Rule learning algorithm has generally two stages, initially rules are induced and later, in the second stage, they are optimized using com- plex global optimization procedures. PART algorithm uses di- vide and conquer technique. It constructs a rule and eliminates its cover, it is repeated until all input instances are covered. PART generates minimal decision tree. The rule optimization step, which is computationally expensive is eliminated in PART. We have trained EFMN for appropriate values of θ. Then the knowledge acquired in the EFMN network is passed to PART algorithm. The generated rules are smaller in number, readable, easy to understand. Hence the drawback of EFMN based model being 'black box models' can be removed. FEMN network is capable of learning in single pass through the input data. This capability combined with PART based rule extraction makes this model to be used in sensitive applications as the induced rules can be used for the justification of the prediction given by EFMN. As part of future work, for more accurate results, the membership function used in EFMN can be modified such that it gives gradually decreasing membership value for an input sample as its distance from hyperbox increases. Also the EFMN algorithm works with continuous attributes only. Hence it can be modified to work with discrete attributes too.

TABLE III: RULE FRO M FMN FOR IRIS DATA SET

| Rule for FMN |
|---|
| If petalwidth >0.075949 AND petallength >0.622725 then class 3, else if petallength <= 0.240506 then 1, else if sepallength >0.620253 then class 2, else class 3 |

TABLE IV: RULES FROM EFMN FOR IRIS DATA SET

| Rule for EFMN |
|---|
| If petalwidth >0.075949 AND petallength >0.620253 then class 3 else if petalwidth <= 0.075949 then class 1 else if petalwidth <= 0.202532 then class 2 else class3 |

TABLE V: STATISTICS BASED ON RULES FOR FMN AND EFMN

| Parameter | FMN | EFMN |
|---|---|---|
| Total Number of samples | 150 | 150 |
| Correct Classification Percentage | 142 (94.67) | 147(98) |
| Incorrect Classification Percentage | 8 (5.33) | 3( 2) |
| Kappa statistic | 0.92 | 0.97 |
| Mean absolute error | 0.0525 | 0.0227 |

## VI. REFERENCES


[1] J. M. Benitez, J. L. Castro, and I. Requena. Are artificial neural networks black boxes? IEEE Transactions on Neural Networks, 8(5):1156–1164, Sep 1997.

[2] Wei-Sen Chen and Yin-Kuan Du. Using neural networks and data mining techniques for the financial distress prediction model. Expert Syst. Appl.,36(2):4075–4086, March 2009.

[3] R. Fisher. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7(2):1179–188, 1936.

[4] Eibe Frank and Ian H. Witten. Generating accurate rule sets without global optimization. In Proceedings of the Fifteenth International Conference on Machine Learning, ICML '98, pages 144–151, 1998.

[5] H. Ishibuchi and M. Nii. Generating fuzzy if-then rules from trained neural networks: linguistic analysis of neural networks. 2:1133–1138 vol.2, Jun 1996.

[6] M. F. Mohammed and C. P. Lim. An enhanced fuzzy min max neural network for pattern classification. IEEE Transactions on Neural Networks and Learning Systems, 26(3):417–429, March 2015.

[7] A. Quteishat, C. P. Lim, and K. S. Tan. A modified fuzzy min max neural network with a genetic-algorithm-based rule extractor for pattern classification. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 40(3):641–650, May 2010.

[8] Anas Quteishat and Chee Peng Lim. A modified fuzzy min-max neural network with rule extraction and its application to fault detection and classification. Appl. Soft Comput., 8(2):985–995, March 2008.

[9] P. K. Simpson. Fuzzy min-max neural networks. i. classification. Trans.Neur. Netw., 3(5):776–786, September 1992.

[10] A. B. Tickle, R. Andrews, M. Golea, and J. Diederich. The truth will come to light: directions and challenges in extracting the knowledge embedded within trained artificial neural networks. IEEE Transactions on Neural Networks, 9(6):1057–1068, Nov 1998.

[11] L. A. Zadeh. Fuzzy sets. Information and Control, 8(3):338–353, June 1965.


 