International Journal of Knowledge-based and Intelligent Engineering Systems 8 (2004) 1–8 KES9 1
IOS Press

An inclusion/exclusion fuzzy hyperbox classifier

Andrzej Bargiela[a,*], Witold Pedrycz[b,c] and Masahiro Tanaka[d]
[a]*The Nottingham Trent University, Nottingham NG1 4BU, UK*
[b]*University of Alberta, Edmonton, Canada*
[c]*Systems Research Institute, Polish Academy of Sciences, Poland*
[d]*Konan University, 8-9-1 Okamoto, Higashinada-ku, Kobe, Japan*

Abstract. In this study we consider the classification (supervised learning) problem in $[0\ 1]^n$ that utilizes fuzzy sets as pattern classes. Each class is described by one or more fuzzy hyperbox defined by their corresponding minimum- and maximum vertices and the hyperbox membership function. Two types of hyperboxes are created: inclusion hyperboxes that contain input patterns belonging to the same class, and exclusion hyperboxes that contain patterns belonging to two or more classes, thus representing contentious areas of the pattern space. With these two types of hyperboxes each class fuzzy set is represented as a union of inclusion hyperboxes of the same class minus a union of exclusion hyperboxes. The subtraction of sets provides for efficient representation of complex topologies of pattern classes without resorting to a large number of small hyperboxes to describe each class. The proposed fuzzy hyperbox classification is compared to the original Min-Max Neural Network and the Gene ral Fuzzy Min-Max Neural Network and the origins of the improved performance of the proposed classification are identified. These are verified on a standard data set from the Machine Learning Repository.

Keywords: Pattern classification, fuzzy hyperbox, min-max neural networks, information granulation

1. Introduction and problem statement

Fuzzy hyperbox classification derives from the original idea of Zadeh [9] of using fuzzy sets for representation of real-life data. Such data frequently is not *crisp* (has a binary inclusion relationship) but rather has a property of a *degree of membership*. In this case the use of traditional set theory introduces unrealistic constraint of forcing binary decisions where the graded response is more appropriate. An early application of fuzzy sets to the pattern classification problem [2] proves the point that fuzzy sets represent an excellent tool simplifying the representation of complex boundaries between the pattern classes while retaining the full expressive power for the representation of the core area for each class. By having classes represented by fuzzy set membership functions it is possible to describe the degree to which a pattern belongs to one class or another.

Bearing in mind that the purpose of classification is the enhancement of interpretability of data or, in other words, derivation of a good abstraction of such data the use of hyperbox fuzzy sets as a description of pattern classes provides clear advantages. Each hyperbox can be interpreted as a fuzzy rule. However, the use of a single hyperbox fuzzy set for each pattern class is too limiting in that the topology of the original data is frequently quite complex (and incompatible with the convex topology of the hyperbox). This limitation can be overcome by using a collection (union) of hyperboxes to cover each pattern class set [4,7,8]. Clearly, the smaller the hyperboxes the more accurate cover of the class set can be obtained. Unfortunately, this comes at the expense of increasing the number of hyperboxes thus eroding the original objective of interpretability of the classification result. We have therefore a task of balancing the requirements of accuracy of coverage of

*Corresponding author. E-mail: andre@doc.ntu.ac.uk.

ISSN 1327-2314/04/$17.00

the origin al data (which translates on the minimization of misclassifications) with the interpretability of class sets composed of many hyperboxes.

The tradeoff originally proposed by Simpson [7] was the optimization of a single parameter defining the maximum hyperbox size as a function of misclassification rate. However, the use of a single maximum hyperbox size is somewhat restrictive. For class sets that are well separated from each other the use of large hyperboxes is quite adequate while for the closely spaced class sets, with a complex partition boundary, there is a need for small hyperboxes, so as to avoid high misclassification rates. One solution to this problem, proposed in [4], is the adaptation of the size of hyperboxes so that it is possible to generate larger hyperboxes in some areas of the pattern space while in the other areas the hyperboxes are constrained to be small to maintain low misclassification rates. The adaptation procedure requires however several presentations of data to arrive at the optimum sizes of hyperbox sizes for the individual classes.

In this paper we take an alternative approach to achieving low misclassification rate while maintaining good interpretability of the classification results. Rather than trying to express the class sets as a union of fuzzy hyperbox sets [4,7], we represent them as a difference of two fuzzy sets. The first set is a union of hyperboxes produced in the standard way and the second set is a union of intersections of all hyperboxes that belong to different classes. We will refer to the first type of hyperboxes as inclusion hyperboxes and the second type as exclusion hyperboxes. By subtracting the exclusion hyperboxes from the inclusion ones it is possible to express complex topologies of the class set using fewer hyperboxes. Also, the three steps of the Min-Max clustering [4,7], namely *expansion, overlap test and contraction* can be reduced to two, namely *expansion* and *overlap tests*. Expansion step results in generating inclusion hyperboxes and the overlap test results in exclusion h yperboxes.

This paper is organized as follows. The fuzzy Min-Max classification algorithm is overviewed in Section 2. In Section 3 we discuss problems inherent to the Min-Max algorithm. This gives basis for the development of a new exclusion-inclusion fuzzy hyperbox classification algorithm described in Section 4. Section 5 describes a numerical example illustrating the algorithm.

2. Fuzzy Min-Max classification

The fuzzy Min-Max classification neural networks are built using hyperbox fuzzy sets. A hyperbox defines a region in R^n, or more specifically in $[0\ 1]^n$ (since the data is normalized to [0 1]) and all patterns contained within the hyperbox have full class membership. A hyperbox B is fully defined by its minimum V and maximum W vertices. So that, $B = [V, W]$ $[0\ 1]^n$ with $V, W \in [0\ 1]^n$.

Fuzzy hyperbox B is described by a membership function (in addition to its minimum and maximum vertices), which maps the universe of discourse (X) into a unit interval

$$B : X \to [0, 1] \tag{1}$$

Formally, B(x) denotes a degree of membership that describes an extent to which x belongs to B. If $B(x) = 1$ then we say that x fully belongs to B. If $B(x)$ is equal to zero, x is fully excluded from B. The values of the membership function that are in-between 0 and 1 represent a partial membership of x to B. The higher the membership grade, the stronger is the association of the given element to B. In this paper we will use an alternative notation for the hyperbox membership function $b(X, V, W)$ which gives an explicit indication of the min- and max- points of the hyperbox. The hyperbox fuzzy set will then be denoted as $B = X, V, W, b(X, V, W)$. Note that X is an input pattern that in general represents a class-labelled hyperbox in $[0\ 1]^n$. To put it formally

$$X = [X^l\ X^u], d \tag{2}$$

where X^l and X^u represent min and max points of the input hyperbox X and $d \in 1, \ldots, p$ is the index of the classes that are present in the data set.

While it is possible to define various hyperbox membership functions that satisfy the boundary conditions with regard to full inclusion and full exclusion, it is quite intuitive to adopt a function that ensures monotonic (linear) change in-between these extremes. Following the suggestion in [4] we adopt here

$$b_j(X_y h) = \min_{i=1,\ldots,n} (\min([1 - f(x^u_{hi} - w_{ji}, \gamma_i)], [1 - f(\nu_{ji} - x^l_{hi}, \gamma_i)])) \tag{3}$$

where $f(r, \gamma) = \begin{cases} 1 & if \quad r\gamma > 1 \\ r\gamma & if \quad 0 \leqslant r\gamma \leqslant 1 \\ 0 & if \quad r\gamma < 0 \end{cases}$ is a two parameter function in which r represents the distance of the test pattern X_h from the hyperbox $[VW]$ and $\gamma = [\gamma_1, \gamma_2, \ldots, \gamma_n]$ represents the gradient of change

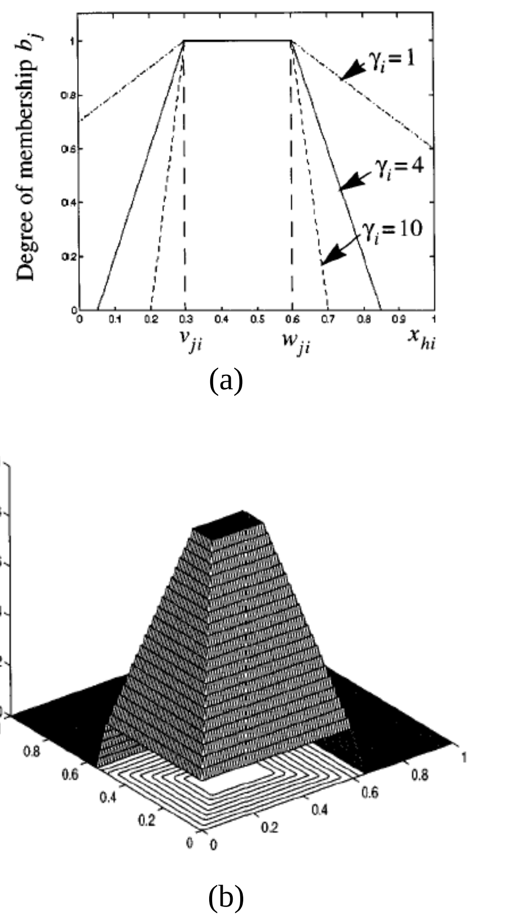


Fig. 1. One-dimensional (a) and two-dimensional (b) fuzzy membership function evaluated for a point input pattern X_h.

of the fuzzy membership function. This is illustrated in Fig. 1.

The fuzzy Min-Max algorithm is initiated with a single point hyperbox $[V_j\ W_j] = [0\ 0]$. However, this hyperbox does not persist in the final solution. As the first input pattern $X_h = [X_h^1\ X_h^u], d$ is presented the initial hyperbox becomes $[V_j\ W_j] = [X_h^1\ X_h^u]$. Presentation of subsequent input patterns has an effect of creating new hyperboxes or modifying the size of the existing ones. A special case occurs when a new pattern falls inside an existing hyperbox in which case no modification to the hyperbox is needed.

Hyperbox expansion: When the input pattern X_h is presented the fuzzy membership function for each hyperbox is evaluated. This creates a preference order for the inclusion of X_h in the existing hyperboxes. However the inclusion of the pattern is subject to two conditions: (a) the new pattern can only be included in the hyperbox if the class label of the pattern and the hyperbox are the same and (b) the size of the expanded hyperbox that includes the new pattern must not be greater in any dimension than the maximum permitted size. To put it formally the expansion procedure involves the following

$$\text{if class}(B_j) = \begin{cases} d_h & \Rightarrow \text{test if } B_j \text{ satisfies the max imum size constraint} \\ else & \Rightarrow \text{take another } B_j \end{cases} \quad (4)$$

with the size constraint in Eq. (4) defined as

$$\forall_{i=1,\dots,n}(\max(w_{ji}, x_{hi}^u) - \min(\nu_{ji}, x_{hi}^l)) \leqslant \Theta \quad (5)$$

If expansion can be accomplished then the hyperbox min and max points are updated as

$$\nu_{ji} = \min(\nu_{ji}, x_{hi}^l), \text{ for each } i = 1, \dots, n$$

$$w_{ji} = \max(w_{ji}, x_{hi}^u), \text{ for each } i = 1, \dots, n$$

The parameter Θ can either be a scalar, as suggested in [7], or a vector defining different maximum hyperbox sizes in different dimensions [4]. It can be shown that the latter can result in fewer hyperboxes defining each pattern class but requires some a-priori knowledge about the topology of individual class sets or multiple presentations of data to facilitate adaptation.

Overlap test: The expansion of the hyperboxes can produce hyperbox overlap. The overlap of hyperboxes that have the same class labels does not present any problem but the overlap of hyperboxes with different class labels must be prevented since it would create ambiguous classification. The test adopted in [7] and [4] adopts the principle of minimal adjustment, where only the smallest overlap for one dimension is adjusted to resolve the overlap. This involves consideration of four cases for each dimension

Case 1: $\nu_{ji} < \nu_{ki} < w_{ji} < w_{ki}$

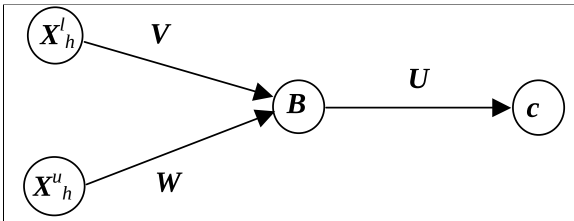
Case 2: $\nu_{ki} < \nu_{ji} < w_{ki} < w_{ji}$

Case 3: $\nu_{ji} < \nu_{ki} < w_{ki} < w_{ji}$

Case 4: $\nu_{ki} < \nu_{ji} < w_{ji} < w_{ki}$

The minimum value of overlap is remembered together with the index i of the dimension, which is stored as variable Δ. The procedure continues until no overlap is found for one of the dimensions (in which case there is no need for subsequent hyperbox contraction) or all dimensions have been tested.

Hyperbox contraction: The minimum overlap identified in the previous step provides basis for the im-



$\boldsymbol{X}^l_h = [x^l_{h1}, \ldots, x^l_{hn}]$, $\boldsymbol{X}^u_h = [x^u_{h1}, \ldots, x^u_{hn}]$

$\boldsymbol{V}=[V_1,\ldots,V_m]$, $\boldsymbol{W}=[W_1,\ldots,W_m]$

$\boldsymbol{B}=[B_1,\ldots,B_m]$

$\boldsymbol{U}=[u_{11},\ldots,u_{1p};\ldots;\ u_{m1},\ldots,u_{mp}]$

$\boldsymbol{c}=[\ c_1,\ldots,c_p]$

Fig. 2. The three-layer neural network implementation of the GFMM algorithm.

plementation of the contraction procedure. Depending on which case has been identified the contraction is implemented as follows:

$$\text{Case 1: } \nu^{new}_{k\Delta} = w^{new}_{j\Delta} = \frac{\nu^{old}_{k\Delta} + w^{old}_{j\Delta}}{2}$$

$$\text{or alternatively } (w^{new}_{j\Delta} = \nu^{old}_{k\Delta})$$

$$\text{Case 2: } \nu^{new}_{j\Delta} = w^{new}_{k\Delta} = \frac{\nu^{old}_{j\Delta} + w^{old}_{k\Delta}}{2}$$

$$\text{or alternatively } (\nu^{new}_{j\Delta} = w^{old}_{k\Delta})$$

$$\text{Case 3: } if\ w_{k\Delta} - \nu_{j\Delta} \leqslant w_{j\Delta} - \nu_{k\Delta}$$

$$then\ \nu^{new}_{j\Delta} = w^{old}_{k\Delta} \text{ otherwise } w^{new}_{j\Delta} = \nu^{old}_{k\Delta}$$

$$\text{Case 4: } if\ w_{k\Delta} - \nu_{j\Delta} \leqslant w_{j\Delta} - \nu_{k\Delta}$$

$$then\ w^{new}_{k\Delta} = \nu^{old}_{j\Delta} \text{ otherwise } \nu^{new}_{k\Delta} = w^{old}_{j\Delta}$$

The above three steps of the fuzzy Min-Max classification can be expressed as training of a three-layer neural network. The network, represented in Fig. 2, has a simple feed-forward structure and grows adaptively according to the demands of the classification problem. The input layer has $2 * n$ processing elements, the first n elements deal with the min point of the input hyperbox and the second n elements deal with the max point of the input hyperbox $X_h = [x^1_h\ X^u_h]$. Each second-layer node represents a hyperbox fuzzy set where the connections of the first and second layers are the min-max points of the hyperbox including the given pattern and the transfer function is the hyperbox membership function. The connections are adjusted using the expansion, overlap test, ontraction sequence described above. Note that the min points matrix V is modified only by the vector of lower bounds X^1_h of the input pattern and the max points matrix W is adjusted in response to the vector of upper bounds.

The connections between the second- and third-layer nodes are binary values. They are stored in matrix U. The elements of U are defined as follows:

$$u_{jk} = \begin{cases} 1 & if\ B_j \text{ is a hyperbox for class } c_k \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where B_j is the j_{th} second-layer node and c_k is the kth 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

$$c_k = \max_{j=1}^{m} B_j u_{jk} \quad (7)$$

for each of the p third-layer nodes. The outputs of the class layer nodes can be fuzzy when calculated using expression Eq. (7), or crisp when a value of one is assigned to the node with the largest ck and zero to the other nodes.

3. Inherent limitations of the fuzzy Min-Max classification

Training of the Min-Max 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, after expansion is accomplished it is necessary to perform an *overlap test* since it is possible that the expansion resulted 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 expanded hyperbox does not require any further adjustment 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 one or several overlapping coordinates and the consequent adjustment of the min- and max-points of the overlapping hyperboxes. 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 in the same time retaining

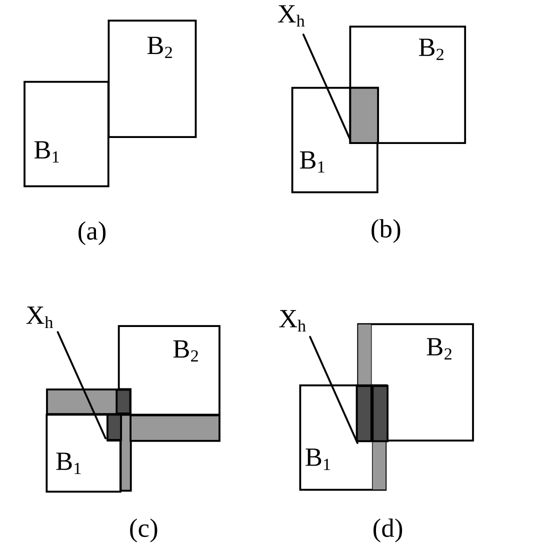


Fig. 3. Training of the fuzzy Min-Max neural network. (a) Hyperboxes belonging to two different classes $class(B_1) \neq class(B_2)$; (b) Inclusion of pattern $X_h, class(B_2)$ in B_2 implying overlap with B_1; (c) Contraction of B_1 and B_2 with adjustment along two coordinates; (d) Contraction of B_1 and B_2 with adjustment along one coordinate.

some of the contentious part of the pattern space in each of the hyperboxes. This is illustrated in Fig. 3.

By inspecting Fig. 3 it is clear that the contraction step of the fuzzy Min-Max network training resolves only part of the problem created by the expansion of the hyperbox B_2. Although the hyperboxes B_1 and B_2 no longer overlap after the contraction has been completed Figs 3(c) and 3(d), some part of the original hyperbox B1 remains included in and similarly some part of the hyperbox B_2 remains included in the contracted B_1. The degree of this residual inclusion depends on the contraction method that is chosen but it is never completely eliminated. Incidentally, it is worth noting that the intuitive approach proposed in [7] of subdividing overlapping hyperboxes along a single coordinate with the smallest overlap does produce worse residual inclusion problem than the alternative subdivision along all overlapping coordinates (compare Figs 3(c) and 3(d)).

Another problem inherent to the contraction procedure is that it unnecessarily eliminates parts of the original hyperboxes. These eliminated portions are marked in Fig. 3 with diagonal pattern lines. The elimination of these parts of hyperboxes implies that the contribution to the training of the Min-Max neural network of the data contained in these areas is nullified. If the neural network training involves only one pass through the data, then this is an irreversible loss that demonstrates itself in a degraded classification performance.

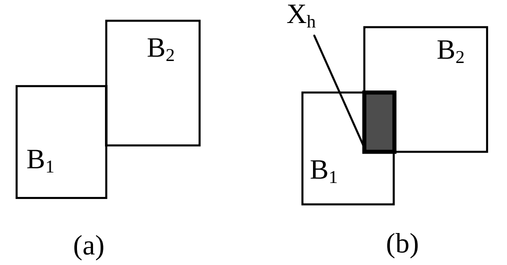


Fig. 4. The concept of the exclusion/inclusion fuzzy hyperboxes. (a) Hyperboxes belonging to two different classes $class$ (B_1) $class$ (B_2); (b) Inclusion of pattern $Xh, class(B_2)$ in B_2 implying overlap with B_1 and consequent identification of the exclusion hyperbox.

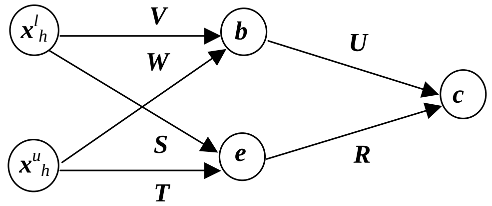
The problem can be somewhat alleviated by allowing multiple presentations of data in the training process, as in [4], or reducing the maximum size of hyperboxes. In either case the result is that additional hyperboxes are created to cover the eliminated portions of the original hyperboxes. Unfortunately, the increased number of hyperboxes reduces the interpretability of classification so that there is a limit as to how far this problem can be resolved in the context of the standard Min-Max expansion/contraction procedure.

Finally, it is worth noting that the training pattern $X_h,\ class(B_2)$ continues to be misclassified in spite of the contraction of the hyperboxes. This means that a 100% correct classification rate is not always possible even with the multiple-pass Min-Max neural network training.

4. 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 Fig. 4. The original hyperbox b_1 and the expanded hyperbox b_2 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.

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



$\boldsymbol{x}^l_h = [x^l_{h1}, \ldots, x^l_{hn}]$, $\boldsymbol{x}^u_h = [x^u_{h1}, \ldots, x^u_{hn}]$

$\boldsymbol{V}=[v_1,\ldots,v_m]$, $\boldsymbol{W}=[w_1,\ldots,w_m]$, $\boldsymbol{S}=[s_1,\ldots,s_m]$, $\boldsymbol{T}=[t_1,\ldots,t_m]$

$\boldsymbol{b}=[b_1,\ldots,b_m]$, $\boldsymbol{e}=[e_1,\ldots,e_q]$

$\boldsymbol{U}=[u_{10},\ldots, u_{1p}, u_{1(p+1)};\ldots; u_{m0},\ldots,u_{mp}, u_{m(p+1)}]$

$\boldsymbol{R}=[r_{10},\ldots, r_{1p}, r_{1(p+1)};\ldots; r_{q0},\ldots,r_{qp}, r_{q(p+1)}]$

$\boldsymbol{c}=[c_1,\ldots,c_p,c_{p+1}]$

Fig. 5. Exclusion/Inclusion Fuzzy Classification Network.

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:

$$r_{lk} = \begin{cases} 1\ if\ e_l \text{ overlapped hyperbox of lass } c_k \\ \quad \text{and } 1 < k < p \\ 1 \quad\quad if k = p+1 \\ 0 \quad\quad \text{otherwise} \end{cases} \quad (8)$$

Note that the third layer has $p + 1$ nodes $[c_1, \ldots, c_p, c_{p+1}]$ with the node cp+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:

$$c_k = \max_{k=1}^{p+1}(\max_{j=1}^{m} b_j u_{jk} - \max_{i=1}^{q} e_i r_{ik}) \quad (9)$$

The second component in Eq. (9) cancels out the contribution from the overlapping hyperboxes that belonged to different classes.

5. Numerical example

The EFC was applied to a number of synthetic data sets and demonstrated improvement over the GFMM and the original FMM [7]. As a representative example, we illustrate the performance of the network using the IRIS data-set from the Machine Learning Repository [11]. It is important to emphasise however that the use of a single specific data set does not detract from the essence of the topological argument that we are making in this paper; that is, that the pattern classes are covered more efficiently by the difference of fuzzy sets compared to the usual covering with the union of fuzzy sets. Using the IRIS data set we have trained the network on 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 [7]. The results are summarized in Table 1.

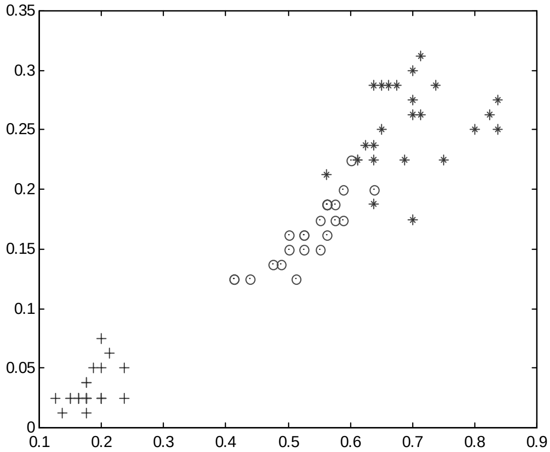


Fig. 6. IRIS data projected onto petal-length/petal-width two -dimensional space.

The results reported in Table 1 deserve some additional commentary. First we need to point out that the EFC algorithm would normally start without any constraint on the hyperbox size, i.e. $\Theta = 1$. However, the two other algorithms that we intended to compare EFC to do require precise control of the maximum hyperbox size. So, in the interest of comparability of the results we have run the EFC algorithm with Θ equal to 0.03, 0.06, 0.2 and 0.4.

The detailed results obtained with EFC for other values of the parameter Θ (0.1, 0.25, 0.35 and 0.45) are illustrated in Figs 6–10. Figure 6 shows the projection of the IRIS data onto a two-dimensional space of petal-length/petal-width. Subsequent figures show the effect of the gradual increase of the value of the maximum hyperbox size parameter Θ. Although it is clear that for $\Theta = 0.10$ (Fig. 7) the covering of the data with hyperboxes is more accurate than for $\Theta = 0.45$ (Fig. 10), we argue that this is achieved at a too great expense of reduced interpretability of the classification. The large number of rules, implied by the individual hyperboxes, is clearly counterproductive. From the viewpoint of the interpretability of classification the result illustrated in

Table 1
Comparison of performance of FMM, GFMM and EFC

Performance criterion	FMM [5]	GFMM [3]	EFC
Correct classification rate (range)	97.33–92%	100–92%	100–97%
Number of hyperboxes (max.size 0.03)	56	49	34
Number of hyperboxes (max. size 0.06)	32	29	18
Number of hyperboxes (max. size 0.20)	16	12	7
Number of hyperboxes (max. size 0.40)	16	12	4*

*the smallest number of classes; the number is not affected by the increase of the maximum size of hyperbox Θ

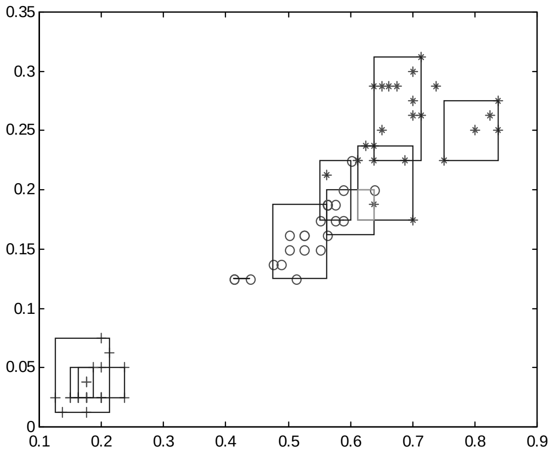


Fig. 7. Exclusion/inclusion hyperboxes evaluated for $\theta = 0.10$.

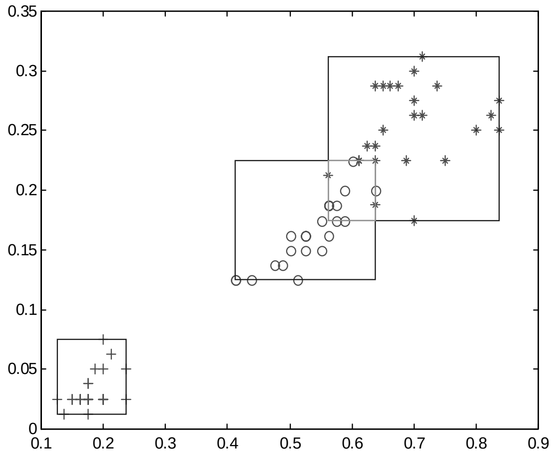


Fig. 9. Exclusion/inclusion hyperboxes evaluated for $\theta = 0.35$.

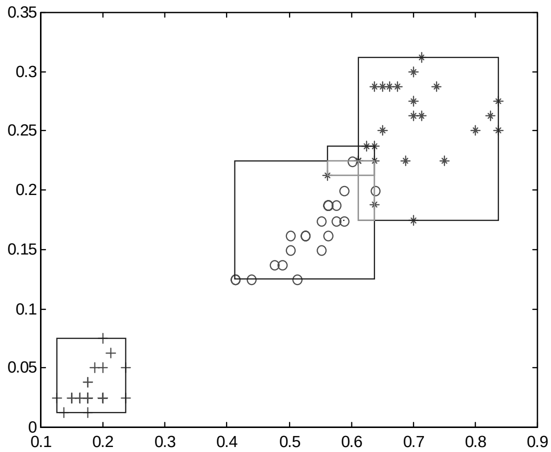


Fig. 8. Exclusion/inclusion hyperboxes evaluated for θ=0.25.

Fig. 10, is much preferred. Also, by comparing Figs 9 and 10, we note that for large Θ the result of classification is no longer dependent on the value of the parameter but is exclusively defined by the data itself. This in it self is a very desirable feature of the proposed algorithm.

Another point worth emphasizing is that the number of classes identified by the EFC is $p + 1$ where p is the number of classes identified by the FMM and GFMM. This implies that the calculation of the "classification rate" is not identical in all three cases. We have taken the view that the exclusion hyperbox(es) offer a positive identification of the patterns that are ambiguous. In this sense the fact of having some test data fall into the exclusion hyperbox is not considered a misclassification. Clearly, this has an effect of improving the classification rate of the EFC with respect to the other two methods. However, to do otherwise and to report all data falling into the exclusion hyperboxes as misclassified would be also misleading since we already have a knowledge about the nature of the exclusion hyperbox and it would effectively make no use of the "$p + 1^{st}$" pattern class.

Of course we do need to balance the assessment of the EFC algorithm by highlighting the importance of the ratio of the volumes of the exclusion and inclusion hyperbox sets. If this ratio is small (e.g. 1/35 in the case of the IRIS dataset) the classification results are very good. However, if the ratio increases significantly, the classification is likely to return a large proportion of patterns as belonging to the "exclusion" class. This

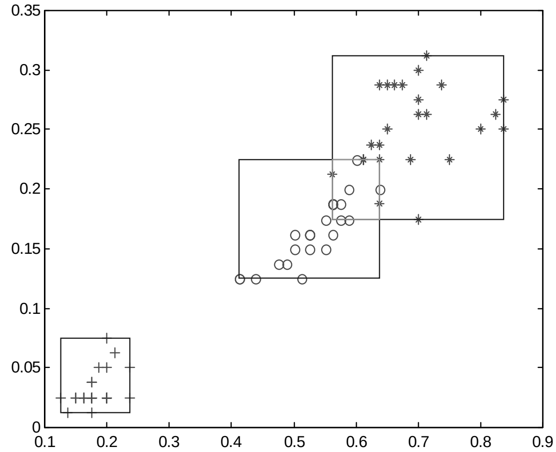


Fig. 10. Exclusion/inclusion hyperboxes evaluated for $\theta = 0.45$..

in itself offers a constructive advice on the reduction of the maximum size of hyperboxes.

6. Conclusion

The paper presented a new algorithm for pattern classification that is based on novel representation of class sets as a difference of two types of fuzzy sets (the union of hyperboxes belonging to the given class and the union of hyperboxes belonging to different classes). It has been shown that, compared to the standard hyperbox paving approaches, the proposed algorithm results in a more efficient generation of complex topologies that are necessary to describe the pattern classes. The consequence of the adoption of the exclusion/inclusion framework is a greater interpretability of the classification results (smaller number of hyperboxes needed to cover the data). It has been shown that in the proposed approach the size of the hyperboxes does not need to be pre-determined and is indeed defined by the data itself. This is a very beneficial feature as it frees the analyst from making arbitrary choices with regard to the parameters of the algorithm. The low misclassification rate and good interpretability of the results of the proposed algorithm is achieved at the expense of rejecting a proportion of patterns that fall into the exclusion hyperbox set. If this proportion is small the algorithm provides an optimum mix of good classifier features. However, if the exclusion set becomes comparable in size to the inclusion set the maximum size of hyperboxes needs to be reduced. This is analogous to the standard hyperbox paving approaches but unlike in the standard approaches we do not use the misclassifications rate (that is dependent on the test data set) but instead use the ratio of exclusion to inclusion hyperbox sets (evaluated with training data only) as an indicator of how small hyperboxes need to be.

A general point raised by this investigation is that of a benefit of a richer vocabulary of topological constructs in describing data sets in multi-dimensional pattern spaces.

Acknowledgments

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

References

[1] A. Bargiela and W. Pedrycz, *Granular clustering with partial supervision*, Proc European Simulation Multiconference, ESM2001, Prague, June 2001, 113–120.
[2] R.E. Bellman, R. Kalaba and L. Zadeh, Abstraction and pattern classification, *J. Math. Anal. Appl.* **13** (1966), 1–7.
[3] D. Dubois and H. Prade, Fuzzy relation equations and causal reasoning, *FSS* **75** (1995), 119–134.
[4] B. Gabrys and A. Bargiela, General fuzzy min-max neural network for clustering and classification, *IEEE Trans. Neural Networks* **11** (2000), 769–783.
[5] W. Pedrycz, Fuzzy neural networks with reference neurons as pattern classifiers, *IEEE Trans Neural Networks* **5** (1992), 770–775.
[6] W. Pedrycz and F. Gomide, *An Introduction to Fuzzy Sets*, Cambridge, MIT Press, Cambridge, MA, 1998.
[7] P.K. Simpson, Fuzzy min-max neural networks – Part1: Classification, *IEEE Trans. Neural Networks* **3**(5) (1992), 776–786.
[8] P.K. Simpson, Fuzzy min-max neural networks – Part 2: Clustering, *IEEE Trans Neural Networks* **4**(1) (1993), 32–45.
[9] L.A. Zadeh, Fuzzy sets, *Inform. And Control* **8** (1965), 189–200.
[10] L.A. Zadeh, 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.
[11] *Machine Learning Repository*, http://www.ics.uci.edu/mlearn /MLRepository.html