THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Diabetes Diagnosis using Fuzzy Min-Max Neural Network with Rule Extraction and Apriori Algorithm

Swati Shinde
Professor, Pimpri Chinchwad College of Engineering, Pune, India
Sheetal Devram Waghole
Student, Pimpri Chinchwad College of Engineering, Pune, India
Musarrat Munaf Bare
Student, Pimpri Chinchwad College of Engineering, Pune, India
Preetam Pradip Patil
Student, Pimpri Chinchwad College of Engineering, Pune, India
Pooja Mallikarjun Humnabade
Student, Pimpri Chinchwad College of Engineering, Pune, India

Abstract:
This paper proposes a modified neural network called Fuzzy Min-Max Neural network (FMMN) that forms hyperboxes for classification and prediction and applied to Pima Indians Diabetes (PID) dataset. The modifications are made to improve its classification performance when a small number of large hyperboxes are formed in the network. This system is composed of formation of hyperbox, Pruning and Prediction and Rule Extraction. The hyperbox is formed by calculating its confidence factor. The user defined threshold is used to prune the hyper box with low confidence factors. The advantage of pruning is that it improves the FMMN performance during the large network of hyperbox formation also it facilitates the extraction of a compact rule set from FMMN to verify its prediction. Apriori algorithm is used to find closely related attributes using support and confidence factor. From closely related attributes a number of rules are generated. The proposed algorithm is applied on the PID dataset from standard University of California, Irvine (UCI) which demonstrates that the proposed system is useful for diabetes diagnosis and classification tool in real environments.

Key words: *Diabetes diagnosis and classification, Fuzzy min–max neural network, Hyperbox fuzzy sets, Rule extraction, Apriori algorithm, Attribute selection*

1. Introduction

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

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

There are several applications of ANN for pattern classification such as

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

Various real life applications of FMMN are:

- Breast cancer cell image classification
- Precision direct mailing
- Credit scoringetc.

FMMN is a pattern classification network based on aggregates of fuzzy hyperboxes. A fuzzy hyperbox is a n-dimensional box defined by a min point and a max point with a corresponding membership function. The 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.

1.1. Basic Concept
Following are the Basic Concepts of fuzzy min-max neural network:

1.1.1. Hyperbox
A hyperbox is a region of n-dimensional pattern space defined by a min point and a max point with a corresponding membership function.

1.1.2. Fuzzy logic
Fuzzy logic is a form of many-valued logic which 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.

1.1.3. Membership function
Membership function describes the degree to which an input pattern fits within the hyperbox. It decides whether the presented input pattern belongs to a particular class, thus whether the corresponding hyperbox is to be expanded, depends mainly on the membership value.

1.1.4. Pruning
Pruning is a technique used in machine learning that reduces the size of neural network by removing hyperboxes having low confidence factor. The objective of pruning is to extract a small, compact set of rules and, at the same time, to achieve a high accuracy rate when large hyperboxes are formed in FMM.

1.2. Fuzzy Min-Max Neural Network
The FMM network is built using hyperboxes with fuzzy sets. The FMMN 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. The FMMN formation phase consist a series of expansion, overlap test and contraction processes. It helps to 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_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)

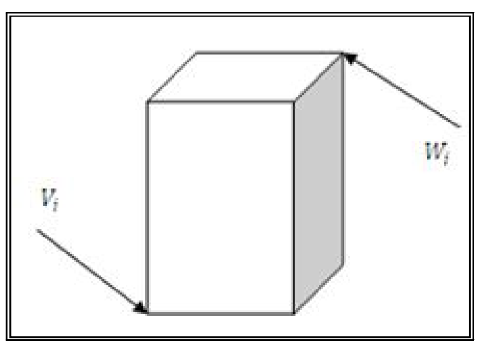

Figure 1: min- max hyperbox $B_{j=}\{ V_j, W_j \}$

The membership function for the j^{th} hyperbox $b_j\langle A_h\rangle, 0 \le b_j\langle A_j\rangle \le 1$ 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:

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

Where $A_h = (a_{h1}, a_{h2}, .. a_{hn}) \in I^n$ ….(3)
is the h^{th} input pattern $V_j = (v_{j1}, v_{j2}, .. v_{jn})$ 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.

2. Related Work

There have been many different types of applications and studies associated with Artificial Neural Network for the pattern classification and pattern detection. Chia Chong proposed fuzzy min-max hyperbox classifier to solve M-class classification problems[4].Fuzzy min-max classification neural networks are developed using hyperbox fuzzy data sets.
Patrick Simpson showed the work on comparing the fuzzy min-max classification neural network with other neural network. 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 PNN and fuzzy min-max neural network classifier are as follows:

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

3. Methodology

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

3.1. Hyperbox Formation

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

3.1.1. Expansion

Identify the hyperbox that can expand the existing hyperbox. If an expandable hyperbox is not found then add a new hyperbox for that class.
For the hyperbox *Bj* to expand to include X_h, the following condition must be met :
$n_\theta \geq \sum_{i=1}^{n} (\max(w_{ji}, x_{hi}) - \min(v_{ji}, x_{hi}))$..(4)
If expansion criterion has been met for hyperbox *Bj,* then main point of the hyperbox is adjusted using following equation:
$v_{ji}^{new} = min(v_{ji}^{old}, x_{hi}) \forall i = 1,2, \ldots, n$..(5)
If expansion criterion has been met for hyperbox *Bj,* then main point of the hyperbox is adjusted using following equation:
$w_{ji}^{new} = min(w_{ji}^{old}, x_{hi}) \forall i = 1,2, \ldots n$ ..(6)

3.1.2. Overlap Test

Overlaping test determine if any overlap exists between hyperboxes from different classes. If this expansion created any overlap, then 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:
$v_{ji} < v_{ki} < w_{ji} < w_{ki}$, $\partial^{new} = \min(w_{ji} - v_{ki}, \partial^{old})$
Case 2:
$v_{ki} < v_{ji} < w_{ki} < w_{ji}$, $\partial^{new} = \min(w_{ki} - v_{ji}, \partial^{old})$
Case 3:
$v_{ji} < v_{ki} < w_{ki} < w_{ji}$, $\partial^{new} = \min(\min(w_{ki} - v_{ji}, w_{ji} -), \partial^{old})$
Case 4:
$v_{ki} < v_{ji} < w_{ji} < w_{ki}$, $\partial^{new} = \min(\min(w_{ji} - v_{ki}, w_{ki} -), \partial^{old})$
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 next dimension.

3.1.3. Contraction

If $\Delta > 0$, Δ^{th} then Δ^{th} dimensions of two hyperboxes are adjusted and then for determining proper adjustment, the same four cases are examined.
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 is the antecedent feature value and CF_j is the confidence factor of the corresponding hyperbox[7].
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 hyperbox with the new input pattern are calculated. The hyperbox with membership function greater than threshold (User defined value) are selected. The Euclidean distance between input pattern and the centroid of the hyperbox is exploited ("6"). In modified FMMN, the centroid of all input patterns falling in each hyperbox is recorded as follows:

$C_{ji} = C_{ji} + \frac{|a_{hi} - C_{ji}|}{N_j}$ …………..… (7)

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

$E_{jh} = \sqrt{\sum_{i=1}^{n}(C_{ji} - a_{hi})^2}$ ………….. (8)

Suppose hyperbox 1 and 2 are selected using their high membership function values towards 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.

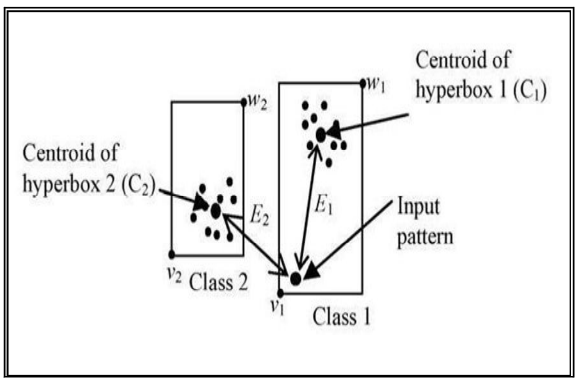

Figure 2: Classification process of an input pattern using euclidean distance

3.3. Pruning

In this phase, number of hyper boxes is reduced using pruning. 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:

$CF_j = (1-\gamma) Uj + \gamma A_j$ ………… (9)

Where Uj is the usage of hyperbox j, Aj the accuracy of hyperbox j and $\gamma \in [0, 1]$ is the weighing factor of hyperbox. The hyperbox with the 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 (pruned hyperboxes) the 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 process 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:

$V_q \frac{q-1}{Q-1}$, where $q = 1..Q$ …… (10)

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

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 is the antecedent feature value and CF_j is the confidence factor of the corresponding hyperbox("7").

4. Experimental Result

We have considered Pima Indian Diabetes (PID) Dataset. Number of instances 200 out of which 55% is used for training, 30% for prediction, and 25% for the testing network. Firstly hyperboxes are formed using the training dataset. The size of a hyperbox is controlled by θ which is varied between 0 and 1. If θ is small, more hyperboxes are created. If θ is large, the number of hyperboxes created is small. Confidence Factor (CF) is used to prune hyperbox i.e. first CF of all hyperbox are calculated then hyperboxes with less CF are pruned.

Θ	**Total Hyperbox Created**	**After Pruning Total Hyperbox**
0.2	42	23
0.3	30	19
0.4	20	13
0.5	15	9
0.6	12	7
0.7	10	6
0.8	12	7
0.9	8	5

Table 1: Table for hyperbox creation

Table 1 represents the result of total number of hyperboxes created before and after pruning when the θ value differs from 0.2 to 0.9. Then during prediction stage the membership function and Euclidean distance for FMMN is calculated to predict its target class. Hyperboxes with high membership function values is then selected. Number of hyperboxes selection can be based on a user-defined threshold value, e.g. highest 15% membership function values. Then after that, the Euclidean distances between the selected hyperboxes and the input pattern are calculated, and the hyperbox with the shortest Euclidean distance is selected as the winner. After that pruning of hyperboxes is performed and rules are extracted for pruned hyperboxes which gives justification to the result that we obtained.

	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

Table 2: Rules Generated

In Table 2 C is class of hyperbox,R is referred to as rules, whereas CF is the confidence factor, and 1, 2… 7, 8 are attributes of PID dataset. Once the rules are generated, attribute selection process is carried out, where each attribute are selected and rules are extracted with individual attribute as well as with combination of different attribute.

Here the attributes are described through 1, 2, 3....8. The support factor is not only calculated for individual attribute but also for combination of attribute. The support factor describes how well the attributes are connected or related to each other. This helps us to know which attribute are strongly associated.

Attributes	Support Factor
1	0.692307
2	0.923076
3	0.769230
4	0.538461
5	0.692307
6	1.0
7	0.923076
8	0.923076
1,2	0.615384
1,3	0.615384
1,4	0.384615
1,5	0.615384
1,6	0.692307
1,7	0.692307
1,8	0.692307
1,2,3	0.538461
1,3,6	0.615384
1,2,4,7	0.307692
1,2,5,8	0.615384
1,2,5,7,8	0.615384
1,3,6,7,8	0.615384
1,3,4,5,6,8	0.307692
1,4,5,6,7,8	0.307692
1,2,3,5,6,7,8	0.538461
1,2,3,4,5,6,7,8	0.307692

Table 3:.Attributes Support factor

5. Conclusion

This paper proposes an innovative work on Fuzzy Min Max Neural Network for PID dataset. Initially dataset is divided into three stages viz. training, prediction and testing. The trained dataset is set to hyperbox creation and then 30% dataset set to predication and pruning and finally 25% dataset set to testing. Confidence factor is calculated for each and every hyperbox on its usage frequency and its predictive accuracy on the prediction dataset. Thesmall number of rules extracted from FMMN allows the network to provide justification for the output generated by FMMN. Finally, Apriori algorithm is used to find closely related attributes using support and confidence measures. From closely related attributes a number of rules are mined. The evaluation result shows that proposed system is applicable to the modified FMM network for diabetes diagnosis and classification task has been demonstrated.

6. Acknowledgements

We take this opportunity to thank our Head of Information Technology Department, Prof. Dr. Sudeep Thepade and Project Coordinator Prof. Mrs. S. V. Shinde for opening the doors of the department towards the realization of this paper, for their help, 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.

7. 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