RESEARCH ARTICLE OPEN ACCESS

Pruning on Fuzzy Min-Max Neural Networks

Sanchita Rothe, Shruti Shete, Rachana Varyani, Sukanya Mulay
[1,2,3,4] Information Technology, Pimpri Chinchwad College of Engineering, Pune, India

ABSTRACT
This paper works on fuzzy min-max classifier neural network implementation. Fuzzy min-max classifier creates hyperboxes for classification. Each fuzzy set hyperbox is an n-dimensional pattern space defined by a min point and max point with a corresponding membership function. The fuzzy min-max algorithm is used to determine the min-max points of the hyperbox. The use of a fuzzy set approach to pattern classification inherently provides degree of membership information that is extremely useful in higher level decision making. A confidence factor is calculated for every FMM hyperbox, and a threshold value defined by user is used to prune the hyperboxes with low confidence factors. This will describe the relationships between fuzzy sets and pattern classification.
Keywords-Fuzzy min–max neural network, Hyperbox, Membership function, Pattern classification, Pruning

I. Introduction

Artificial Neural Network(ANN) has emerged as a research applications tool including classification and regression.ANN are successfully applied across an range of problem domains in areas as diverge as finance, medicine, engineering, physics and biology. They are powerful tool for modeling especially when the underlying data relationship is unknown. ANNs are useful for solving pattern classification problems in many different fields, e.g. medical prognosis and diagnosis, industrial fault detection and diagnosis, etc. In the medical field, ANNs are expanded as diagnostic decision support. The Neural networks (NN) is a wide area of research for the practitioners. NN are useful for the solving the problems of regression and pattern classification. Pattern Recognition techniques provide a unified frame work to study a variety of techniques, in statistics and computer science that are individually useful in many different applications. Pattern classification is the key problem to many engineering solutions such as sonar, radar, seismic, and diagnostic applications that requires the ability to accurately classify a situation. Pattern classification may be defined as the organization of patterns into groups of patterns sharing the same set of properties. Pattern classification covers a wide spectrum of discipline such as Cybernetics, Computer Science, Mathematics, Logic etc. There are several applications of pattern recognition such as Life form analysis, Image processing, Information management system, Weather prediction, Character recognition, Speech and speaker recognition etc.

Artificial neural network is one of the most commonly used classifier technique. The reason for being commonly used is due to some properties such as learning from examples and exhibiting some capability for generalization beyond the training data. Also they have universal approximation property. But the problem with traditional neural network techniques is that they take more training time because of scanning of input data for many times. This problem is solved in fuzzy min-max neural network (FMMN) proposed by Simpson [1].

This paper will represent classifier that creates classes by aggregating several smaller fuzzy sets into a single fuzzy set class. Fuzzy min-max classification neural networks are built using hyperbox fuzzy sets which were initially defined by Zadeh [3]. Fuzzy min-max classification neural networks are built using hyperbox fuzzy sets. A hyperbox defines a region of the n-dimensional pattern space that has patterns with full class membership.

A hyperbox is completely defined by its min-point and its max-point, and a membership function is defined with respect to these hyperbox min-max points. The min-max (hyperbox) membership function combination defines a fuzzy set, hyperbox fuzzy sets are aggregated to form single fuzzy set classes and the resulting structure fits into a neural network framework; hence this classification system is called a fuzzy min-max classification neural network.

Pruning is a technique introduced in FMM to reduce the complexity of network. It is used to remove the hyperboxes with low confidence factor when compared with a user defined threshold.

II. Related Work

There have been several studies associated with Artificial Neural Network for the pattern classification. C.Chong proposed fuzzy min-max hyperbox classifier to solve M-class classification problems. A learning procedure is proposed to generate a fuzzy classifier by adding min-max hyperboxes as needed to ensure that all training patterns are correctly recognized. Fuzzy systems have been used to represent and manipulate data that are fuzzy rather than precise. Ishibuchi presented a heuristic method for generating fuzzy rules was applied to the grid-type fuzzy partitions, and a rule selection method, based on genetic algorithms was then employed to select relevant fuzzy rules from generated fuzzy rules for classifying training patterns in the considered classification problem. Fuzzy min-max classification neural networks are built using hyperbox fuzzy sets.

A fuzzy set A is a subset of the universe of discourse X that admits partial membership. The fuzzy set A is defined as the ordered pair

$$A=\{x,m_A(x)\}$$

Where $x \,\varepsilon\, X$ and $0 \le m_A(x) \le 1$. The membership function $m_A(x)$ describes the degree to which the object x belongs to the set A where $m_A(x) = 0$ represents no membership and $m_A(x)=1$ represents full membership.

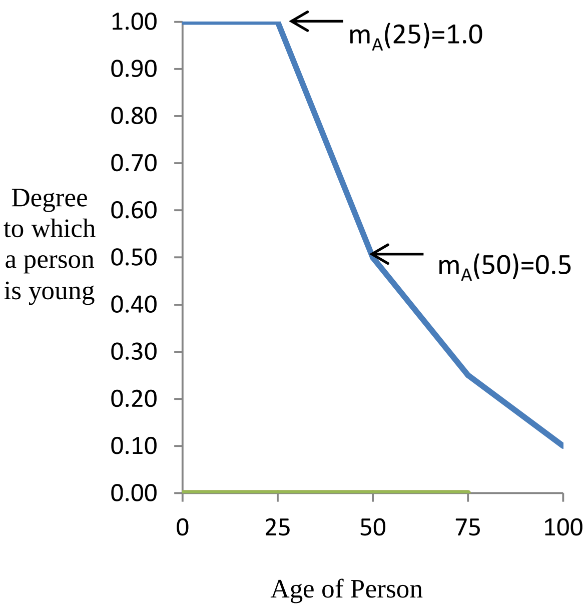

Fig1. Membership Function showing how young a person is

The membership function shown above describes the relationship between person's age and the degree to which a person is considered to be young. This membership function determines that a 25 year old person belongs to A twice as much as a 50 year old person.

2.1 K- Nearest Neighbor Classifier

The fuzzy min-max neural network classifier collapses to the k-nearest-neighbor classifier[2] when the size of the hyperboxes are set to 0 (i.e., $\Theta = 0$). This could be the expansion of point to a hyperbox has several computing advantage.

2.2 Probabilistic NN

The PNN[7] is similar to the fuzzy min max neural network in that it associates membership function with pattern classes, it utilizes a union operation, and it grows to meet the needs of the problem. However, the main difference between the two is PNN and the fuzzy min max neural network classifier uses a Hamming distance based membership function.

III. Methodology

The methodology is divided into three modules:

3.1 Fuzzy Min-Max Classification Algorithm

The Fuzzy Min Max neural network is formed using hyperboxes with fuzzy sets. It defines a region of the n-dimensional pattern space that has patterns with full class membership. The hyperbox can be described using its minimum and maximum points and their corresponding membership functions are used to create fuzzy subsets in the n-dimensional pattern space. The learning process in FMM comprises a series of expansion and contraction processes that fine tune the hyperboxes in the network to establish boundaries among classes. If overlapped hyperboxes of different classes occurred in the pattern space, contraction is performed to eliminate the overlapped areas. We define membership function with respect to the minimum and maximum points of a hyperbox. A pattern which is contained in the hyperbox has the membership function of one.

The definition of each hyperbox fuzzy set[6] as given in Bj, is

$$B_j = \{X, V_j, W_j, f(X, V_j, W_j)\} \quad \forall X \in I^n$$

Where the input pattern is

$$X=(x, x_2 \ldots\ldots x_n),$$

the minimum And maximum points of B_j are

$$V_j=\{v_{j1}, vj_2 \ldots\ldots, v_{jn}\}$$

and $W_j = \{w_{j1}, wj_2 \ldots\ldots w_{jn}\}$ respectively.

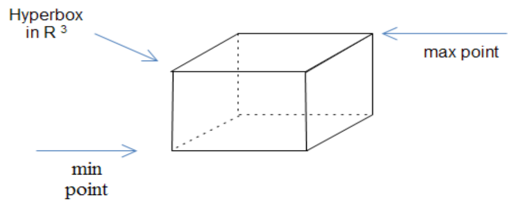

Fig2. A min max hyperbox B_j= { V_j,W_j}

Applying the definition of the hyperboxed fuzzy set, the combined fuzzy set that classifies the Kth pattern class, C_k, is defined as

$$C_k = \bigcup_{j \in k} B_j$$

The membership function for the j^{th} hyperbox b_j (A_h), 0 $\leq$bj (Aj) $\leq$1 measures the degree to which h^{th} input pattern. A_h falls outside hyperbox B_j. As b_j (A_h) approaches 1. The pattern is said to be more ‘‘contained’’ by the hyperbox. The resulting membership function[1] is

$b_j(A_h) = \frac{1}{2n}\sum_{i=1}^{n}[\max(0, 1 - \max(0, \gamma \min(1, a_{hi} - wji + \max(0, 1 - \max(0, \gamma \min(1, vji - ahi)))]$

3.2 Hyperbox Formation

The training set *V* consists of a set of M ordered pairs {X_h,d_h}, where X_h= (X_{h1},X_{h2},..,X_{hn}) □$I^n$is the input pattern and d_h□ {1, 2, ... ,*m*} is the index of one of the m classes.
The fuzzy min-max classification learning algorithm is a three-step process:

3.2.1 Expansion: Identify the hyperbox that can expand and expand it. If an expandable hyperbox cannot be found, add a new hyperbox for that class.

For the hyperbox *Bj* to expand to include X_h, the following constraint must be met[1]

$$n\square \geq \sum_{i=1}^{n}\left(max\left(w_{ji}, x_{hi}\right) - min\left(v_{ji}, x_{hi}\right)\right)$$

If the expansion criterion has been met for hyperbox *Bj*, the main point of the hyperbox is adjusted using the equation.
If the expansion criterion has been met for Hyperbox B_j,min point of the hyperbox is adjusted using equation

$$v_{ji}^{new} = \min(v_{ji}^{old}, x_{hi}) \quad \forall i = 1,2,3, \ldots, n,$$

and the max point is adjusted using the equation

$$w_{ji}^{new} = \max(w_{ji}^{old}, x_{hi}) \quad \forall i = 1,2,3, \ldots, n,$$

3.2.2 Overlap Test: Determine if any overlap exists between hyperboxes from different classes. To determine if this expansion created any overlap, a dimension by dimension comparison between hyperboxes is performed.

Assuming $\square^{old}$ = 1 initially, the four test cases and the corresponding minimum overlap value for the i^{th} dimension are as follows[1]

Case 1: $v_{ji} < v_{ki} < w_{ji} < w_{ki}$,

$$\delta^{new} = \min(w_{ji} - v_{ki}, \delta^{old})$$

Case 2: $v_{ki} < v_{ji} < w_{ki} < w_{ji}$,

$$\delta^{new} = \min(w_{ki} - v_{ji}, \delta^{old})$$

Case 3: $v_{ji} < v_{ki} < w_{ki} < w_{ji}$,

$$\delta^{new} = \min(\min(w_{ki} - v_{ji}, w_{ji} - v_{ki}), \delta^{old})$$

Case 4: $v_{ki} < v_{ji} < w_{ji} < w_{ki}$,

$$\delta^{new} = \min(\min(w_{ji} - v_{ki}, w_{ki} - v_{ji}), \delta^{old})$$

If σ^{0ld} - σ^{new}> 0, then Δ= i and σ^{0ld} = 8^{new}, signifying that there was overlap for the Δ^{th}dimension and overlap testing will proceed with the next dimension.

3.2.3 Contraction: If Δ >0, Δth then Δthdimensions of the two hyperboxes are adjusted. To determine the proper adjustment to make, the same four cases are examined[1]

Case 1: $v_{j\Delta} < v_{k\Delta} < w_{j\Delta} < w_{k\Delta}$,

$$w_{j\Delta}^{new} = v_{k\Delta}^{new} = \frac{w_{j\Delta}^{old} + v_{k\Delta}^{old}}{2}$$

Case 2: $v_{k\Delta} < v_{j\Delta} < w_{k\Delta} < w_{j\Delta}$,

$$w_{k\Delta}^{new} = v_{j\Delta}^{new} = \frac{w_{k\Delta}^{old} + v_{j\Delta}^{old}}{2}$$

Case 3a: $v_{j\Delta} < v_{k\Delta} < w_{k\Delta} < w_{j\Delta}$ and

$$(w_{k\Delta} - v_{j\Delta}) < (w_{j\Delta} - v_{k\Delta})$$

$$w_{k\Delta}^{old} = v_{j\Delta}^{new}$$

Case 3b: $v_{j\Delta} < v_{k\Delta} < w_{k\Delta} < w_{j\Delta}$ and

$$(w_{k\Delta} - v_{j\Delta}) > (w_{j\Delta} - v_{k\Delta})$$

$$w_{j\Delta}^{new} = v_{k\Delta}^{old}$$

Case 4a: $v_{k\Delta} < v_{j\Delta} < w_{j\Delta} < w_{k\Delta}$ and

$$(w_{k\Delta} - v_{j\Delta}) < (w_{j\Delta} - v_{k\Delta})$$

$$w_{k\Delta}^{new} = v_{j\Delta}^{old}$$

Case 4b: $v_{k\Delta} < v_{j\Delta} < w_{j\Delta} < w_{k\Delta}$ and

$$(w_{k\Delta} - v_{j\Delta}) > (w_{j\Delta} - v_{k\Delta})$$

$$w_{j\Delta}^{old} = v_{k\Delta}^{new}$$

3.3Pruning

An effective pruning algorithm is a crucial component of any neural network. Pruned network also serves to filter the noise that might be present in the data. Such noise could be data samples that are outliers or incorrectly labeled. The proposed classification model assumes that the trained network has been pruned to remove the less useful hyperboxes based on their confidence factor to improve the overall system performance. The confidence factor identifies good hyperboxes that are frequently used and generally correct, as well as that are rarely used but extremely accurate. The confidence factor for each hyperbox node is defined in terms of its usage frequency and its predictive accuracy on the predicting set is given by:

$$CF_j = (1 - \alpha)U_j + \alpha A_j$$

Where $U_j \in [0,1]$ is the usage of hyperbox j, $A_j \in [0,1]$ is its accuracy, and $\alpha \in [0,1]$ is a weighting factor.

The value of Uj is defined as the number of patterns in the prediction set classified by any hyperbox j, divided by the maximum number of patterns in the prediction set classified by any hyperbox with the same classification class.

$U_j = C_j$ / max { C_f : hyperbox f predicts class k}

Where C_j is the number of patterns classified by hyperbox j for class k, C_f is the number of patterns classified any hyperbox f for class k.

On the other hand, the value of A_j is defined as the number of correctly predicted set of patterns classified by any hyperbox j, divided by the maximum correctly classified patterns with the same classification class.

$A_j = P_j$ / max{ P_f : hyperbox f correctly predicts class k}

where$P_j$is the number of patterns correctly classified by hyperbox j for class k, P_f is the number of patterns correctly classified by any hyperbox f for class k.

The hyperboxes with a confidence factor less than or equal to some user defined threshold are pruned. Hyperboxes with confidence factor 1 are highly accurate and have high usage. As the confidence factor decreases the usage and accuracy of the corresponding hyperbox also decreases.

IV. EXPERIMENTAL RESULTS

We have considered PID Dataset. Number of Instances 786 which divided into training and testing.Initially Hyperboxes are formed using training dataset, The size of a hyperbox is controlled by θ, that is varied between 0 and 1. once θ is small, more hyperboxes are created.When θ is large, the number of hyperboxes is small. During Pruning, Confidence factor of all hyperboxes are calculated and the hyperbox with the confidence factor less than threshold (user defined) are pruned.

Theta	Hyperbox Number(Before pruning)	Hyperbox Number(After pruning)	Accuracy(In Percentage)
0.1	17	8	61.01
0.3	13	6	66.10
0.5	7	3	67.79
0.9	7	3	67.79

V. CONCLUSION

In this paper, we implemented fuzzy min-max classifier that uses aggregated fuzzy set classes. The proposed system is capable to perform the function of searching reasonable decision boundaries in the overlapping classes, learn highly nonlinear decision boundaries and provide results on a standard data set that was equivalent to other neural and traditional classifiers. Due to the application of pruning algorithms on fuzzy min-max neural network the number of unwanted hyperboxes or rather the hyperbox which do not contribute in decision making are removed. Thus it makes decision making much easier as the complexity of the network is reduced. Our future work will include a mathematical description of decision boundaries that are formed, a study of the effects of the maximum hyper box size *B* on classification accuracy □ on classification accuracy.

VI. ACKNOWLEDGEMENTS

We would like to express our special appreciation and thanks to our Head of Information Technology Department Prof. Sudeep Thepade and Project Coordinator Prof. S.V. Shinde for opening the doors of the department towards the realization of this paper, for many insightful suggestions that helped in properly representing this work and for encouraging our research. Most importantly we would also like to express our sincere gratitude to all who rendered their valuable help, along with all those unseen people across the internet for maintaining those valuable resources for the successful completion of the paper.

REFRENCES

[1] P. Simpson, “Fuzzy min–max neural networks-Part 1: Classification,” IEEE Trans. Neural Networks, Vol. 3,No. 5, September 1992.

[2] J. Bezdek, S. Chuah, and D. Leep, "Generalized k-nearest neighbor rules," *Fuzzy Sets and Systems*, vol. 18, pp. 237-256, 1986..

[3] L. Zadeh , “Fuzzy Sets,” Inform. and control,vol.8,pp.338-353,1965.[4] P. Simpson, “Fuzzy min–max classification with neural networks,” Heuristics, vol. 4, no. 1, pp. 1-9, 1991.

[4] P. Simpson, “Fuzzy min–max neural networks,” in Proc. 1991 Int.Joint Conf. Neural Networks (Singapore).

[5] P. Simpson, “Fuzzy min–max neural networks-Part 2: Clustering,” submitted to IEEE Trans. Fuzzy Syst.

[6] R. Bellman, R. Kalaba, and L. Zadeh, "Abstraction and pattern classification," *RAND Memorandum*, RM-4307-PR, Oct. 1964

[7] D. Specht, “Probabilistic neural networks,” Neural Networks, Vol. 3, No. 1, pp. 109-118, 1990.

[8] AnasQuteishat and CheePeng Lim, “Application of the Fuzzy Min-Max Neural Networks to Medical Diagnosis.”