RESEARCH ARTICLE OPEN ACCESS

Cancer Diagnosis Using Fuzzy Min-Max Neural Network WithRule Extraction

PritamPatil, SheetalWaghole, Musarrat Bare, PoojaHumnabade
(Department of Information Technology, Pune University, India)
(Department of Information Technology, Pune University, India)
(Department of Information Technology, Pune University, India)
(Department of Information Technology, Pune University, India)

ABSTRACT
This paper proposes a modified neural network called Fuzzy Min-Max Neural network (FMMN) that creates hyperboxes for classification and prediction and applied to Women Breast Cancer (WBC) dataset. A modification has been done in an attempt to improve its performance when a large network of hyperboxes forms in the network. The system is composed of formation of hyperbox, Pruning and Prediction. A rule extraction algorithm is also been used in the system. For each hyperbox formation, its confidence factor is calculated and user defined threshold is used to prune the hyper box with low confidence factors. The advantage of this work is that it improves the FMMN performance during the large network of hyperbox formation. It facilitates the extraction of a compact rule set from FMMN to verify its prediction. The evaluation result shows that the proposed system is useful for cancer diagnosis and classification tool in real environments.
Keywords- cancer diagnosis and classification, Fuzzy min–max neural network, Hyperbox fuzzy sets, Rule extraction.

I. INTRODUCTION

Artificial Neural Network (ANN) has emerged as an research applications tool including classification and regression [1][2].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 specially 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 systems that help pathologists diagnose diseases [3]. Following are the characteristics of neural network:

1. It exhibits mapping capabilities. It can map input patterns to their associated output patterns.
2. ANN's are robust systems and are fault tolerance. It can thus, recall full patterns from incomplete, partial or noisy patterns.
3. ANN's can process information in parallel at high speed, and in distributed manner.

There are several applications of ANN for pattern classification such as

- Voice Recognition and Voice synthesis.
- Categorization of radar/sonar signals.
- Remote sensing and image classification
- Credit card applications
- Data mining and information retrieval

Various real life applications of FMMN are:

- Credit scoring
- Breast cancer cell image classification
- Precision direct mailing etc.

With advantages and wide applications, we modified FMMN. It is a pattern classification network based on aggregates of fuzzy hyperboxes. A fuzzy hyperbox defines a region in an n dimensional pattern space. Each input pattern is classified based on the degree of membership to the corresponding hyperboxes. A smaller hyperbox size means that the hyperbox can contain only a smaller number of patterns, which will increase the network complexity. A large size hyperbox means that the hyperbox can contain a larger number of patterns, and will decrease the network complexity.

1.1Basic concept

Basic Concepts of fuzzy min-max neural network are as follows:

1.1.1Fuzzy logic

Fuzzy logic is a form of many-valued logic; it deals with reasoning that is approximate rather than a particular. On comparing with traditional binary sets (where variables may take on true or false values) fuzzy logic variables may have a truth value that ranges in degree between 0 and 1.

1.1.2Hyperbox

A fuzzy set hyperbox is an n-dimensional box defined by a min point and a max point with a corresponding membership function.

1.1.3 Membership function

The membership function of a fuzzy set is a generalization of the indicator function in classical sets. It represents the degree of truth as an extension of valuation in fuzzy logic.

1.1.4Pruning

Pruning is a technique in machine learning that reduces the size of neural network by removing sections of the tree that provide little power to classify instances.

1.2 Fuzzy Min-Max Neural Network

The FMM 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 phase in FMMN consist a series of expansion and contraction processes that fine tune the hyper boxes in the network to establish boundaries among classes. 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*Bj*, is :

$Bj=\{X, Vj, Wj, f(X, Vj, Wj)\} \forall \ X \in I^n$ .. *(1)*

when the input pattern is $X = (\mathrm{x}_2, \mathrm{x}_2, .. \mathrm{x}_\mathrm{n})$, the minimum and maximum points of $B_\mathrm{j}$are :

$Vj = (\mathrm{v}_{\mathrm{j}1}, \mathrm{v}_{\mathrm{j}2}, .. \mathrm{v}_{\mathrm{jn}})$and

$Wj = (\mathrm{w}_{\mathrm{j}1}, \mathrm{w}_{\mathrm{j}2}, .. \mathrm{w}_{\mathrm{jn}})$respectively.

$C_k = \cup_{j \in k} B_j$ …. (2)

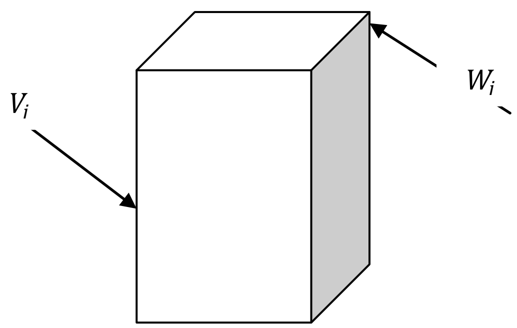

Fig1.min- max hyperbox$B_{j=}\{ V_{j,} W_j \}$

The membership function for the j^{th}hyperbox$b_j \langle A_h \rangle, 0 \leq b_j \langle A_j \rangle \leq 1$ measures the degree to which h^{th}input pattern. $A_h$falls outside hyperbox B_j. As $b_j \langle A_h \rangle$ approaches 1. The pattern is said to be more 'contained' by the hyperbox. Resulting membership function is:

$$b_j \langle A_h \rangle = \frac{1}{2n} \sum_{i-1}^{n} \left[\max(0,1 - \max(0, \gamma \min(1, a_{hi} - w_{ji}))) + \max(0,1 - \max(0, \gamma \min(1, v_{ji} - a_{hi})))\right]$$

where$A_h = (\mathrm{a}_{\mathrm{h}1}, \mathrm{a}_{\mathrm{h}2}, .. \mathrm{a}_{\mathrm{hn}}) \in I^n$ ….(3)

is the h^{th} input pattern $V_j = (\mathrm{v}_{\mathrm{j}1}, \mathrm{v}_{\mathrm{j}2}, .. \mathrm{v}_{\mathrm{jn}})$ is the minimum point of $B_{j\,,}$ and the γ is the sensitivity parameter that regulates how fast the membership values decrease as the distance between A_h and B_j increases.

II. RELATED WORK

There have been several studies associated with Artificial Neural Network for the pattern classification. Chia Chong proposed fuzzy min-max hyperbox classifier to solve M-class classification problems [4]. 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[4]. Fuzzy min-max classification neural networks are built using hyperbox fuzzy sets.

Patrick simpson presented the work on comparing the fuzzy min-max classification neural network with other artificial neural network. The Probabilistic Neural Network (PNN) is similar to the fuzzy min-max neural network in that it associates membership functions with pattern classes, it uses a union operation, and it grows to meet the needs of the problem [5]. The differences between the PNN and the fuzzy min-max neural network classifier are that:

- The PNN stores each data set pattern in the network and the fuzzy min-max neural network classifier utilizes hyper-boxes.
- The Probability Neural Network uses a Euclidian distance metric and the probability density function and the fuzzy min-max neural network classifier uses a Hamming-distance-based membership function.
- The Probability Neural Network normalizes its data to unit length, which destroys the relative magnitude information, and the fuzzy min-max neural network classifier only re-scales the data and retains the relative magnitude information.

III. METHODOLOGY

The Methodology is divided into three major modules that are as follows

3.1 Hyperbox Formation

The training set *V* consists of a set of *M* ordered pairs$\{X_h, d_h\}$, where$X_h = \{x_{h1}, x_{h2}, .. x_{hn}\} \in I^n n$ is the input pattern and $d_h \in \{1,2,..m\}$ is the index of one of the *m* classes.

The fuzzy min-max neural network classification learning algorithm is a three-step process.

3.1.1 Expansion

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

For the hyperbox*Bj*to expand to include X_h, the following constraint must be met :

$$n_\theta \geq \sum_{i=1}^{n}(\max(w_{ji}, x_{hi}) - \min(v_{ji}, x_{hi})) \text{..(4)}$$

If expansion criterion has been met for hyperbox*Bj*, the main point of the hyperbox is adjusted using the equation :

$$v_{ji}^{new} = \min(v_{ji}^{old}, x_{hi}) \; \forall i = 1,2, \dots, n \text{..(5)}$$

If expansion criterion has been met for hyperbox*Bj*, the main point of the hyperbox is adjusted using the equation :

$$w_{ji}^{new} = \min(w_{ji}^{old}, x_{hi}) \; \forall i = 1,2, \dots n \text{..(6)}$$

3.1.2 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 $\sigma^{old} = 1$ initially, the four test cases and the corresponding minimum overlap value for the i^{th}dimension are as follows.

Case1:
$v_{ji} < v_{ki} < w_{ji} < w_{ki}$, $\partial^{new} = \min(w_{ji} -, \partial^{old})$

Case 2:
$v_{ki} < v_{ji} < w_{ki} < w_{ji}$, $\partial^{new} = \min(w_{ki} - v_{ji}, \partial^{old})$

Case 3:
$v_{ji} < v_{ki} < w_{ki} < w_{ji}$, $\partial^{new} = \min(\min(w_{ki} - vji, wji -, \partial old)$

Case 4:
$v_{ki} < v_{ji} < w_{ji} < w_{ki}$, $\partial^{new} = \min(\min(w_{ji} - vki, wki -, \partial old)$

If $\partial^{old} - \partial^{new} > 0$, then $\Delta = i$ and $\partial^{old} = \partial^{new}$, signifying that there was overlap for the Δ^{th}dimension and overlap testing will proceed with the next dimension.

3.1.3 Contraction

If $\Delta > 0$, Δ^{th}thenΔ^{th} dimensions of the two hyperboxes are adjusted. To determine the proper adjustment to make, the same four cases are examined.

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:$v_{j\Delta} < v_{k\Delta} < w_{k\Delta} < w_{j\Delta} \, and \, (w_{k\Delta} - v_{j\Delta}) < (w_{j\Delta} - v_{k\Delta})$,
$v_{j\Delta}^{new} = w_{k\Delta}^{old}$.

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: $v_{k\Delta} < v_{j\Delta} < w_{j\Delta} < w_{k\Delta} \, and \, (w_{k\Delta} - v_{j\Delta}) > (w_{j\Delta} - v_{k\Delta})$,
$v_{k\Delta}^{new} = w_{j\Delta}^{old}$.

3.2 Prediction

The membership functions of all thehyperbox new input pattern is calculated. The hyperbox with membership function greater than threshold (User defined) are selected. The Euclidean distance between the input pattern and the centroid of the hyperbox is exploited [6]. In modified FMMN, the centroids of all input patterns falling in each hyperbox is recorded as follows:

$$C_{ji} = C_{ji} + \frac{|a_{hi} - C_{ji}|}{N_j} \text{ ……………. (8)}$$

where$C_{ij}$is the centroid of the j^{th}hyperbox in the i^{th}dimension, N_j is the number of patterns included in the j^{th}hyperbox. Euclidean distance between the centroid of thej^{th}hyperbox and the h^{th} input pattern, *Ejh* is calculated using:

$$E_{jh} = \sqrt{\sum_{i=1}^{n}(C_{ji} - a_{hi})^2} \text{ ………….. (9)}$$

Suppose hyperboxes 1 and 2 are selected owing to their high membership function values towards the input pattern. *E1* and *E2* are the distances between the input pattern and the centroids of hyperboxes 1 and 2, respectively. Since *E2 < E1*, the input pattern is predicted as belonging to class 2, even though it is contained in hyperbox.

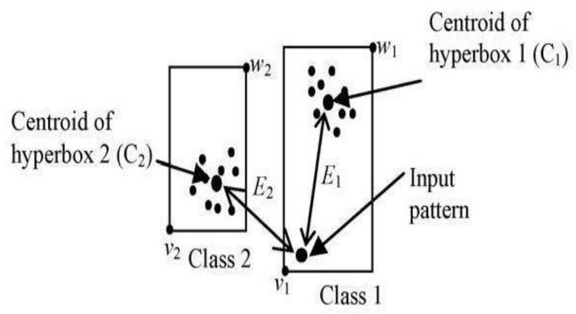

Fig 2.classification process of an input pattern using the euclidean distance.

3.3 Pruning

In this phase, the number of hyper boxes is reduced using pruning. In this phase the confidence factor is calculated. The data set is divided into three phases: Training set, prediction set and test set. The confidence factor can be evaluated as:

$$CF_j=(1-\gamma)U_j+\gamma A_j \text{………… (10)}$$

where Uj is the usage of hyperbox j, Aj the accuracy of hyperbox j and γ $\in[0, 1]$ is a weighing factorhyperboxes with a confidence factor lower than a user defined threshold is pruned. The confidence factor is tagged to each fuzzy if–then rule that is extracted from the corresponding hyperbox.

After pruning the hyperboxes with low confidence factors fuzzy if–then rules are extracted from the remaining hyperboxes using equation 10. The quantization of the minimum and maximum values of the hyperboxes is conducted.

A quantization level Q is equal to the number of feature values in the quantized rule. Quantization is done by the round-off method, in which the interval [0, 1] is divided into Q intervals and assigned to quantization points evenly with one at each end point using:

$$V_q = \frac{q-1}{Q-1}, \text{ where } q = 1..Q \text{…… (11)}$$

The fuzzy if–then rules extracted are in the following form :

Rule R_j IF x_1 is A_q and . . . x_{pn} is A_q

Then x_1 is class C_j with $CF = CF_j$

$j = 1,2, \ldots, N$ where N is the number of hyperboxes $x_p = \{x_{p1}, x_{p2}, .. x_{nn}\}$ is an n dimensional pattern vector,

A_q the antecedent feature value and CF_j is the confidence factor of the corresponding hyperbox[7].

IV. EXPERIMENTAL RESULT

We have considered WBC Dataset. Number of instances 150 out of which 50% is used for training, 30% for prediction, 20% for 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. Confidence Factor (CF)is used to prune hyperbox i.e. CF of all hyperbox are calculated then hyperbox with CF less then threshold are pruned.

Table 1Frequency table for hyperbox creation

Θ	Total Hyperbox Created	Total Hyperbox After Pruning
0.2	29	16
0.3	14	8
0.4	12	7
0.5	10	6
0.6	12	7
0.7	12	7
0.8	15	9
0.9	8	5

Table 1 shows the result of total hyperboxes created before and after pruning when θ value varies from 0.2 to 0.9.

Then during prediction stage the membership function and the Euclidean distance for FMMN to predict its target output. Hyperboxes that have high membership function values is then chosen. The number of hyperboxes chosen can be based on a user-defined threshold, e.g. the highest 10% membership function values. After that, the Euclidean distances between the selected hyperboxes and the input pattern are computed, and the hyperbox with the shortest Euclidean distance is chosen as the winner, and the input pattern belongs to class represented by the hyperbox.

Table 2 Rules Extracted

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**C**	**CF**
R1	5	1-2	5	5	3-4	3-4	4-5	5	4	0.5
R2	3-4	3-4	2-3	1-2	5	5	5	5	2	0.95
R3	5	5	5	1-2	4-5	3-4	4-5	5	4	6.25
R4	3-4	5	5	4-5	5	5	5	5	2	4.31
R5	5	5	5	1-2	3-4	4-5	4-5	5	4	3.0
R6	2-3	4-5	3-4	5	2-3	4-5	3-4	5	4	0.0
R7	3-4	4-5	3-4	5	3-4	4-5	3-4	5	4	0.0

In Table 2 **R** is referred to as rules, whereas **C** is the class of the hyperbox, **CF** is the confidence factor, and 1,2,…,8 are attributes of WBC dataset.
In this, the rules are extracted from the hyperboxes that are pruned and confidence factor of the corresponding hyperbox.

V. CONCLUSION

This paper proposes an innovative work on Fuzzy Min Max Neural Network for WBC dataset. Initially the dataset is divided into two stages viz. training and testing. The trained dataset is set to hyperbox formation and then set to pruning. The confidence factor is calculated for each hyperbox on its usage frequency and its predictive accuracy on the prediction set. Finally, we achieve the input set as the automatic classification in different categories. The evaluation result shows that the proposed system is applicable to the modified FMM network for cancer diagnosis and classification task has been demonstrated.

VI. ACKNOWLEDGEMENTS

We take this opportunity to thank our Head of Information Technology Department Prof. SudeepThepade and Project Coordinator Prof. S.V. Shinde for opening the doors of the department towards the realization of this paper, for their guidance and encouragement. 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.

REFERENCES

[1] P. Devijver, P. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice-Hall, NJ, 1982.
[2] R.O. Duda, P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
[3] N. Belacel, P. Vincke, J.M. Scheiff, M.R. Boulassel, *Acute leukemia diagnosis aid using multicriteria fuzzy assignment methodology*, Comput. Methods Progr.Biomed. 64 (2001) 145–151.
[4] Patrick K. Simpson, *Fuzzy min–max neural networks. Part 1. Classification*, IEEE Trans. Neural Networks 3 (1992) 776–786.
[5] Design of a *fuzzy min-max hyperbox classifier using a supervised learning method CS06.pdf*
[6] D.Specht, *Probabilistic neural networks*, Neural Networks, *vol. 3, no. 1*, pp. 109-118, 1990.
[7] AnasQuteishat, CheePeng Lim, *A modified fuzzy min-max neural network with rule extraction and its application to fault detection and classification*, Elsevier, 2007.