MINING CLASSIFICATION RULES FROM ENHANCED FUZZY MIN-MAX NEURAL NETWORK

Jaishri M. Waghmare¹, Dr. U. V. Kulkarni²

¹Research Scholar
²Professor, Department of Computer Engineering, SGGS Institute of Engineering & Technology Nanded, India

ABSTRACT:

Artificial neural networks (ANNs) have capability in solving complicated problems. But they are “black box” models. So for safety critical prognostic and diagnostic tasks, domain experts often hesitate to use these models as solution. Nevertheless, a “black box” can be converted into a “white box” by translating their internal contents into a set of comprehensible and meaningful rules. In Enhanced fuzzy min max neural network (EFMM) many enhancements as compared to the original fuzzy min max neural network (FMM) are proposed. Hence it is more accurate than original FMM and other related classification algorithms. Like other ANNs, EFMM is also like a black box and the knowledge is expressed in terms of min-max values of the hyperboxes and associated class labels. Hence the justification of classification decision produced by EFMM needs to be obtained in order to make it more suitable to the real world applications. In this paper, we have proposed a simple model to extract predictions in the form of if-then from trained EFMM the knowledge encoded in the network. These rules justify the classification decision given by EFMM without loss of the accuracy. EFMM is trained for the appropriate value of θ such that 100% accuracy is obtained with least number of hyperboxes. The min max values of all the hyperboxes and their respective class labels are used as input to a program which form the if-then rules. These rules are as accurate as the EFMM decision, 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 and compared with PART based rule extraction approach.

Keywords: Rule extraction, enhanced fuzzy min max neural network(s).
INTRODUCTION
Fuzzy sets [5] were introduced by Lotfi A. Zadeh and Dieter Klaua in 1965 as an extension of the classical notion of set. By definition of crisp classic sets, an element is either present in the set or absent. On the other hand, fuzzy sets allow a member to have partial belonging in the set as their boundaries are vague and ambiguous. Hence they are more suitable for genuine real-world systems. Hence fuzzy sets are used as solution in many machine learning and pattern recognition. The concept of ANN has been inspired by biological neural networks. It consists of large number of processing units called nodes or neurons. They are capable of executing in parallel. Nodes are interconnected by synapses called as connecting links and each of the connection links has weight which can be updated. ANNs learn from the inputs provided. Neurofuzzy frameworks are made by combining fuzzy logic and ANNs. Computational efficiency of ANNs and complex class boundaries representation capability of fuzzy logic make FMM an uncommon type of a neurofuzzy framework. The efficiency of neurofuzzy systems is high as compared to the other machine learning techniques. FMM network involves a dynamic network structure. It can learn new patterns online. Hence as like ANNs, it does not need retraining. As per [4] there are several notable properties associated with FMM learning, 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 FMM, there are certain shortcomings in the FMM learning algorithm. Specifically, the present hyperbox expansion criterion can affect FMM 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 FMM, [1] proposed three heuristic rules.

The drawback of all the FMMs is that they need to have explanation capability. The justification of the classification results given by EFMM is not readily available. Hence they are considered as black boxes [7]. So for safety critical prognostic and diagnostic tasks, domain specialists many times hesitate to utilize these models as solution [6]. It encourages the need of classification rules extraction from EFMM that provides the explanation of the classification prediction. [9] and [10] are methods for rule extraction that extracts fuzzy rules in form of if then from trained FMMs. 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. [6] proposed a two-stage system. Modified FMM 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 FMM classification model. The rules are optimized using genetic algorithm with don’t care approach. But these approaches modify the classification model constructed at the end of training. Hence it affects the accuracy of the rules.

In this paper, EFMM is used to train input patterns with different values of \( \theta \) and for each value of \( \theta \) classification accuracy is checked. The maximum value of \( \theta \) which has resulted in maximum
classification accuracy with minimum number of hyperboxes is used. The corresponding knowledge acquired in the EFMM for that highest classification accuracy value of $\theta$ is passed to rule extraction program. PART constructs partial decision tree in order to extract rule. Such extracted rules clarify the classification prediction of EFMM for input test data. The proposed model is able to extract accurate rules which measured in terms rule accuracy and rule fidelity.

The paper is organized as follows. In section II, subsection A presents description of FMM, subsection B gives details of EFMM and subsection C describes rule extraction basics. Section III describes the proposed approach for rule extraction. Experimental results and discussions are given in Section IV. Finally, conclusions are presented in Section V.

[II] BACKGROUND

A. Fuzzy min-max neural network

FMM is a machine learning technique that was introduced in 1992 by Simpsons[4]. 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 FMM 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, $\theta$ 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 hyperbox 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 [1] many modifications to the learning steps of FMM are suggested. As result of these modifications proposed in EFMM, classification performance is improved. These proposed modifications are as follows:

1) Hyperbox expansion rule: The present expansion criterion of FMM can cause overlapping among the hyperboxes of different classes. The new constraint proposed avoids such expansions. Specifically, it checks each dimension separately whether it exceeds $\theta$. If yes, then the expansion is not performed.

2) Hyperbox overlap test rule: The overlap test cases presented in FMM are insufficient to identify all overlapping cases. To tackle this problem, additional cases are introduced to detect other possible overlapping areas.

3) Hyperbox contraction rule: Appropriate contraction criterion is given in EFMM for each of the overlapping cases which ensures minimum disturbance to the hyperboxes involved in the overlap.

C. Rule Extraction

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. There-fore 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 EFMM and extract eclectic type if then rules. For this, the knowledge acquired in the EFMM in the form of min max values is used as input and the global rules giving the relationships between input attribute and output class information of the EFMM network are extracted.

[III] PROPOSED APPROACH FOR RULE EXTRACTION

EFMM network indeed gives improved results as compared to FMM 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 EFMM 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].

The If-then rules having antecedents and consequents parts are extracted from EFMM model. Initially EFMM is trained using complete dataset. As every hyperbox created in this algorithm belongs to meaningful data samples, we have not used any type of pruning method here. The min max points V and W along with class labels are used to passed to a rule formation program. The V
and W values are used to set limits for the corresponding attributes and are used in antecedents part of the rule. The class label is used in the consequent part of rule to predict the class. The experimental setup is discussed in the next section.

[IV] EXPERIMENTAL RESULTS AND DISCUSSIONS

The data set selected for experiments is the Fisher iris data [2]. 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. Table-I presents description of Iris data set.

Table I: Iris Data Set Description

<table>
<thead>
<tr>
<th>Iris Features and Classes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sepal length</td>
<td></td>
</tr>
<tr>
<td>sepal width</td>
<td></td>
</tr>
<tr>
<td>petal length</td>
<td></td>
</tr>
<tr>
<td>petal width</td>
<td></td>
</tr>
<tr>
<td>Class: Iris Setosa (1)</td>
<td></td>
</tr>
<tr>
<td>Iris Versicolour (2)</td>
<td></td>
</tr>
<tr>
<td>Iris Virginica (3)</td>
<td></td>
</tr>
</tbody>
</table>

Rule generation: The UCI datasets Iris is used in this experiment. EFMM is trained using complete dataset. As every hyperbox created by EFMM algorithm belongs to meaningful data sample, we have not used any type of pruning method for the hyperboxes. The normalized min max points V and W are used to limit the values of the corresponding attributes. The maximum value of 0 which gave 100% classification accuracy for the training data is used.

The rules generated from trained EFMM are shown in Table II where 0=0.12 is used EFMM.

Table II: Rules generated using proposed approach

<table>
<thead>
<tr>
<th>Rules are</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. if 0.54 &lt;= A1 &lt;= 0.66 &amp; 0.37 &lt;= A2 &lt;= 0.48 &amp; 0.13 &lt;= A3 &lt;= 0.24 &amp; 0.01 &lt;= A4 &lt;= 0.08 then class=1</td>
</tr>
<tr>
<td>2. else if 0.80 &lt;= A1 &lt;= 0.89 &amp; 0.35 &lt;= A2 &lt;= 0.42 &amp; 0.56 &lt;= A3 &lt;= 0.63 &amp; 0.16 &lt;= A4 &lt;= 0.22 then class=2</td>
</tr>
<tr>
<td>3. else if 0.80 &lt;= A1 &lt;= 0.91 &amp; 0.32 &lt;= A2 &lt;= 0.43 &amp; 0.65 &lt;= A3 &lt;= 0.76 &amp; 0.20 &lt;= A4 &lt;= 0.32 then class=3</td>
</tr>
<tr>
<td>4. else if 0.92 &lt;= A1 &lt;= 0.97 &amp; 0.33 &lt;= A2 &lt;= 0.38 &amp; 0.77 &lt;= A3 &lt;= 0.87 &amp; 0.23 &lt;= A4 &lt;= 0.29 then class=3</td>
</tr>
</tbody>
</table>
The rule set is validated by testing using all the 150 patterns. All the patterns are correctly classified by using the rules.

### Table III: Rules from EFMM for Iris using PART [3]

<table>
<thead>
<tr>
<th>Rules for EFMM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>If</strong> petalwidth &gt; 0.075949 AND petallength &gt; 0.620253 <strong>then</strong> class 3,</td>
</tr>
<tr>
<td>else if petalwidth &lt;= 0.075949 <strong>then</strong> class 1,</td>
</tr>
<tr>
<td>else if petalwidth &lt;= 0.202532 <strong>then</strong> class 2,</td>
</tr>
<tr>
<td>else class 3</td>
</tr>
</tbody>
</table>
In [3] the rule generation part was carried out by using PART algorithm of the WEKS tool. Min max values of hyperboxes and respective class labels of the trained EFMM are passed to PART algorithm and pruned rule sets are generated. The $\theta$ value used was 0.1. The rules generated for are shown in Table III.

**Table IV: Comparison of Results from [3]**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Accuracy using rules</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PART</td>
</tr>
<tr>
<td>Total number of samples</td>
<td>150</td>
</tr>
<tr>
<td>% of correct classification</td>
<td>147(98)</td>
</tr>
<tr>
<td>% of Incorrect classification</td>
<td>3(2)</td>
</tr>
<tr>
<td>No of rules</td>
<td>4</td>
</tr>
</tbody>
</table>

Table IV gives the comparison of the results of [3] with the proposed method. It can be observed that proposed method of rule extraction is accurate and do not modify the constructed classifier in any way. Hence the predictions using rules truly reflect the classification decision.

[V] CONCLUSIONS

We have presented the simple *if-then* form rule extraction method from trained EFMM in this paper. As per the results, EFMM network is more accurate as compared to FMM. We have trained EFMM for appropriate values of $\theta$. Then the knowledge acquired in the EFMM network is passed to a rule mining program. The main feature of the generated rules is that they truly reflect the classification decision of the EFMM model. Moreover the predictions using rules matches with classifier results. Hence this proposed method of rule extraction can be used as readable justification of the classification problem. The generated rules are as accurate as the classifier model, readable, easy to understand. Hence the drawback of EFMM based model being ‘black box models’ can be removed. This makes this model to be used in sensitive applications as the induced rules can be used for the justification of the prediction given by EFMM.
REFERENCES