Exclusion/Inclusion Fuzzy Classification Network

Andrzej BARGIELA [1], Witold PEDRYCZ [2] and Masahiro TANAKA [3]

[1] *The Nottingham Trent University, Nottingham NG1 4BU, UK*
(andre@doc.ntu.ac.uk)

[2] *University of Alberta, Edmonton, Canada*
[2] *Systems Research Institute, Polish Academy of Sciences, Poland*
(pedrycz@ee.ualberta.ca)

[3] *Konan University, 8-9-1 Okamoto, Higashinada-ku, Kobe, Japan*
(m_tanaka@konan_u.ac.jp)

Abstract The paper introduces an exclusion/inclusion fuzzy classification neural network. The network is based on our GFMM [3] and it allows for two distinct types of hyperboxes to be created: inclusion hyperboxes that correspond directly to those considered in GFMM, and exclusion hyperboxes that represent contentious areas of the pattern space. The subtraction of the exclusion hyperboxes from the inclusion hyperboxes, implemented by EFC, provides for a more efficient coverage of complex topologies of data clusters.

1. Introduction and problem statement

Fuzzy hyperbox classification of data has been shown to be a powerful algorithmic approach to deriving intuitive interpretation of data [5, 6, 3, 2, 4, 7]. However, the interpretability of individual hyperboxes comes with their inherent limitation of incompatibility of their shape with the shape of many real-life data clusters. In order to overcome this incompatibility it is necessary to use sets of appropriately sized hyperboxes to cover the more complex topologies of the actual data. The quality of this coverage, measured as a misclassification rate, depends on the maximum size of hyperboxes. The smaller is the maximum size of hyperboxes the more accurate coverage can be obtained. Unfortunately, a direct consequence of that is that the increase of the number of hyperboxes erodes the interpretability of the results. It is therefore necessary to balance the requirement of interpretability with the classification accuracy.

The tradeoff originally proposed by Simpson [5, 6] was the optimization of a parameter defining the maximum hyperbox size as a function of misclassification rate. However, the use of a single maximum hyperbox size proved too restrictive in that for some data clusters there was a need for several hyperboxes while for the other clusters, with a more complex topology, there was still a significant misclassification possible with hyperboxes of such a size. One solution to this problem, proposed in [3], is the adaptation of the size of hyperboxes so that it is possible to generate larger hyperboxes in some areas of the pattern space without sacrificing the recognition rate, while in the other areas the hyperboxes are kept small to afford accurate coverage of the data topology. A similar effect has been also obtained in the context of a partially supervised hyperbox clustering [1].

In this paper we propose a development of the General Fuzzy Min-Max Neural Network (GFMM) [3] architecture that generates two types of hyperboxes. The first type, called inclusion hyperboxes, is the type of hyperboxes that we have considered so far. The second type, called exclusion hyperboxes, represent hyperboxes that contain data belonging to different classes. So, the exclusion hyperboxes represent areas of the pattern space in which the classification is ambiguous. Using these two types of hyperboxes as the basic building blocks it is possible to represent complex topologies of data clusters with a reduced overall number of hyperboxes. Also, the three steps of the GFMM algorithm, namely Expansion, Overlap test and Contraction can be reduced to two i.e. Expansion and Overlap tests.

This paper is organized as follows. Section 2 gives an overview of the GFMM. The new Exclusion/Inclusion Fuzzy Classification (HEFC) network is introduced in section 3. Section 4 provides a brief comparison of the HEFC to FSS and GFMM using the IRIS data set.

2. General Fuzzy Min-Max Neural Network

The neural network that implements the GFMM algorithm is shown in Figure 1. It is a three-layer feed forward neural network that grows adaptively to meet the demands of the classification problem. The input layer has *2*n* processing elements, two for each of the *n* dimensions of the input pattern X_h - $[x_h^l x_h^u]$. Each second-layer node represents a hyperbox fuzzy set where the connections of the first and second layers are the min-max points and the transfer function is the hyperbox membership function. The connections are adjusted using the algorithm described in [3]. The min points matrix **V** is applied to the first *n* input nodes representing the vector of lower bounds x_h^l of the input pattern and the max points matrix **W** is applied to the second *n* input nodes representing the vector of upper bounds x_h^u. The connections between the second- and third-layer nodes are binary values. They are stored in the matrix **U**. The elements of **U** are defined as follows:

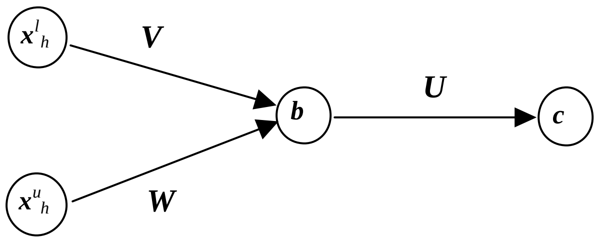
$$u_{jk} = \begin{cases} 1, & if\ b_j\ is\ a\ hyperbox\ for\ class\ c_k \\ 0, & otherwise \end{cases} \quad (1)$$

where b_j is the *j*th second-layer node and c_k is the *k*th third-layer node. Each third-layer node represents a class. The output of the third-layer node represents the degree to which the input pattern X_h fits within the class *k*. The transfer function for each of the third-layer nodes is defined as

$$c_k = \max_{j=1}^{m} b_j u_{jk} \quad (2)$$

for each of the *p+1* third-layer nodes. Node c_0 represents all unlabelled hyperboxes from the second layer. The outputs of the class layer nodes can be fuzzy when calculated using expression (2), or crisp when a value of one is assigned to the node with the largest c_k and zero to the other nodes.

The topology of the network depicted in Figure 1 is almost identical to the original fuzzy min-max neural network proposed by Simpson [5] except for two changes. First, the number of input nodes has been increased from *n* to *2*n*. This has eliminated the need for double connections from input nodes to second-layer nodes. Second, an additional node c_0 representing all the unlabelled hyperboxes from the second layer has been included in the output layer. This allows combined consideration of the classification and clustering of data.



$$\boldsymbol{x}^l_h = [x^l_{h1}, \ldots, x^l_{hn}],\ \boldsymbol{x}^u_h = [x^u_{h1}, \ldots, x^u_{hn}]$$
$$\boldsymbol{V}=[v_1,\ldots,v_m],\ \boldsymbol{W}=[w_1,\ldots,w_m]$$
$$\boldsymbol{b}=[b_1,\ldots,b_m]$$
$$\boldsymbol{U}=[u_{10},\ldots,u_{1p};\ldots;\ u_{m0},\ldots,u_{mp}]$$
$$\boldsymbol{c}=[\ c_0,\ c_1,\ldots,c_p]$$

Figure 1. The three-layer neural network implementation of the GFMM algorithm.

The training of the neural network involves adaptive construction of hyperboxes guided by the class labels. The input patterns are presented in a sequential manner and are checked for a possible inclusion in the existing hyperboxes. If the pattern is fully included in one of the hyperboxes no adjustment of the min- and max-point of the hyperbox is necessary, otherwise a hyperbox *expansion* is initiated. However, before the expansion can be confirmed it is necessary to perform an *overlap test* since it is possible that the postulated expansion could result in some areas of the pattern space belonging simultaneously to two distinct classes, thus contradicting the classification itself. If the overlap test is negative, the postulated hyperbox expansion is confirmed and the next input pattern is being considered. If, on the other hand, the overlap test is positive the hyperbox *contraction* procedure is initiated. This involves subdivision of the hyperboxes along the overlapping coordinates and the consequent adjustment of the min- and max-points. However, the contraction procedure has an inherent weakness in that it inadvertently eliminates from the two hyperboxes some part of the pattern space that was unambiguous while retaining some of the contentious part of the pattern space in each of the hyperboxes. This is illustrated in Figure 2.

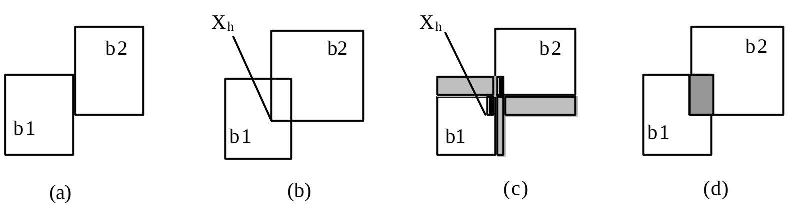
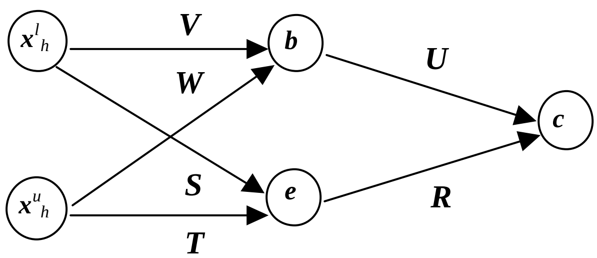


Figure 2. GFMM hyperbox expansion/contraction (a)-(c) and the proposed exclusion/inclusion hyperboxes (d)

The attempted expansion of hyperbox *b2* (because the pattern X_h belongs to the same class as hyperbox *b2*) is shown in Figure 2(b). However, the expansion creates an overlap with the hyperbox *b1,* which is assumed here to belong to a different class than hyperbox *b2*. This initiates the contraction procedure that is illustrated in Figure 2(c). The result of the contraction is that a proportion of the original hyperboxes *b1* and *b2* is discarded (areas highlighted with diagonal lines) while part of the new hyperboxes *b1* and *b2* still remains contentious (areas marked with a solid black fill). Note that the pattern X_h is still included in the new hyperbox *b1* (see Figure 2(c)) therefore it is being misclassified. Clearly there is a scope for a fine-tuning of the contraction procedure, however the result is always a comp romise between the loss of correctly classified patterns and the inclusion of areas of the pattern space where the misclassification occurs.

3. Exclusion/Inclusion Fuzzy Classification network (EFC)

The solution proposed here is the explicit representation of the contentious areas of the pattern space as *exclusion hyperboxes*. This is illustrated in Figure 2(d). The original hyperbox *b1* and the expanded hyperbox *b2* do not lose any of the undisputed area of the pattern space but the patterns contained in the exclusion hyperbox are eliminated from the relevant classes in the $\{c_1,\ldots,c_p\}$ set and are instead assigned to class c_{p+1} (contentious area of the pattern space class). This overruling implements in effect the subtraction of hyperbox sets which allows for the representation of non-convex topologies with a relatively few hyperboxes.



$$\boldsymbol{x}^l_h = [x^l_{h1}, \ldots, x^l_{hn}],\ \boldsymbol{x}^u_h = [x^u_{h1}, \ldots, x^u_{hn}]$$
$$\boldsymbol{V}=[v_1,\ldots,v_m],\ \boldsymbol{W}=[w_1,\ldots,w_m],\ \boldsymbol{S}=[s_1,\ldots,s_m],\ \boldsymbol{T}=[t_1,\ldots,t_m]$$
$$\boldsymbol{b}=[b_1,\ldots,b_m],\ \boldsymbol{e}=[e_1,\ldots,e_q]$$
$$\boldsymbol{U}=[u_{10},\ldots, u_{1p}, u_{1(p+1)};\ldots; u_{m0},\ldots,u_{mp}, u_{m(p+1)}]$$
$$\boldsymbol{R}=[r_{10},\ldots, r_{1p}, r_{1(p+1)};\ldots; r_{q0},\ldots,r_{qp}, r_{q(p+1)}]$$
$$\boldsymbol{c}=[c_0, c_1,\ldots,c_p,c_{p+1}]$$

Figure 3. Exclusion/Inclusion Fuzzy Classification Network.

The additional second-layer nodes ***e*** are formed adaptively in a similar fashion as for nodes ***b***. The min-point and the max-point of the exclusion hyperbox are identified when the overlap test is positive for two hyperboxes representing different classes. These values are stored as new entries in matrix ***S*** and matrix ***T*** respectively. If the new exclusion hyperbox contains any of the previously identified exclusion hyperboxes, the included hyperboxes are eliminated from the set ***e***. The connections between the nodes ***e*** and nodes ***c*** are binary values stored in matrix ***R***. The elements of ***R*** are defined as follows:

$$r_{lk} = \begin{cases} 1, & \text{if } e_l \text{ is a hyperbox overlapping with class } c_k \text{ and } 1 < k < p \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Note that the third layer has $p+2$ nodes [c_0, c_1,…, c_p, c_{p+1}] with the node c_{p+1} representing the new exclusion hyperbox class. The output of the third-layer is now moderated by the output from the exclusion hyperbox nodes ***e*** and the values of matrix ***R.*** The transfer function for the third-layer nodes is defined as:

$$c_k = \max\left(0,\ \max_{j=1}^{m} b_j u_{jk} - \max_{i=1}^{q} e_i r_{ik}\right) \quad (4)$$

The second component in (4) cancels out the contribution from the overlapping hyperboxes that belonged to different classes. Since we allow both fuzzy and crisp outputs of the class layer we restrict the minimum value of c_k to 0.

4. Numerical example

The EFC was applied to a number of synthetic data sets and demonstrated improvement over the GFMM and the original FMM [5]. As a representative example, we illustrate the performance of the network using the Iris data-set. The network was trained using the first 75 patterns and the EFC performance was checked using the remaining 75 patterns. The results for FMM have been obtained using our implementation of the FMM algorithm which produced results consistent with those reported in [5]. The results are summarized in Table 1.

Table 1. Comparison of performance of FMM, GFMM and EFC

Performance criterion	**FMM [5]**	**GFMM [3]**	**EFC**
Recognition rate (range)	97.33-92%	100-92%	100-97%
Number of hyperboxes (max. size 0.06)	32	29	18
Number of hyperboxes (max. size 0.03)	56	49	34

It is clear that the number of misclassifications has been significantly reduced while the number of hyperboxes has been reduced as well. This is due to the increased expressive power of the combined exclusion and inclusion hyperboxes.

Further results will be reported at the conference.

Acknowledgments

Support from the Engineering and Physical Sciences Research Council (UK), the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Konan University is gratefully acknowledged.

References

1. Bargiela A., Pedrycz W., Granular clustering with partial supervision, *Proc. European Simulation Multiconference, ESM2001*, Prague, June 2001, 113-120.
2. D. Dubois and H. Prade, Fuzzy relation equations and causal reasoning, *FSS*, 75, 1995, pp. 119-134
3. Gabrys B., Bargiela A., General fuzzy min-max neural network for clustering and classification, *IEEE Trans. Neural Networks*, vol.11, 2000, 769-783.
4. W. Pedrycz, F. Gomide, *An Introduction to Fuzzy Sets*, Cambridge, MIT Press, Cambridge, MA, 1998.
5. Simpson PK., Fuzzy min-max neural networks - Part1: Classification, *IEEE Trans. Neural Networks*, vol.3, 5, 1992, 776-786.
6. Simpson PK., Fuzzy min-max neural networks – Part 2: Clustering, *IEEE Trans. Neural Networks*, vol.4, 1, 1993, 32-45.
7. Zadeh LA, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems (FSS)*, 90, 1997, 111-117.