Accelerated learning algorithms of general fuzzy min-max neural network using a branch-and-bound-based hyperbox selection rule

Thanh Tung Khuat\textsuperscript{a,*}, Bogdan Gabrys\textsuperscript{a}

\textsuperscript{a}Advanced Analytics Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, NSW, Australia

Abstract

This paper proposes a method to accelerate the training process of general fuzzy min-max neural network. The purpose is to reduce the unsuitable hyperboxes selected as the potential candidates of the expansion step of existing hyperboxes to cover a new input pattern in the online learning algorithms or candidates of the hyperbox aggregation process in the agglomerative learning algorithms. Our proposed approach is based on the mathematical formulas to form a branch-and-bound solution aiming to remove the hyperboxes which are certain not to satisfy expansion or aggregation conditions, and in turn decreasing the training time of learning algorithms. The efficiency of the proposed method is assessed over a number of widely used data sets. The experimental results indicated the significant decrease in training time of proposed approach for both online and agglomerative learning algorithms. Notably, the training time of the online learning algorithms is reduced from 1.2 to 12 times when using the proposed method, while the agglomerative learning algorithms are accelerated from 7 to 37 times on average.

Keywords: General fuzzy min-max neural network, branch-and-bound-based hyperbox selection, online learning, agglomerative learning, accelerated learning algorithms

1. Introduction

General fuzzy min-max (GFMM) neural network (GFMMNN)\textsuperscript{[2]} is a generalization framework of fuzzy min-max neural network for classification\textsuperscript{[6]} and clustering\textsuperscript{[7]}. The GFMM model can handle both labelled and unlabelled data as well as crisp and fuzzy input samples in a single model. One of the remarkable characteristics of the GFMMNN is that it is able to explain the predictive results based on the rule sets extracted directly or indirectly from hyperboxes\textsuperscript{[4]}. This interpretable property of the GFMMNN is essential so that it can be used for high-stakes applications such as medical diagnosis, self-driving cars, and financial investment\textsuperscript{[5]}. Interpretability helps fuzzy min-max neural networks to overcome the black-box drawbacks of the traditional neural networks.

\textsuperscript{*}Corresponding author

Email addresses: thanhtung.khuat@student.uts.edu.au (Thanh Tung Khuat), bogdan.gabrys@uts.edu.au (Bogdan Gabrys)
There are two types of learning algorithms for the general fuzzy min-max neural network, i.e., incremental (online) learning (Onln-GFMM) [2] and agglomerative (batch) learning [1]. The online learning algorithm accommodates new input patterns by extending the current existing hyperboxes or creating a new hyperbox. In contrast, the agglomerative learning algorithm starts with all training samples and conducts a process of merging hyperboxes satisfying aggregation criteria to generate larger sized hyperboxes. To take advantages of the strong points of the online and agglomerative learning algorithms, a recent study proposed an improved version of online learning algorithm (IOL-GFMM) [3] to avoid the hyperbox contraction process which is more likely to cause the classification errors in the online learning algorithm.

However, all of the current learning algorithms for the general fuzzy min-max neural network have the same drawback in the selection of expandable or mergeable hyperbox candidates. The creation of a new hyperbox in the online learning algorithms only occurs when all existing hyperboxes with the same class as the input patterns cannot satisfy the expansion condition to cover the new input pattern. The expansion condition is the maximum hyperbox size and the non-overlap of hyperboxes representing different classes if using the IOL-GFMM. Similarly, in the agglomerative learning algorithm, the process of hyperbox merging only terminates if all pairs of hyperbox candidates are examined with regard to the aggregation criteria but the aggregation process cannot be performed. The aggregation conditions include maximum hyperbox size, minimum similarity value, and the non-overlap of hyperboxes with different classes. The consideration of expansion or merging conditions for all hyperbox candidates leads to a waste of time. Therefore, in this study, we provide a lower bound on the membership functions and similarity measures to reduce the considered hyperbox candidates for the expansion or merging process. This method contributes to decreasing the training time of the learning algorithms.

Our contributions in this paper can be summarized as follows:

- We propose and prove the lemmas to reduce significantly the considered hyperboxes for both online and batch learning algorithms for the general fuzzy min-max neural network. To the best of our knowledge, this is the first study tackling this issue.

- The effectiveness of the proposed method is assessed on widely used datasets. Experimental results confirmed the strong points of the method in decreasing the training time of the algorithms.

The rest of this paper is structured as follows. Section 2 presents an overall architecture and learning algorithms of general fuzzy min-max neural network. Section 3 shows our proposed method. The experimental results and discussion are described in section 4. Section 5 concludes the main findings.
2. General fuzzy min-max neural network and learning algorithms

2.1. An overall architecture

General fuzzy min-max neural network contains three layers as shown in Fig. 1. The input layer includes $2n$ nodes, where $n$ is the number of dimensions of the input pattern. The first $n$ nodes are the lower bounds of the input, while the remaining $n$ nodes correspond to the upper bounds. These input nodes are connected to $m$ hyperbox nodes in the hidden layer, in which the connection weights of the lower bound input nodes are stored in a matrix $V$ and the connection weights between upper bound input nodes and hyperbox nodes are saved in a matrix $W$. These weights correspond to minimum points and maximum points of hyperboxes, and their values are adjusted during the training process. Each hyperbox node $B_i$ is associated with an activation function, which is also known as the membership function defined by Eq. (1).

$$b_i(X) = \min_{j=1}^{n}(\min([1 - f(x_j^u - w_j, \gamma_j]), [1 - f(v_j - x_j^l, \gamma_j)]))$$  \hspace{1cm} (1)

where $f(x, \gamma) = \begin{cases} 
1, & \text{if } z\gamma > 1 \\
z\gamma, & \text{if } 0 \leq z\gamma \leq 1 \\
0, & \text{if } z\gamma < 0 
\end{cases}$ is the ramp function, $\gamma = [\gamma_1, \ldots, \gamma_n]$ is a sensitivity parameter describing the speed of decreasing of the membership function, and $X = [X^l, X^u]$ is an input pattern with lower bounds $X^l$ and upper bounds $X^u$. 

Figure 1: An architecture of general fuzzy min-max neural network
Each hyperbox $B_i$ in the hidden layer is connected to each output (class) node $c_j$ by a binary-valued parameter $u_{ij}$ computed using Eq. (2). There are $p + 1$ output nodes corresponding to $p$ classes, where node $c_0$ is linked to all unlabelled hyperboxes in the hidden layer. The transfer function of $c_j$ is determined by the maximum membership value of all hyperboxes with same class as $c_j$ and is shown in Eq. (3).

$$u_{ij} = \begin{cases} 
1, & \text{if hyperbox } B_i \text{ represents class } c_j \\
0, & \text{otherwise}
\end{cases} \quad (2)$$

$$c_j = \max_{i=1}^{m} b_i \cdot u_{ij} \quad \quad \quad (3)$$

where $m$ is the number of hyperboxes in the middle layer. The output of each class node can be a fuzzy value calculated directly from Eq. (3) or a crisp value if the node associated with the highest membership value gets the value of one, and the others are assigned zero values [2].

Although the GFMM model can be applied for labelled and unlabelled datasets, this paper only considers the classification problems. Therefore, the learning algorithms in the next sections are presented for labeled training data.

2.2. Online learning algorithm

The incremental (online) learning algorithm, proposed in [2], adjusts the size of existing hyperboxes or create new hyperboxes to accommodate new coming input patterns. There are three main steps in the algorithms including hyperbox expansion/creation, hyperbox overlap test, and hyperbox contraction. The pseudo code of the original online learning algorithm is given in Algorithm 1.

Assuming that each input pattern is represented in the form of $X = [X^l, X^u, l_X]$, where $l_X$ is a class label and $X^l$ and $X^u$ are lower and upper bounds. The online learning algorithm first selects all existing hyperboxes with the same class as $l_X$. After that, the algorithm performs the computation of the membership values between these selected hyperboxes and the input pattern $X$, and then these membership values are sorted in a descending order (lines 8-9). Next, the algorithm traverses in turn each hyperbox $B_i$ in the list of selected hyperboxes starting from the hyperbox with the maximum membership value to choose a hyperbox candidate aiming to expand and cover the input pattern. This process terminates when there is a hyperbox satisfying the expansion condition or the membership value is one (lines 12-28). Otherwise, a new hyperbox will be created with the same co-ordinates and label as the input pattern (lines 29-31). The expansion condition relates to the maximum hyperbox size threshold in each dimension as shown in Eq. (4).

$$\max(w_{ij}, x^u_j) - \min(v_{ij}, x^l_j) \leq \theta, \quad \forall j \in [1, n] \quad (4)$$

If this criterion is met, the hyperbox $B_i$ is extended to accommodate the input pattern $X$ using Eqs. (5) and
Algorithm 1 The original online learning algorithm

Require:
- \( \theta \): The maximum hyperbox size threshold
- \( \gamma \): The speed of decreasing of the membership function

Ensure:
A list \( \mathcal{H} \) of hyperboxes with minimum-maximum values and classes

1: Initialize an empty list of hyperboxes: min-max values \( V = W = \emptyset \), hyperbox classes: \( L = \emptyset \)
2: for each input pattern \( X = [x^l, x^u, l_X] \) do
3: \( n \leftarrow \) The number of dimensions of \( X \)
4: if \( V = \emptyset \) then
5: \( V \leftarrow x^l; \quad W \leftarrow x^u; \quad L \leftarrow l_X \)
6: else
7: \( \mathcal{H}_1 = [V_1, W_1, L_1] \leftarrow \) Find hyperboxes in \( \mathcal{H} = [V, W, L] \) representing the same class as \( l_X \)
8: \( \mathcal{M} \leftarrow \text{ComputeMembershipValue}(X, V_1, W_1, L_1) \)
9: \( \mathcal{H}_d \leftarrow \text{SortByDescending}(\mathcal{H}_1, \mathcal{M}(\mathcal{H}_1)) \)
10: \( \mathcal{T}_1 \leftarrow \mathcal{H} \setminus \mathcal{H}_1 \)
11: \( \text{flag} \leftarrow \text{false} \)
12: for each \( h = [V_h, W_h, l_h] \in \mathcal{H}_d \) do
13: \( \text{if } \mathcal{M}(h) = 1 \text{ then} \)
14: \( \text{flag} = \text{true} \)
15: \( \text{break} \)
16: end if
17: \( \text{if } \max(w_{hj}, x_{uj}) - \min(v_{hj}, x_{lj}) \leq \theta, \forall j \in [1, n] \text{ then} \)
18: \( W_h^t \leftarrow \max(W_h, X^u); \quad V_h^t \leftarrow \min(V_h, X^l) \)
19: for each hyperbox \( [V_i, W_i, l_i] \in \mathcal{T}_1 \) do
20: \( \text{isOver} \leftarrow \text{OverlapTest}(V_h^t, W_h^t, V_i, W_i) \)
21: \( \text{if } \text{isOver} = \text{true} \text{ then} \)
22: \( \text{DoContraction}(V_h^t, W_h^t, V_i, W_i) \)
23: end if
24: end for
25: \( \text{flag} = \text{true} \)
26: \( \text{break} \)
27: end if
28: end for
29: \( \text{if } \text{flag} = \text{false} \text{ then} \)
30: \( V \leftarrow V \cup X^l; \quad W \leftarrow W \cup X^u; \quad L \leftarrow L \cup l_X \)
31: end if
32: end if
33: end for
34: return \( \mathcal{H} = [V, W, L] \)

\[
\begin{align*}
v_{ij}^{\text{new}} &= \min(v_{ij}^{\text{old}}, x_{lj}) \quad (5) \\
w_{ij}^{\text{new}} &= \max(w_{ij}^{\text{old}}, x_{uj}), \quad \forall j \in [1, n] \quad (6)
\end{align*}
\]

If the expansion step of the hyperbox candidate is carried out, the extended hyperbox \( B_i \) is verified for the overlap with the hyperboxes \( B_k \) representing other classes. For each dimension \( j \), four following conditions
are checked (initially $\delta^{old} = 1$):

- $v_{ij} < v_{kj} < w_{ij} < w_{kj}$: $\delta^{new} = \min(w_{ij} - v_{kj}, \delta^{old})$
- $v_{kj} < v_{ij} < w_{kj} < w_{ij}$: $\delta^{new} = \min(w_{kj} - v_{ij}, \delta^{old})$
- $v_{ij} < v_{kj} \leq w_{kj} < w_{ij}$: $\delta^{new} = \min(\min(w_{kj} - v_{ij}, w_{ij} - v_{kj}), \delta^{old})$
- $v_{kj} < v_{ij} \leq w_{ij} < w_{kj}$: $\delta^{new} = \min(\min(w_{ij} - v_{kj}, w_{kj} - v_{ij}), \delta^{old})$

If $\delta^{new} < \delta^{old}$, then we set $\Delta = i$ and $\delta^{old} = \delta^{new}$ to show an overlapping area on the $\Delta$th dimension, and the testing procedure is repeated for the next dimension. In contrast, there is no overlap region between two considered hyperboxes, and the hyperbox contraction step will not be performed ($\Delta = -1$). If $\Delta \neq -1$, the contraction procedure is applied on the $\Delta$th dimension to remove the overlapping area between two hyperboxes. The overlapping region is eliminated by tuning the value of the dimension with the smallest overlap. If $\Delta > 0$, this dimension is adjusted according to the four following cases:

Case 1: $v_{i\Delta} < v_{k\Delta} < w_{i\Delta} < w_{k\Delta}$: $v_{i\Delta}^{new} = w_{i\Delta}^{new} = \frac{w_{i\Delta}^{old} + v_{i\Delta}^{old}}{2}$

Case 2: $v_{k\Delta} < v_{i\Delta} < w_{k\Delta} < w_{i\Delta}$: $v_{k\Delta}^{new} = w_{k\Delta}^{new} = \frac{w_{k\Delta}^{old} + v_{k\Delta}^{old}}{2}$

Case 3: $v_{i\Delta} < v_{k\Delta} \leq w_{k\Delta} < w_{i\Delta}$:

\[
\begin{align*}
v_{i\Delta}^{new} & = w_{i\Delta}^{old}, \quad \text{if } w_{k\Delta} - v_{i\Delta} \leq w_{i\Delta} - v_{k\Delta} \\
v_{i\Delta}^{new} & = v_{i\Delta}^{old}, \quad \text{if } w_{k\Delta} - v_{i\Delta} > w_{i\Delta} - v_{k\Delta}
\end{align*}
\]

Case 4: $v_{k\Delta} < v_{i\Delta} \leq w_{i\Delta} < w_{k\Delta}$:

\[
\begin{align*}
w_{k\Delta}^{new} & = v_{i\Delta}^{old}, \quad \text{if } w_{k\Delta} - v_{i\Delta} \leq w_{i\Delta} - v_{k\Delta} \\
v_{k\Delta}^{new} & = v_{i\Delta}^{old}, \quad \text{if } w_{k\Delta} - v_{i\Delta} > w_{i\Delta} - v_{k\Delta}
\end{align*}
\]

2.3. Agglomerative learning algorithm

The online learning algorithm creates or adjusts the size of hyperboxes whenever an input sample comes in the network. Therefore, its performance depends on the data presentation order. In [1], Gabrys proposed an agglomerative learning algorithm based on the full similarity matrix (AGGLO-SM) to reduce the impact of the data presentation order on the accuracy of the learning algorithm. In contrast to the online learning algorithm, the AGGLO-SM algorithm for the classification problems starts with all of the training samples. The idea is to merge hyperboxes with the same class, possessing the similarity values larger than a given threshold, and not generating the overlapping areas with existing hyperboxes representing other classes. The main steps of the AGGLO-SM algorithm are shown in Algorithm 2.

Firstly, the algorithm initializes a matrix $V$ of minimum points and a matrix $W$ of maximum points using the lower bounds $X^l$ and upper bounds $X^u$ of all training samples. Next, the algorithm performs a
Algorithm 2 The agglomerative algorithm with full similarity matrix - AGGLO-SM

**Require:**
- \( X = [X^1, X^n] \): A list of training features
- \( L \): A vector of pattern classes
- \( \theta \): The maximum hyperbox size threshold
- \( \gamma \): The speed of decreasing of the membership function

**Ensure:**
A list \( H \) of hyperboxes with minimum-maximum values and classes

1: Initialize a list of hyperboxes: min-max values \( V = X^l \), \( W = X^u \), hyperbox classes: \( L = L \)
2: \( loop \leftarrow true; n \leftarrow \) the number of features of \( X \)
3: while \( loop = true \) do
4: \( loop \leftarrow false \)
5: \( S \leftarrow \text{ComputeOrUpdateSimilarityValPairWithinEachClass}(V, W, L) \)
6: \( S \leftarrow S \setminus \{s \in S | s < \sigma \} \)
7: \( I, K, S \leftarrow \text{SortByDescending}(S, V, W, L) \)
8: for each \( [i, k, s] \in [I, K, S] \) do
9: if \( \max(w_{ij}, w_{kj}) - \min(v_{ij}, v_{kj}) \leq \theta, \forall j \in [1, n] \) then
10: \( W_t \leftarrow \max(W_i, W_k); \ V_t \leftarrow \min(V_i, V_k) \)
11: \( H_1 \leftarrow \) A list of hyperboxes with classes different from \( \ell_i \in L \)
12: \( \text{isOver} \leftarrow \text{IsOverlap}(V_t, W_t, H_1) \)
13: if \( \text{isOver} = false \) then
14: \( loop \leftarrow true \)
15: \( V_i \leftarrow V_t; \ W_i \leftarrow W_t \)
16: \( V \leftarrow V \setminus V_k; \ W \leftarrow W \setminus W_k; \ L \leftarrow L \setminus L_k \)
17: break
19: end if
20: end if
21: end for
22: return \( H = [V, W, L] \)

repeated training process of aggregating hyperboxes starting from the computation of a similarity matrix of hyperboxes for each class. There are three measures possible to be used to find the similarity value of each pair of hyperboxes \( B_i \) and \( B_k \) as follows:

- The first similarity measure is based on maximum points or minimum points of two hyperboxes. For simplifying, we call this measure as “middle distance” in this paper, though the similarity measures are not distance measures:
  \[ s_{ik} = s(B_i, B_k) = \min_{j=1}^n (\min(1 \cdot f(w_{kj} - w_{ij}, \gamma_j), 1 - f(v_{ij} - v_{kj}, \gamma_j))) \]
  It can be seen that \( s_{ik} \neq s_{ki} \), thus the similarity value between \( B_i \) and \( B_k \) may receive the minimum or maximum value between \( s_{ik} \) and \( s_{ki} \). If the maximum value is used, this measure is called “mid-max distance”; otherwise, the name “mid-min distance” is used.

- The second similarity measure employs the smallest gap between two hyperboxes \( B_i \) and \( B_k \), namely
“shortest distance” in this paper:
\[
\tilde{s}_{ik} = \tilde{s}(B_i, B_k) = \min_{j=1}^{n}(\min(1 - f(v_{kj} - w_{ij}, \gamma_j), 1 - f(v_{ij} - w_{kj}, \gamma_j))) \tag{7}
\]

- The third similarity measure is based on the longest possible distance between two hyperboxes \(B_i\) and \(B_k\), called “longest distance” for short, defined as follows:
\[
\hat{s}_{ik} = \hat{s}(B_i, B_k) = \min_{j=1}^{n}(\min(1 - f(w_{kj} - v_{ij}, \gamma_j), 1 - f(w_{ij} - v_{kj}, \gamma_j)))
\]

We can observe that both \(\tilde{s}_{ik}\) and \(\hat{s}_{ik}\) possess the symmetrical property.

From the similarity matrix of hyperboxes with the same class, the algorithm will merge sequentially hyperboxes by seeking for a pair of hyperboxes with the maximum similarity value. It is noted that the algorithm only considers pairs of hyperboxes with similarity values larger than or equal to a minimum similarity threshold (\(\sigma\)): \(s_{ih} \geq \sigma\) (line 6 - Algorithm 2). Assuming that these two hyperboxes are \(B_i\) and \(B_k\), the following conditions will be checked before aggregating:

(a) Maximum hyperbox size:
\[
\max(w_{ij}, w_{kj}) - \min(v_{ij}, v_{kj}) \leq \theta, \quad \forall j \in [1, n]
\]

(b) Overlap test. Newly aggregated hyperbox from \(B_i\) and \(B_k\) does not overlap with any existing hyperboxes belonging to other classes. The overlap checking conditions between two hyperboxes are shown in subsection 2.2. If any overlapping area exists, another pair of hyperboxes is selected.

If all above constraints are met, the hyperbox aggregation process is carried out as follows:

(a) Updating the coordinates of \(B_i\) so that \(B_i\) represents the coordinates of the merged hyperbox (line 15).

(b) Removing \(B_k\) from the current set of hyperboxes (line 16) and updating the similarity matrix.

This training process is iterated until there are no pairs of hyperboxes to aggregate.

Training process of the AGGLO-SM algorithm takes a very long time to complete, especially for massive datasets, due to the fact that we need to compute and sort the similarity matrix for all pairs of hyperboxes. To lower the training time of the AGGLO-SM algorithm, Gabrys [1] introduced the second agglomerative algorithm (AGGLO-2) removing the usage of the full similarity matrix when selecting and merging hyperboxes. The main steps of the AGGLO-2 are shown in Algorithm 3.

The AGGLO-2 algorithm traverses and selects in turn each hyperbox in the current list of hyperboxes to perform the hyperbox merging process. For the first selected hyperbox candidate \(B_i\), the similarity values of \(B_i\) and the remaining hyperboxes with the same class as \(B_i\) are computed. The hyperbox \(B_k\) with the highest similarity value is chosen as the second candidate for the aggregation. The aggregation constraints and the aggregation process of \(B_i\) and \(B_k\) are the same as in the AGGLO-SM algorithm. If \(B_i\) and \(B_k\) do not meet
the constraints, the hyperbox with the second largest similarity value is selected, and the above checking and merging steps are repeated until the agglomeration happens. The learning algorithm stops when no pair of hyperboxes can be aggregated (the variable \( \text{loop} = \text{false} \) in Algorithm 3).

**Algorithm 3** The agglomerative algorithm version two - AGGLO-2

Require:
- \( X = [X^l, X^u] \): A list of training features
- \( L \): A vector of pattern classes
- \( \theta \): The maximum hyperbox size threshold
- \( \gamma \): The speed of decreasing of the membership function

Ensure:
- A list \( H \) of hyperboxes with minimum-maximum values and classes

1: Initialize a list of hyperboxes: \( \min\-\max\) values \( V = X^l, W = X^u \), hyperbox classes: \( L = L \)
2: \( \text{loop} \leftarrow \text{true}; n \leftarrow \text{the number of features of} X \)
3: while \( \text{loop} = \text{true} \) do
4: \( \text{loop} \leftarrow \text{false}; i \leftarrow 1 \)
5: while \( i \leq |L| \) do
6: \( H_1 = [V_i, W_i, L_i] \leftarrow \text{Find hyperboxes in} \ [V, W, L] \text{ representing the same class as} l_i \in L \)
7: \( S \leftarrow \text{ComputeOrUpdateSimilarityValPair}(V_i, W_i, H_1) \)
8: \( S \leftarrow S \{ s \in S | s < \sigma \} \)
9: \( K, S \leftarrow \text{SortByDescending}(S, V_i, W_i, L_i) \)
10: for each \([k, s] \in [K, S]\) do
11: if \( \max(w_{ij}, w_{kj}) - \min(v_{ij}, v_{kj}) \leq \theta, \forall j \in [1, n] \) then
12: \( W_t \leftarrow \max(W_i, W_k); V_t \leftarrow \min(V_i, V_k) \)
13: \( H_t \leftarrow \text{A list of hyperboxes with classes different from} l_i \in L \)
14: \( \text{isOver} \leftarrow \text{IsOverlap}(V_t, W_t, H_t) \)
15: if \( \text{isOver} = \text{false} \) then
16: \( \text{loop} \leftarrow \text{true} \)
17: \( V_i \leftarrow V_t; W_i \leftarrow W_t \)
18: \( V \leftarrow V \setminus V_t; W \leftarrow W \setminus W_t; L \leftarrow L \setminus L \)
19: if \( i > k \) then
20: \( i \leftarrow i - 1 \)
21: end if
22: break
23: end if
24: end if
25: end for
26: \( i \leftarrow i + 1 \)
27: end while
28: end while
29: return \( H = [V, W, L] \)

2.4. An improved online learning algorithm

The improved online learning algorithm (IOL-GFMM) is proposed in [3] to tackle the disadvantages of the original online learning algorithms related to the hyperbox contraction and equal membership value.
Similarly to the agglomerative learning algorithm, the IOL-GFMM algorithm prevents the overlap between the expanded hyperbox and any existing hyperboxes representing other classes, so it removes the contraction step from the learning algorithm. The learning process contains only two main steps, i.e., hyperbox expansion/creation and overlap test. The details of the IOL-GFMM are given in Algorithm 4.

**Algorithm 4 The IOL-GFMM algorithm**

**Require:**
- \( \theta \): The maximum hyperbox size threshold
- \( \gamma \): The speed of decreasing of the membership function

**Ensure:**
A list \( H \) of hyperboxes with minimum-maximum values and classes

1: Initialize an empty list of hyperboxes: min-max values \( V = W = \emptyset \), hyperbox classes: \( L = \emptyset \)
2: for each input pattern \( X = [X^l, X^u, l_X] \) do
3: \( n \leftarrow \) The number of dimensions of \( X \)
4: if \( V = \emptyset \) then
5: \( V \leftarrow X^l; \; W \leftarrow X^u; \; L \leftarrow l_X \)
6: else
7: \( H_1 = [V_1, W_1, L_1] \leftarrow \) Find hyperboxes in \( [V, W, L] \) representing the same class as \( l_X \)
8: \( M \leftarrow \text{ComputeMembershipValue}(X, V_1, W_1, L_1) \)
9: \( H_d \leftarrow \text{SortByDescending}(H_1, M(H_1)) \)
10: \( \bar{H}_1 \leftarrow H \setminus H_1 \)
11: \( \text{flag} \leftarrow \) false
12: for each \( h = [V_h, W_h, l_h] \in H_d \) do
13: if \( M(h) = 1 \) then
14: \( \text{flag} = \) true
15: break
16: end if
17: if \( \max(w_{h,j}, x_j^u) - \min(v_{h,j}, x_j^l) \leq \theta, \forall j \in [1, n] \) then
18: \( W'_h \leftarrow \max(W_h, X^u); \; V'_h \leftarrow \min(V_h, X^l) \)
19: \( \text{isOver} \leftarrow \text{IsOverlap}(W'_h, V'_h, \bar{H}_1) \)
20: if \( \text{isOver} = \) false then
21: \( V_h \leftarrow V'_h; \; W_h \leftarrow W'_h \)
22: \( \text{flag} \leftarrow \) true
23: Increase the number of samples contained in \( h \)
24: break
25: end if
26: end if
27: end if
28: if \( \text{flag} = \) false then
29: \( V \leftarrow V \cup X^l; \; W \leftarrow W \cup X^u; \; L \leftarrow L \cup l_X \)
30: end if
31: end if
32: end for
33: return \( H = [V, W, L] \)

**Expansion of hyperboxes.** When an input sample \( X = [X^l, X^u, l_X] \) comes to the network, the algorithm first filters all existing hyperboxes representing the same class as \( l_X \). Next, the membership values between \( X \)
and all selected hyperboxes are computed and sorted in a descending order (lines 8-9). If the maximum membership value is one, the learning process continues with another input sample (lines 13-16). Otherwise, the algorithm traverses in turn each hyperbox starting from the hyperbox with the maximum membership value to verify the expansion criteria. If all constraints are satisfied, the size of the selected hyperbox and the number of pattern included in that hyperbox are updated and the learning process continues with next input samples (lines 17-26). If none of the hyperbox candidates satisfies the conditions, a new hyperbox is generated to accommodate the input pattern and added to the current set of hyperboxes (lines 28-30). Two expansion criteria include the maximum hyperbox size shown in Eq. (4) and overlap. If the maximum hyperbox size constraint is satisfied, then the non-overlapping condition is verified as follows:

Overlap test. The overlap test is performed between the newly extended hyperbox and the hyperboxes belonging to other classes. If there is any overlapping regions occurring, the next hyperbox candidate is considered. Otherwise, the chosen hyperbox is updated by the new coordinates of the expanded hyperbox, and the learning steps continue with another input sample. The overlap test between each pair of hyperboxes is conducted in the same way as the steps shown in subsection 2.2.

Classification phase. In the classification phase, when an unseen sample $X$ comes to the network, the membership degrees between $X$ and all existing hyperboxes in the model are calculated. The input $X$ will be classified to the class of the hyperbox with the maximum membership value. In the original online learning algorithm, if there are at least two hyperboxes with the same maximum membership value but different classes, the algorithm will select the predicted class randomly. In contrast, in the IOL-GFMM algorithm, if many hyperboxes belonging to $K$ different classes output the same maximum membership degree ($b_{\text{win}}$), we need to deploy an additional criterion to specify the suitable class for $X$. If $b_{\text{win}} = 1$ and $\exists i : n_i = 1$, then the class of $X$ is the class of $B_i$. Otherwise, the predicted class of $X$ is the class $c_k$ with the highest value of $P(c_k|X)$ defined by:

$$P(c_k|X) = \frac{\sum_{j \in I_{\text{win}}^k} n_j \cdot b_j}{\sum_{i \in I_{\text{win}}} n_i \cdot b_i} \quad (8)$$

where $k \in [1, K]$ and $I_{\text{win}} = \{i, \text{if } b_i = b_{\text{win}}\}$ contains the indexes of all hyperbox with the same maximum membership value, $I_{\text{win}}^k = \{j, \text{if } \text{class}(B_j) = c_k \text{ and } b_i = b_{\text{win}}\}$ is a subset of $I_{\text{win}}$ created by indexes of the $k^{th}$ class, and $n_i$ is the number of patterns included in the hyperbox $B_i$.

3. Proposed method

In order to alleviate the computational issues, we present an approach to training algorithms that drastically reduce the number of considered expandable hyperbox candidates by omitting hyperboxes certain not to satisfy the expansion or aggregation conditions.
3.1. Accelerated online learning algorithms

It is observed that in online learning algorithms of GFMMNN, a new hyperbox is only created to cover the input pattern if all hyperbox candidates cannot satisfy the conditions to be expanded for accommodating the new input pattern. However, in the current versions of the online learning algorithms, there is no way to reduce the considered hyperbox candidates. In this paper, therefore, we provide a lemma to narrow down the expandable hyperboxes during the training process. This solution is given in Lemma 1.

**Lemma 1.** When finding the candidates of expandable hyperboxes to cover an input pattern \( X \), we only need to consider the hyperboxes \( (h) \) with the same class as \( X \) and having a membership degree \( b_h(X) \) to the new input pattern satisfying: 
\[
b_h(X) \geq 1 - \theta \cdot \gamma_{\text{max}}, \text{ where } \gamma_{\text{max}} = \max_{j=1}^{n} \gamma_j; \gamma_j > 0
\]

**Proof.** See the proof in the appendix A

This lemma shows the relationship between the membership function and the maximum hyperbox size parameter if we keep the sensitivity parameter \( \gamma \) fixed. By using this lemma, we can reduce the number of hyperbox candidates considered for the expansion step based on their membership values. We can modify the Algorithm 1 into Algorithm 5 to accelerate the original online learning algorithm of GFMM model. The only change in this algorithm is that we use the proposed lemma to limit the number of hyperboxes considered for each input pattern. The other steps are the same as in the original version.

**Algorithm 5** The accelerated original online learning algorithm

1: Lines 1-8 from the Algorithm 1
2: \( H_s \leftarrow \{h : h \in H_1, M(h) \geq 1 - \theta \cdot \gamma_{\text{max}}\} \)
3: \( H_d \leftarrow \text{SortByDescending}(H_s, M(H_s)) \)
4: Lines 10-34 from Algorithm 1

Similarly, we can also change the steps of the IOL-GFMM shown in Algorithm 4 to Algorithm 6 to accelerate the IOL-GFMM procedure. With the Lemma 1, we can reduce the extendable hyperbox candidates to cover the new input pattern. The remaining operations are the same as the original version of the IOL-GFMM algorithm.

**Algorithm 6** The accelerated IOL-GFMM algorithm

1: Lines 1-8 from the Algorithm 4
2: \( H_s \leftarrow \{h : h \in H_1, M(h) \geq 1 - \theta \cdot \gamma_{\text{max}}\} \)
3: \( H_d \leftarrow \text{SortByDescending}(H_s, M(H_s)) \)
4: Lines 10-33 from Algorithm 4

As will be illustrated in the experimental section, these changes to the algorithms and the use of the proved Lemma 1 resulted in from 2 to 3.5 times reduction of learning time on average.
3.2. Accelerated agglomerative learning algorithms

In the original agglomerative learning algorithms, the hyperbox aggregation process considers all pairs of hyperboxes for which their similarity values are larger than or equal to a given minimum similarity threshold. If this threshold is set too small, then there might be many candidates considered, and so the training process can be long. However, when the minimum similarity condition is met and two hyperboxes could be merged, the newly aggregated hyperbox has to still be checked for the maximum hyperbox size constraint. In this paper, we will show the dependency of the similarity value with the maximum hyperbox size parameter. Based on this identified relationship, we can remove immediately the pairs of hyperboxes for which the hyperbox aggregated from these candidates cannot, with absolute certainty, satisfy the maximum hyperbox size condition. The details of the proposed method are described in Lemma 2.

**Lemma 2.** Regardless of the similarity measure used, the hyperbox aggregation process only considers pairs of hyperbox candidates that their similarity values satisfy the following condition: \( s(B_i, B_k) \geq \max(\sigma, 1 - \theta \cdot \gamma_{\max}) \), where \( \gamma_{\max} = \max_j(\gamma_j); \gamma_j > 0 \), \( \sigma \) is the minimum similarity threshold, and \( \theta \) is the maximum hyperbox size parameter.

**Proof.** See the proof in the appendix B

By using the Lemma 2, we can change the steps of the AGGLO-SM algorithm in Algorithm 2 to Algorithm 7, and modify the AGGLO-2 algorithm as shown in Algorithm 8. The only change in these accelerated algorithms compared to their original versions is the limitation of pairs of candidates considered during the learning process by a stricter lower bound. The remaining operations are kept unchanged as described in the original algorithms.

**Algorithm 7** The accelerated AGGLO-SM algorithm

1: Lines 1-5 from the Algorithm 2
2: \( S \leftarrow S \setminus \{ s \in S | s < \max(\sigma, 1 - \theta \cdot \gamma_{\max}) \} \)
3: Lines 7-22 from Algorithm 6

**Algorithm 8** The accelerated AGGLO-2 algorithm

1: Lines 1-7 from the Algorithm 3
2: \( S \leftarrow S \setminus \{ s \in S | s < \max(\sigma, 1 - \theta \cdot \gamma_{\max}) \} \)
3: Lines 9-29 from Algorithm 3

As will be shown in the experimental section, these changes to the agglomerative learning algorithms and the usage of the proposed Lemma 2 led to the acceleration of seven times in the training time of the AGGLO-SM algorithm, while the training time of the AGGLO-2 algorithm is reduced from 25 to 37 times on average depending on the similarity measure and dataset deployed.
4. Experiments

4.1. Experimental datasets and parameter settings

To evaluate the effectiveness of the proposed method, we conducted the experiments on 24 datasets taken from the UCI machine learning repository. A summary of these datasets related to the numbers of classes, features, and samples is shown in Table 1.

Table 1: The summary of the used datasets

| ID | Dataset                        | # samples | # features | # classes |
|----|--------------------------------|-----------|------------|-----------|
| 1  | balance scale                  | 625       | 4          | 3         |
| 2  | banknote authentication        | 1372      | 4          | 2         |
| 3  | blood transfusion              | 748       | 4          | 2         |
| 4  | breast cancer wisconsin        | 699       | 9          | 2         |
| 5  | breast cancer coimbra          | 116       | 9          | 2         |
| 6  | climate model crashes          | 540       | 18         | 2         |
| 7  | connectionist bench sonar      | 208       | 60         | 2         |
| 8  | glass                          | 214       | 9          | 6         |
| 9  | haberman                       | 306       | 3          | 2         |
| 10 | heart                          | 270       | 13         | 2         |
| 11 | ionosphere                     | 351       | 33         | 2         |
| 12 | movement libras                | 360       | 90         | 15        |
| 13 | optical digit                  | 5620      | 62         | 10        |
| 14 | page blocks                    | 5473      | 10         | 2         |
| 15 | pendigits                      | 10992     | 16         | 10        |
| 16 | pima diabetes                  | 768       | 8          | 2         |
| 17 | plant species leaves margin    | 1600      | 64         | 100       |
| 18 | plant species leaves texture   | 1600      | 64         | 100       |
| 19 | ringnorm                       | 7400      | 20         | 2         |
| 20 | seeds                          | 210       | 7          | 3         |
| 21 | image segmentation             | 2310      | 19         | 7         |
| 22 | spambase                       | 4601      | 57         | 2         |
| 23 | spectf heart                   | 267       | 44         | 2         |
| 24 | landsat satellite              | 6435      | 36         | 6         |

For each dataset, we carried out 5 times 2-fold cross-validation, and then the average values of the training time and the number of hyperbox candidates considered during the training process are reported in this paper. Experiments were run on a Intel Xeon Gold 6150 2.7GHz computer. For learning algorithms, the maximum hyperbox size is set $\theta = 0.1$ and the sensitivity threshold of $\gamma_j = 1; \forall j \in [1, n]$. In the agglomerative learning algorithms, we set the minimum similarity threshold $\sigma = 0$ to assess the impact of the lower bound related to $\theta$ on the training time of algorithms.

1https://archive.ics.uci.edu/ml/datasets.php
4.2. Experimental results for online learning algorithms

Table 2 shows the average speed-up factor and the number of hyperbox candidates considered during the learning process over 10 iterations (5 times 2-fold cross-validation) for each dataset. The speed-up value is computed by dividing the training time of the algorithm without using the lemma by the training time of the algorithm using the lemma to accelerate the learning process. The training time of online learning algorithms are shown in Table C.7 in the Appendix C.

Table 2: Speed-up factor and the number of hyperbox candidates considered during the training process of online learning algorithms

| Dataset                | Speed-up factor | Number of hyperbox candidates | Number of hyperbox candidates | Speed-up factor | Number of hyperbox candidates |
|------------------------|-----------------|-------------------------------|-------------------------------|-----------------|-------------------------------|
|                        |                 | w/t. lemma    | w. lemma               |                 | w/t. lemma    | w. lemma               |
| balance scale          | 5.7227          | 20880          | 0                       | 5.3978          | 20880          | 0                       |
| banknote authentication| 1.3016          | 5192.5         | 914                    | 1.0563          | 5129           | 904                    |
| blood transfusion      | 1.2487          | 2154.5         | 392                    | 1.0568          | 1542           | 320                    |
| breast cancer wisconsin| 3.8746          | 14655          | 0                       | 3.6556          | 14655          | 0                       |
| breast cancer coimbra  | 2.3571          | 768            | 2                       | 1.9804          | 768            | 2                       |
| climate model crashes  | 6.2478          | 30634          | 0                       | 6.0559          | 30634          | 0                       |
| connectionist bench sonar | 2.8392          | 2664.5         | 0                       | 2.5664          | 2664.5         | 0                       |
| glass                  | 1.3178          | 567            | 49.5                   | 1.1009          | 567            | 49.5                   |
| haberman               | 1.3418          | 960.5          | 143.5                  | 1.122           | 913.5          | 141.5                  |
| heart                  | 3.5575          | 4479.5         | 1                       | 3.1069          | 4479.5         | 1                       |
| ionosphere             | 2.7678          | 5705           | 46                      | 1.6163          | 5705           | 46                      |
| movement libras        | 1.5575          | 676            | 24.5                   | 1.0989          | 676            | 24.5                   |
| optical digit          | 4.5451          | 393452.5       | 0                       | 3.0049          | 393452.5       | 0                       |
| page blocks            | 1.3019          | 21124.5        | 2928.5                 | 1.0458          | 20825.5        | 2909.5                 |
| pendligits             | 4.2866          | 892103         | 3421.5                 | 1.0656          | 892103         | 3421.5                 |
| pima diabetes          | 4.9452          | 26995.5        | 197.5                  | 1.8977          | 26995.5        | 197.5                  |
| plant species leaves margin | 2.0362          | 2800            | 0                       | 1.3029          | 2800           | 0                       |
| plant species leaves texture | 7.2026          | 309308         | 16                      | 7.417           | 309308         | 16                      |
| ringnorm               | 12.7486         | 2986751.5      | 4108.5                 | 3.0377          | 2986751.5      | 4108.5                 |
| seeds                  | 1.7213          | 959.5          | 42.5                   | 1.2             | 959.5          | 42.5                   |
| image segmentation     | 1.8082          | 26153          | 1052.5                 | 1.0332          | 26153          | 1052.5                 |
| spambase               | 3.3182          | 329619.5       | 2815.5                 | 1.2982          | 329619.5       | 2815.5                 |
| spectf heart           | 3.8481          | 5929.5         | 0                       | 3.7564          | 5929.5         | 0                       |
| landsat satellite      | 2.76            | 349032.5       | 12759                  | 1.0733          | 349032.5       | 12759                  |
| Average                | 3.526963        | 226398.6       | 1204.75                | 2.372788        | 226355.8       | 1200.167               |

It can be easily observed that the online learning algorithms using the proposed lemma are much faster than ones without deploying the lemma. These figures can be explained based on the number of hyperboxes considered during the training process. We can see that the use of lemma has significantly reduced the unsuitable hyperbox candidates that the original versions have to verify. In several datasets, the number of hyperbox candidates is zero in the case of using the proposed lemma because all existing hyperboxes cannot
be extended to cover the new input patterns. It means that the resulting models only contain hyperboxes with one data point. In this case, the speed-up of learning process using the proposed lemma is obvious.

In general, the proposed method contributes to the acceleration of IOL-GFMM algorithm more significantly than the Onln-GFMM. This is because the training time of IOL-GFMM is usually faster than Onln-GFMM algorithm with the small value of \( \theta \) (\( \theta = 0.1 \) in this work) [3]. Therefore, when the number of candidates considered in the IOL-GFMM reduces, the overlap test operation between the extended hyperbox and the existing hyperboxes is conducted much faster. Meanwhile, the original online learning algorithm needs to check overlap and find the dimension to conduct the contraction for each pair of hyperboxes. These operations occupy most of the computational expense of the Onln-GFMM algorithm, so the obtained speed-up of Onln-GFMM is smaller than that of IOL-GFMM algorithm.

### 4.3. Experimental results of agglomerative learning algorithms

This part reports the experimental results of agglomerative learning algorithms with and without using the proposed lemma. Table 3 presents the speed-up factor of the AGGLO-2 algorithm compared to one without using the lemma for four similarity measures. These values are

| Dataset                  | Longest distance | Shortest distance | Mid-max distance | Mid-min distance |
|--------------------------|------------------|-------------------|------------------|------------------|
| balance scale            | 31.7626          | 31.9529           | 21.156           | 21.2067          |
| banknote authentication   | 3.5408           | 2.6579            | 2.6302           | 2.7469           |
| blood transfusion        | 4.3349           | 2.7389            | 2.7078           | 3.2302           |
| breast cancer wisconsin  | 21.3396          | 21.1669           | 14.7005          | 14.5726          |
| breast cancer coimbra    | 8.9              | 9.0253            | 6.0635           | 6.0397           |
| climate model crashes    | 32.5457          | 32.2009           | 20.9304          | 20.89             |
| connectionist bench sonar| 12.1624          | 12.0424           | 7.829            | 7.699            |
| glass                    | 4.1902           | 3.6753            | 2.9623           | 3.1639           |
| haberman                 | 5.0756           | 3.5084            | 3.1492           | 3.7729           |
| heart                    | 17.5549          | 17.5233           | 11.4509          | 11.3709           |
| ionosphere               | 14.5455          | 12.7882           | 8.9468           | 9.6922           |
| movement libras          | 3.84             | 3.4975            | 2.8035           | 2.9697           |
| optical digit            | 272.9711         | 259.0673          | 176.7774         | 183.3464         |
| page blocks              | 5.5846           | 3.4114            | 3.5063           | 4.1349           |
| pendigits                | 76.6387          | 59.2468           | 51.1295          | 57.0118          |
| pima diabetes            | 35.3624          | 28.1336           | 21.2801          | 23.7888          |
| plant species leaves margin| 5.8103          | 5.725             | 4.1252           | 4.1257           |
| plant species leaves texture| 59.9784         | 60.189            | 37.6131          | 37.3868          |
| ringnorm                 | 154.348          | 110.4403          | 89.7151          | 100.4713         |
| seeds                    | 6.2366           | 5.2175            | 4.1833           | 4.6059           |
| image segmentation       | 13.6811          | 9.7371            | 8.6332           | 10.1224          |
| spambase                 | 52.2312          | 29.2803           | 28.7548          | 34.9655          |
| spectf heart             | 20.6977          | 20.6945           | 13.2112          | 13.239           |
| landsat satellite        | 46.4068          | 16.1142           | 20.3232          | 32.471           |
| **Average**             | **37.9058**      | **31.6681**       | **23.52427**     | **25.54268**     |

This table presents the speed-up factor of the AGGLO-2 algorithm compared to one without using the lemma for four similarity measures. These values are
calculated from the average training time shown in Table C.8 in the Appendix C. Table 4 describes the number of considered hyperbox candidates of the AGGLO-2 algorithm for each similarity measure. Table 5 shows the speed-up factors of the AGGLO-SM algorithm on the experimental datasets, which are computed from their training time in Table C.9 in the Appendix C. The number of hyperbox candidates considered during the training process of the AGGLO-SM algorithm with four similarity measures is presented in Table 6.

Table 4: The number of hyperboxes considered during the training process of the AGGLO-2 algorithm

| Dataset                        | Longest distance | Shortest distance | Mid-max distance | Mid-min distance |
|--------------------------------|------------------|-------------------|-----------------|-----------------|
|                                | w/t. lemma       | w. lemma          | w/t. lemma      | w. lemma        | w/t. lemma      | w. lemma        |
| blance scale                   | 41760            | 0                 | 41760           | 0               | 41760           | 0               |
| banknote authentication        | 22340            | 552               | 22105.5         | 2896            | 22300           | 1675.5          | 22385.5         | 1046            |
| blood transfusion              | 17806.5          | 447               | 14332           | 1952.5          | 14470.5         | 1256.5          | 17637.5         | 836.5           |
| breast cancer wisconsin        | 119974           | 106.5             | 119974          | 106.5           | 119974          | 106.5           | 119974          | 106.5           |
| breast cancer coimbra          | 3038.5           | 2                 | 3038.5          | 2               | 3038.5          | 2               | 3038.5          | 2               |
| climate model crashes          | 61268            | 0                 | 61268           | 0               | 61268           | 0               | 61268           | 0               |
| connectionist bench sonar      | 5329             | 0                 | 5329            | 0               | 5329            | 0               | 5329            | 0               |
| glass                          | 3004             | 36.5              | 2954.5          | 120.5           | 2954            | 107.5           | 3005            | 60.5            |
| haberman                       | 5912.5           | 93.5              | 4913.5          | 398.5           | 4905.5          | 275.5           | 5746.5          | 142.5           |
| heart                          | 13311.5          | 1                 | 13311.5         | 1               | 13311.5         | 1               | 13311.5         | 1               |
| ionosphere                     | 33598            | 27                | 26748           | 167.5           | 26747.5         | 134.5           | 35599           | 36              |
| movement                       | 3897             | 34.5              | 3127.5          | 55              | 3127.5          | 54              | 3897            | 36              |
| optical digit                  | 786905           | 0                 | 786905          | 0               | 786905          | 0               | 786905          | 0               |
| page blocks                    | 235607           | 2522              | 164723.5        | 15074.5         | 194023          | 8211.5          | 231764.5        | 5870            |
| pendigits                      | 5080436          | 1276              | 5077333         | 19843           | 5077333         | 11493.5         | 5080439         | 2128.5          |
| pima diabetes                  | 128175.5         | 61.5              | 100825.5        | 576.5           | 100825.5        | 372.5           | 128175.5        | 84              |
| plant species leaves margin    | 5600             | 0                 | 5600            | 0               | 5600            | 0               | 5600            | 0               |
| plant species leaves texture   | 1232425          | 12.5              | 1232425         | 32.5            | 1232425         | 32.5            | 1232425         | 12.5            |
| ringnorm                       | 11629867         | 250               | 11631487        | 29762.5         | 11631487        | 14020.5         | 11629867        | 311             |
| seeds                          | 3333             | 28.5              | 3289.5          | 120             | 3289.5          | 95              | 3333            | 34.5            |
| image segmentation             | 181993           | 570               | 157079.5        | 4098.5          | 157231          | 2709            | 181772          | 1173            |
| spambase                       | 3707848          | 1150              | 3340482         | 51622.5         | 3345250         | 21569.5         | 3707269         | 4208.5          |
| spectf heart                   | 11859            | 0                 | 11859           | 0               | 11859           | 0               | 11859           | 0               |
| landsat satellite              | 2021658          | 1187.5            | 1956507         | 81882.5         | 1956399         | 39069           | 2438703         | 11483           |
| Average                        | 1056539          | 348.25            | 1032807         | 8696.333        | 1034242         | 4216.083        | 1073707         | 1149.25         |
In general, the use of the proposed lemma makes the agglomerative learning algorithm much faster because the number of candidates for the hyperbox aggregation process is considerably reduced. Among two versions of the batch learning algorithms, the influence of the proposed lemma on the performance of the AGGLO-2 is more significant compared to the AGGLO-SM. It is due to the fact that the number of hyperbox candidates for the aggregation process in the original AGGLO-2 is higher than the AGGLO-SM algorithm. For several datasets in the AGGLO-2 algorithm such as optical digit, pendigits, rignorm, and spambase, the use of the proposed lemma accelerates the training process from 50 to nearly 280 times compared to the original version. Although the AGGLO-SM algorithm cannot obtain such speed-up, the training time also reduces considerably when using the proposed lemma. These results confirm that our proposed method is efficient for all learning algorithms of the GFMM neural network.

Table 5: Speed-up factor of the AGGLO-SM algorithm

| Dataset               | Longest distance | Shortest distance | Mid-max distance | Mid-min distance |
|-----------------------|------------------|-------------------|------------------|------------------|
| balance scale         | 17.9903          | 17.9709           | 17.8937          | 17.8841          |
| banknote authentication| 1.1323           | 1.1549            | 1.1503           | 1.1634           |
| blood transfusion      | 1.1521           | 1.1381            | 1.1133           | 1.1198           |
| breast cancer wisconsin| 3.0204           | 3.0734            | 2.9323           | 3.0311           |
| breast cancer coimbra  | 2.6782           | 2.6897            | 2.6782           | 2.6782           |
| climate model crashes  | 21.6677          | 21.4537           | 21.7             | 21.4056          |
| connectionist bench sonar| 7.4505           | 7.4286            | 7.4505           | 7.4615           |
| glass                 | 1.0887           | 1.0678            | 1.0637           | 1.0775           |
| haberman              | 1.1469           | 1.1208            | 1.1308           | 1.1521           |
| heart                 | 1.7236           | 1.7073            | 1.6763           | 1.6911           |
| ionosphere            | 1.2263           | 1.3166            | 1.2595           | 1.2782           |
| movement libras       | 1.0842           | 1.0981            | 1.0841           | 1.0845           |
| optical digit         | 86.5943          | 83.8357           | 75.9623          | 83.9551          |
| page blocks           | 1.1686           | 1.1677            | 1.1389           | 1.1638           |
| pendigits             | 3.3746           | 1.6839            | 1.591            | 3.0278           |
| pima diabetes         | 1.5718           | 1.4924            | 1.519            | 1.5698           |
| plant species leaves margin| 3.6856          | 3.7029            | 3.6874           | 3.6836           |
| plant species leaves texture| 3.3463         | 3.3465            | 3.3581           | 3.3449           |
| rignorm               | 1.5222           | 1.4626            | 1.4874           | 1.5244           |
| seeds                 | 1.1595           | 1.1351            | 1.1399           | 1.1459           |
| image segmentation    | 1.0589           | 1.0521            | 1.0545           | 1.0577           |
| spambase              | 1.0672           | 1.0647            | 1.0658           | 1.0623           |
| spectf heart          | 12.6057          | 12.6098           | 12.5547          | 12.5587          |
| landsat satellite     | 1.2885           | 1.0914            | 1.0982           | 1.2314           |
| Average               | **7.49185**      | **7.286029**      | **6.949579**     | **7.348021**     |

In the AGGLO-2 algorithm, among four similarity measures, the proposed lemma has the most impact on the longest distance measure and the least influence on the mid-max distance-based similarity measure. This is because the number of hyperbox candidates considered during the original training process using the longest distance-based measure is highest, but after using the proposed method, the number of considered
Table 6: Number of hyperbox candidates considered during the training process of the AGGLO-SM algorithm

| Dataset               | Longest distance | Shortest distance | Mid-max distance | Mid-min distance |
|-----------------------|------------------|-------------------|------------------|------------------|
|                       | w/t. lemma       | w. lemma          | w/t. lemma       | w. lemma         |
| balance scale         | 18032            | 0                 | 18032            | 0                |
| banknote authentication| 5003             | 552.5             | 21709.5          | 18143.5          |
| blood transfusion     | 2426             | 572               | 3940             | 2482.5           |
| breast cancer wisconsin| 12731.5         | 106.5             | 12731.5          | 106.5            |
| breast cancer coimbra | 767.5            | 2                 | 767.5            | 2                |
| climate model crashes | 30631            | 0                 | 30631            | 0                |
| connectionist bench sonar | 2652.5       | 0                 | 2652.5           | 0                |
| glass                 | 543.5            | 37                | 679              | 198.5            |
| haberman              | 1018.5           | 95.5              | 1276             | 418.5            |
| heart                 | 419              | 1                 | 419              | 1                |
| ionosphere            | 4385.5           | 26.5              | 4522             | 219              |
| movement libras       | 690              | 34.5              | 727.5            | 77               |
| optical digit         | 268395.5         | 0                 | 268395.5         | 0                |
| page blocks           | 27414            | 3158.5            | 203481           | 188327.5         |
| pendigits             | 801368           | 1266              | 825055.5         | 32153            |
| pima diabetes         | 27268.5          | 60                | 30319.5          | 3535.5           |
| plant species leaves margin | 2792.5   | 0                 | 2792.5           | 0                |
| plant species leaves texture | 300635.5 | 12.5              | 301010           | 45.5             |
| ringnorm              | 3000191          | 241               | 3952685          | 965343.5         |
| seeds                 | 946              | 29.5              | 1152             | 268.5            |
| image segmentation    | 25541            | 571.5             | 68281.5          | 44145.5          |
| spambase              | 328946.5         | 1158              | 665797.5         | 356241.5         |
| spectf heart          | 5916             | 0                 | 5916             | 0                |
| landsat satellite     | 384186.5         | 1179.5            | 1592177          | 1258323          |
| **Average**           | **218870.9**     | **379.3333**      | **333964.6**     | **119584.7**     |

The number of hyperbox candidates considered in the training process with regard to the middle distance-based similarity measures using the proposed lemma is lower than that of the shortest distance-based measure, its speed-up factor on average is still lower compared to the value of the shortest-based measure. This is due to the fact that the middle-based measures are asymmetrical values, so for each pair of candidates \( B_i \) and \( B_k \), the training step has to spend time computing both similarity values \( s_{ik} \) and \( s_{ki} \). The repetition of similarity computation increases the training time and reduces the speed-up.
factor though there is a reduction in the number of considered candidates.

Similarly, the speed-up factor of the training process using the proposed lemma with regard to the longest distance-based similarity measure in the AGGLO-SM is highest because the number of hyperbox candidates after using the proposed lemma is much lower than those of other similarity measures. Hence, its training time is fastest on average. The influence of the proposed lemma on the AGGLO-SM using mid-max distance measure is still smallest among four similarity measures since its training process has to calculate the similarity values for each pair of hyperboxes twice and the number of hyperbox candidates is relatively high after using the proposed lemma. In contrast to the outcomes of the AGGLO-2, the impact of the proposed lemma on the training time of the AGGLO-SM algorithm using the mid-min distance-based measure is ranked in the second place as the number of candidates is much smaller than those of the shortest and mid-max distance-based measures.

5. Conclusion

One of the drawbacks in the current learning algorithms of the general fuzzy min-max neural network is the consideration of too many candidates in the expansion or aggregation process of hyperboxes. Therefore, this paper presented and proved stricter lower bounds for online and agglomerative learning algorithms of the GFMM neural network. The proposed method reduces significantly the unsuitable hyperbox candidates considered during the learning process, especially in the AGGLO-2 algorithm. Therefore, the training operations are accelerated when applying our proposed solutions. Experimental results on many datasets confirmed the effectiveness of our approach. In particular, the acceleration factors of the online learning algorithms are from two to three on average, while the training time of the AGGLO-SM algorithm is reduced about seven times on average. Especially, the speed-up factor in the AGGLO-2 algorithm using the proposed lemma can achieve from 30 to 250 on several datasets when the number of unsuitable hyperbox candidates is considerably reduced.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Thanh Tung Khuat: Conceptualization, Methodology, Validation, Software, Writing - original draft. Bogdan Gabrys: Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration.
Acknowledgement

Thanh Tung Khuat would like to thank the UTS-FEIT for awarding him Ph.D. scholarships.

Appendix A. Proof of Lemma 1

This is a proof of Lemma 1.

Proof. We need to prove that if lemma 1 is violated for any dimension $j \in [1, n]$, then the maximum hyperbox size condition is also not satisfied. For each $j^{th}$ dimension, there are six cases concerning the positions of the input pattern $X = [X^{l}, X^{h}]$ and the hyperbox $h = [V, W]$ as follows:

**Case 1:** $x_j^l \leq v_j \leq x_j^u \leq w_j$. The membership value along the $j^{th}$ dimension is:

$$b_{hj} = b_h(x_j) = \min([1 - f(x_j^u - w_j, \gamma_j)], [1 - f(v_j - x_j^l, \gamma_j)]) = 1 - \min((v_j - x_j^l) \cdot \gamma_j, 1)$$

We only consider $(v_j - x_j^l) \cdot \gamma_j \leq 1$, because in case of $(v_j - x_j^l) \cdot \gamma_j > 1 \Rightarrow b_{hj} = 0$ and $1 < (v_j - x_j^l) \cdot \gamma_j \leq (x_j^u - x_j^l) \cdot \gamma_j \leq \theta \cdot \gamma_i$, thus $1 - \theta \cdot \gamma_j < 0 \Rightarrow b_{hj} = 0 > 1 - \theta \cdot \gamma_j$, and the lemma holds for the $j^{th}$ dimension.

For $(v_j - x_j^l) \cdot \gamma_j \leq 1$, we have:

$$b_{hj} = 1 - (v_j - x_j^l) \cdot \gamma_j \Leftrightarrow v_j - x_j^l = (1 - b_{hj})/\gamma_j$$

If the hyperbox $h$ is expanded, then:

$$v_j^{\text{new}} = \min(v_j, x_j^l) = x_j^l; \quad w_j^{\text{new}} = \max(w_j, x_j^u) = w_j$$

The newly aggregated hyperbox satisfies the maximum hyperbox size $\theta$ on the $j^{th}$ dimension if and only if:

$$w_j^{\text{new}} - v_j^{\text{new}} \leq \theta \Leftrightarrow w_j - x_j^l \leq \theta \Leftrightarrow (w_j - v_j) + (v_j - x_j^l) \leq \theta \Leftrightarrow (w_j - v_j) + (1 - b_{hj})/\gamma_j \leq \theta$$

$$\Leftrightarrow b_{hj} \geq 1 - \theta \cdot \gamma_j + (w_j - v_j) \cdot \gamma_j \quad \text{(because of } \gamma_j > 0)$$

We also have: $1 - \theta \cdot \gamma_j + (w_j - v_j) \cdot \gamma_j \geq 1 - \theta \cdot \gamma_j$, because $(w_j - v_j) \cdot \gamma_j \geq 0$. Therefore, if $b_{hj} < 1 - \theta \cdot \gamma_j$, it guarantees that the maximum hyperbox size condition is not satisfied for the $j^{th}$ dimension. Hence, in this case, we need $b_{hj} \geq 1 - \theta \cdot \gamma_j$ so that the maximum hyperbox size condition is met.

**Case 2:** $v_j \leq x_j^l \leq w_j \leq x_j^u$. The membership value along the $j^{th}$ dimension is:

$$b_{hj} = b_h(x_j) = \min([1 - f(x_j^u - w_j, \gamma_j)], [1 - f(v_j - x_j^l, \gamma_j)]) = 1 - \min((x_j^u - w_j) \cdot \gamma_j, 1)$$

Similarly to case 1, we only consider $(x_j^u - w_j) \cdot \gamma_j \leq 1$, thus we have:

$$b_{hj} = 1 - (x_j^u - w_j) \cdot \gamma_j \Leftrightarrow x_j^u - w_j = (1 - b_{hj})/\gamma_j$$
The remaining proof is similar to case 1, and we also have $b_{h_j} \geq 1 - \theta \cdot \gamma_j$ in this case.

**Case 3:** $x^l_j \leq v_j \leq w_j \leq x^u_j$. The membership value along the $j^{th}$ dimension is:

$$b_{h_j} = b_h(x_j) = \min(1 - f(x^u_j - w_j, \gamma_j), 1 - f(v_j - x^l_j, \gamma_j)) = \min(1 - (x^u_j - w_j) \cdot \gamma_j, 1 - (v_j - x^l_j) \cdot \gamma_j)$$

If the hyperbox $h$ is expanded, then:

$$v^\text{new}_j = \min(v_j, x^l_j) = v_j; \quad w^\text{new}_j = \max(w_j, x^u_j) = x^u_j$$

The newly expanded hyperbox satisfies the maximum hyperbox size $\theta$ if and only if:

$$w^\text{new}_j - v^\text{new}_j \leq \theta \Leftrightarrow x^u_j - v_j \leq \theta \Leftrightarrow (x^u_j - w_j) + (w_j - v_j) \leq \theta \Leftrightarrow (1 - b_{h_j})/\gamma_j + (w_j - v_j) \leq \theta$$

$$\Leftrightarrow b_{h_j} \geq 1 - \theta \cdot \gamma_j + (w_j - v_j) \cdot \gamma_j$$

We have: $1 - \theta \cdot \gamma_j + (w_j - x^l_j) \cdot \gamma_j \geq 1 - \theta \cdot \gamma_j$, because $w_j - x^l_j \geq 0$. The remaining proof is similar to case 1.

**Case 3.1:** $x^u_j - w_j \geq v_j - x^l_j$, we have:

$$b_{h_j} = 1 - (x^u_j - w_j) \cdot \gamma_j \Leftrightarrow x^u_j - w_j = (1 - b_{h_j})/\gamma_j$$

The newly aggregated hyperbox satisfies the maximum hyperbox size $\theta$ if and only if:

$$w^\text{new}_j - v^\text{new}_j \leq \theta \Leftrightarrow x^u_j - x^l_j \leq \theta \Leftrightarrow (x^u_j - w_j) + (w_j - x^l_j) \leq \theta \Leftrightarrow (1 - b_{h_j})/\gamma_j + (w_j - x^l_j) \leq \theta$$

$$\Leftrightarrow b_{h_j} \geq 1 - \theta \cdot \gamma_j + (w_j - x^l_j) \cdot \gamma_j$$

**Case 3.2:** $x^u_j - w_j < v_j - x^l_j$, we have:

$$b_{h_j} = 1 - (v_j - x^l_j) \cdot \gamma_j \Leftrightarrow v_j - x^l_j = (1 - b_{h_j})/\gamma_j$$

The newly aggregated hyperbox satisfies the maximum hyperbox size $\theta$ if and only if:

$$w^\text{new}_j - v^\text{new}_j \leq \theta \Leftrightarrow x^u_j - x^l_j \leq \theta \Leftrightarrow (x^u_j - v_j) + (v_j - x^l_j) \leq \theta \Leftrightarrow (1 - b_{h_j})/\gamma_j + (x^u_j - v_j) \leq \theta$$

$$\Leftrightarrow b_{h_j} \geq 1 - \theta \cdot \gamma_j + (x^u_j - v_j) \cdot \gamma_j$$

We have: $1 - \theta \cdot \gamma_j + (x^u_j - v_j) \cdot \gamma_j \geq 1 - \theta \cdot \gamma_j$, because $x^u_j - v_j \geq 0$. The remaining proof is similar to case 1.

**Case 4:** $v_j \leq x^l_j \leq x^u_j \leq w_j$, we have the membership value: $b_{h_j} = 1 \geq 1 - \theta \cdot \gamma_j$, and we gain $b_{h_j} \geq 1 - \theta \cdot \gamma_j$ in this case.
The lemma holds for the $j^{th}$ dimension:

$$b_{hj} = b_h(x_j) = \min([1 - f(x_j^u - w_j, \gamma_j)], [1 - f(v_j - x_j^l, \gamma_j)]) = 1 - \min((x_j^u - w_j) \cdot \gamma_j, 1)$$

If the hyperbox $h$ is expanded, then:

$$v_j^{new} = \min(v_j, x_j^l) = v_j; \quad w_j^{new} = \max(w_j, x_j^u) = x_j^u$$

**Case 5.1**: $(x_j^u - w_j) \cdot \gamma_j > 1 \Leftrightarrow x_j^u - w_j > 1/\gamma_j$ due to $\gamma_j > 0$. In addition, $b_{hj} = 0$.

The newly aggregated hyperbox satisfies the maximum hyperbox size $\theta$ if and only if:

$$w_j^{new} - v_j^{new} \leq \theta \Leftrightarrow x_j^u - v_j \leq \theta \Leftrightarrow (x_j^u - w_j) + (w_j - v_j) \leq \theta \Leftrightarrow x_j^u - w_j \leq \theta \Leftrightarrow 1/\gamma_j < \theta \Leftrightarrow 1 - \theta \cdot \gamma_j < 0 = b_{hj}$$

The lemma holds for the $j^{th}$ dimension.

**Case 5.2**: $(x_j^u - w_j) \cdot \gamma_j \leq 1$, we have:

$$b_{hj} = 1 - (x_j^u - w_j) \cdot \gamma_j \Leftrightarrow x_j^u - w_j = (1 - b_{hj})/\gamma_j$$

The newly aggregated hyperbox satisfies the maximum hyperbox size $\theta$ if and only if:

$$w_j^{new} - v_j^{new} \leq \theta \Leftrightarrow x_j^u - v_j \leq \theta \Leftrightarrow (x_j^u - w_j) + (w_j - v_j) \leq \theta \Leftrightarrow (1 - b_{hj})/\gamma_j + (w_j - v_j) \leq \theta \Leftrightarrow b_{hj} \geq 1 - \theta \cdot \gamma_j + (w_j - v_j) \cdot \gamma_j \geq 1 - \theta \cdot \gamma_j$$

Hence, we also have $b_{hj} \geq 1 - \theta \cdot \gamma_j$ in this case.

**Case 6**: $x_j^l \leq x_j^u \leq v_j \leq w_j$. The membership value along the $j^{th}$ dimension is:

$$b_{hj} = b_h(x_j) = \min([1 - f(x_j^u - w_j, \gamma_j)], [1 - f(v_j - x_j^l, \gamma_j)]) = 1 - \min((v_j - x_j^l) \cdot \gamma_j, 1)$$

If the hyperbox $h$ is expanded, then:

$$v_j^{new} = \min(v_j, x_j^l) = x_j^l; \quad w_j^{new} = \max(w_j, x_j^u) = w_j$$

**Case 6.1**: $(v_j - x_j^l) \cdot \gamma_j > 1 \Leftrightarrow v_j - x_j^l > 1/\gamma_j$ due to $\gamma_j > 0$. In addition, $b_{hj} = 0$.

The newly aggregated hyperbox satisfies the maximum hyperbox size $\theta$ if and only if:

$$w_j^{new} - v_j^{new} \leq \theta \Leftrightarrow w_j - x_j^l \leq \theta \Leftrightarrow (w_j - v_j) + (v_j - x_j^l) \leq \theta \Leftrightarrow v_j - x_j^l \leq \theta \Leftrightarrow 1/\gamma_j < \theta \Leftrightarrow 1 - \theta \cdot \gamma_j < 0 = b_{hj}$$

The lemma holds for the $j^{th}$ dimension.

**Case 6.2**: $(v_j - x_j^l) \cdot \gamma_j \leq 1$, we obtain:

$$b_{hj} = 1 - (v_j - x_j^l) \cdot \gamma_j \Leftrightarrow v_j - x_j^l = (1 - b_{hj})/\gamma_j$$

The newly aggregated hyperbox satisfies the maximum hyperbox size $\theta$ if and only if:

$$w_j^{new} - v_j^{new} \leq \theta \Leftrightarrow w_j - x_j^l \leq \theta \Leftrightarrow (w_j - v_j) + (v_j - x_j^l) \leq \theta \Leftrightarrow (1 - b_{hj})/\gamma_j + (w_j - v_j) \leq \theta \Leftrightarrow b_{hj} \geq 1 - \theta \cdot \gamma_j + (w_j - v_j) \cdot \gamma_j \geq 1 - \theta \cdot \gamma_j$$
This lemma holds for the $j^{th}$ dimension in this case.

Meanwhile, the membership function for whole dimensions is $b_h(X) = \min b_{hj}$ and $0 \leq \theta \cdot \gamma_j \leq \theta \cdot \gamma_{max}$. We proved that the maximum hyperbox size condition is satisfied on any dimension $j$ if and only if $b_{hj} \geq 1 - \theta \cdot \gamma_j \geq 1 - \theta \cdot \gamma_{max}$. Therefore, the maximum hyperbox size condition is met for every dimension $j$ if and only if $b_h(X) \geq 1 - \theta \cdot \gamma_{max}$. The lemma is proved.

\[ \Box \]

**Appendix B. Proof of Lemma 2**

This is a proof of Lemma 2.

*Proof.* First of all, we need to prove that if the similarity value $s_{ik} = s(B_i, B_k) \geq 1 - \theta \cdot \gamma_{max}$ is violated for any dimension $j \in [1, n]$, then the maximum hyperbox size constraint is also not satisfied for the hyperbox aggregated from $B_i$ and $B_k$. We see that the similarity measure using middle distance between two hyperboxes is the same as the membership value between a hyperbox and an input pattern. Therefore, the proof is the same as in the appendix A. Here, we only prove the above condition for the longest and shortest distance measures.

**Using the shortest distance based similarity measure**

The shortest distance based similarity value for the $j^{th}$ is computed as follows:

\[ \tilde{s}_{ik}^j = \min[1 - f(v_{kj} - w_{ij}, \gamma_j), 1 - f(v_{ij} - w_{kj}, \gamma_j)] \]

For each $j^{th}$ dimension, there are six cases concerning the positions of the hyperbox $B_i = [V_i, W_i]$ and the hyperbox $B_k = [V_k, W_k]$ as follows:

**Case 1:** $v_{ij} \leq v_{kj} \leq w_{ij} \leq w_{kj}$. The similarity value: $\tilde{s}_{ik}^j = 1 \geq 1 - \theta \cdot \gamma_j$ (because of $\theta \cdot \gamma_j > 0$). The condition holds in this case.

**Case 2:** $v_{kj} \leq v_{ij} \leq w_{kj} \leq w_{ij}$. The similarity value: $\tilde{s}_{ik}^j = 1 \geq 1 - \theta \cdot \gamma_j$. The condition also holds in this case.

**Case 3:** $v_{kj} \leq w_{kj} \leq v_{ij} \leq w_{ij}$. The coordinate at the $j^{th}$ dimension of the hyperbox aggregated from $B_i$ and $B_k$ is:

\[ v_{ij}^{new} = \min(v_{ij}, v_{kj}) = v_{kj}; \quad w_{ij}^{new} = \max(w_{ij}, w_{kj}) = w_{ij} \]

The similarity value: $\tilde{s}_{ik}^j = 1 - \min[1, (v_{ij} - w_{kj}) \cdot \gamma_j]$

**Case 3.1:** $(v_{ij} - w_{kj}) \cdot \gamma_j > 1 \Leftrightarrow v_{ij} - w_{kj} > 1/\gamma_j$ (because of $\gamma_j > 0$). In this case: $\tilde{s}_{ik}^j = 0$. The newly aggregated hyperbox must satisfy the maximum hyperbox box size condition:

\[ w_{ij}^{new} - v_{ij}^{new} \leq \theta \Leftrightarrow v_{ij} - v_{kj} \leq \theta \Leftrightarrow (v_{ij} - v_{ij}) + (v_{ij} - w_{kj}) + (w_{kj} - v_{kj}) \leq \theta \Leftrightarrow v_{ij} - w_{kj} \leq \theta \]

\[ \Leftrightarrow 1/\gamma_j < \theta \Leftrightarrow 1 - \theta \cdot \gamma_j < 0 = \tilde{s}_{ik}^j \]
The constraint is correct in this case.

**Case 3.2:** \((v_{ij} - w_{kj}) \cdot \gamma_j \leq 1 \Rightarrow \tilde{s}_{ik}^j = 1 - (v_{ij} - w_{kj}) \cdot \gamma_j \Leftrightarrow v_{ij} - w_{kj} = (1 - \tilde{s}_{ik}^j)/\gamma_j\)

The newly aggregated hyperbox satisfies the maximum hyperbox box size condition if and only if:

\[w_{j}^{new} - v_{j}^{new} \leq \theta \Leftrightarrow w_{ij} - v_{kj} \leq \theta \Leftrightarrow (w_{ij} - v_{ij}) + (v_{ij} - w_{kj}) \leq \theta \Leftrightarrow v_{ij} - w_{kj} \leq \theta\]

\[\Leftrightarrow (1 - \tilde{s}_{ik}^j)/\gamma_j \leq \theta \Leftrightarrow \tilde{s}_{ik}^j \geq 1 - \theta \cdot \gamma_j\]

The constraint is also correct in this case.

**Case 4:** \(v_{ij} \leq w_{ij} \leq v_{kj} \leq w_{kj}\). This case is proved similarly to case 3.

**Case 5:** \(v_{kj} \leq v_{ij} \leq w_{ij} \leq w_{kj}\). The similarity value: \(\tilde{s}_{ik}^j = 1 \geq 1 - \theta \cdot \gamma_j\). The condition also holds in this case.

**Case 6:** \(v_{ij} \leq v_{kj} \leq w_{kj} \leq w_{ij}\). The similarity value: \(\tilde{s}_{ik}^j = 1 \geq 1 - \theta \cdot \gamma_j\). The condition also holds in this case.

**Using the longest distance based similarity measure**

The longest distance based similarity value for the \(j^{th}\) is calculated as follows:

\[\tilde{s}_{ik}^j = \min[1 - f(w_{kj} - v_{ij}, \gamma_j), 1 - f(w_{ij} - v_{kj}, \gamma_j)]\]

For each \(j^{th}\) dimension, we also consider in turn six cases relevant to the positions of the hyperbox \(B_i = [V_i, W_i]\) and the hyperbox \(B_k = [V_k, W_k]\) as follows:

**Case 1:** \(v_{ij} \leq v_{kj} \leq w_{ij} \leq w_{kj}\). The coordinate at the \(j^{th}\) dimension of the hyperbox aggregated from \(B_i\) and \(B_k\) is:

\[v_{j}^{new} = \min(v_{ij}, v_{kj}) = v_{ij}; \quad w_{j}^{new} = \max(w_{ij}, w_{kj}) = w_{kj}\]

The similarity value: \(\tilde{s}_{ik}^j = \min[1 - \min((w_{kj} - v_{ij}) \cdot \gamma_j, 1), 1 - \min((w_{ij} - v_{kj}) \cdot \gamma_j, 1)]\). In this case, we have \(w_{kj} - v_{ij} \geq w_{kj} - v_{kj} \geq w_{ij} - v_{ij} \Rightarrow \min((w_{kj} - v_{ij}) \cdot \gamma_j, 1) \geq \min((w_{ij} - v_{kj}) \cdot \gamma_j, 1)\). Therefore, \(\tilde{s}_{ik}^j = 1 - \min((w_{kj} - v_{ij}) \cdot \gamma_j, 1)\)

**Case 1.1:** \((w_{kj} - v_{ij}) \cdot \gamma_j > 1 \Leftrightarrow w_{ij} - v_{ij} > 1/\gamma_j\) (because of \(\gamma_j > 0\)). In this case: \(\tilde{s}_{ik}^j = 0\). The newly aggregated hyperbox meets the maximum hyperbox box size condition if and only if:

\[w_{j}^{new} - v_{j}^{new} \leq \theta \Leftrightarrow w_{kj} - v_{ij} \leq \theta \Leftrightarrow 1/\gamma_j < \theta \Leftrightarrow 1 - \theta \cdot \gamma_j < 0 = \tilde{s}_{ik}^j\]

The constraint is correct in this case.

**Case 1.2:** \((w_{kj} - v_{ij}) \cdot \gamma_j \leq 1 \Rightarrow \tilde{s}_{ik}^j = 1 - (w_{kj} - v_{ij}) \cdot \gamma_j \Leftrightarrow w_{kj} - v_{ij} = (1 - \tilde{s}_{ik}^j)/\gamma_j\)

The newly aggregated hyperbox satisfies the maximum hyperbox box size condition if and only if:

\[w_{j}^{new} - v_{j}^{new} \leq \theta \Leftrightarrow w_{kj} - v_{ij} \leq \theta \Leftrightarrow (1 - \tilde{s}_{ik}^j)/\gamma_j \leq \theta \Leftrightarrow \tilde{s}_{ik}^j \geq 1 - \theta \cdot \gamma_j\]

The constraint is also correct in this case.
**Case 2**: \( v_{kj} \leq v_{ij} \leq w_{kj} \leq w_{ij} \). This case is proved similarly to case 1.

**Case 3**: \( v_{kj} \leq w_{kj} \leq v_{ij} \leq w_{ij} \). The coordinate at the \( j^{th} \) dimension of the hyperbox aggregated from \( B_i \) and \( B_k \) is:

\[
v_{j}^{\text{new}} = \min(v_{ij}, v_{kj}) = v_{kj}; \quad w_{j}^{\text{new}} = \max(w_{ij}, w_{kj}) = w_{ij}
\]

The similarity value: \( \hat{s}_{ik}^j = 1 - \min((w_{ij} - v_{kj}) \cdot \gamma_j, 1) \).

**Case 3.1**: \((w_{ij} - v_{kj}) \cdot \gamma_j > 1 \iff w_{ij} - v_{kj} > 1/\gamma_j \) (because of \( \gamma_j > 0 \)). In this case: \( \hat{s}_{ik}^j = 0 \). The newly aggregated hyperbox meets the maximum hyperbox box size condition if and only if:

\[
w_{j}^{\text{new}} - v_{j}^{\text{new}} \leq \theta \iff w_{ij} - v_{kj} \leq \theta \iff 1/\gamma_j < \theta \iff 1 - \theta \cdot \gamma_j < 0 = \hat{s}_{ik}^j
\]

The constraint holds in this case.

**Case 3.2**: \((w_{ij} - v_{kj}) \cdot \gamma_j \leq 1 \Rightarrow \hat{s}_{ik}^j = 1 - (w_{ij} - v_{kj}) \cdot \gamma_j \iff w_{ij} - v_{kj} = (1 - \hat{s}_{ik}^j)/\gamma_j \)

The newly aggregated hyperbox satisfies the maximum hyperbox box size condition if and only if:

\[
w_{j}^{\text{new}} - v_{j}^{\text{new}} \leq \theta \iff w_{ij} - v_{kj} \leq \theta \iff (1 - \hat{s}_{ik}^j)/\gamma_j \leq \theta \iff \hat{s}_{ik}^j \geq 1 - \theta \cdot \gamma_j
\]

The constraint also holds in this case.

**Case 4**: \( v_{ij} \leq w_{ij} \leq v_{kj} \leq w_{kj} \). This case is proved similarly to case 3.

**Case 5**: \( v_{kj} \leq v_{ij} \leq w_{ij} \leq w_{kj} \). The coordinate at the \( j^{th} \) dimension of the hyperbox aggregated from \( B_i \) and \( B_k \) is:

\[
v_{j}^{\text{new}} = \min(v_{ij}, v_{kj}) = v_{kj}; \quad w_{j}^{\text{new}} = \max(w_{ij}, w_{kj}) = w_{kj}
\]

The similarity value: \( \hat{s}_{ik}^j = \min(1 - \min((w_{kj} - v_{ij}) \cdot \gamma_j, 1), 1 - \min((w_{ij} - v_{kj}) \cdot \gamma_j, 1)) \).

**Case 5.1**: \( \hat{s}_{ik}^j = 1 - \min((w_{kj} - v_{ij}) \cdot \gamma_j, 1) \)

**Case 5.1a**: \((w_{kj} - v_{ij}) \cdot \gamma_j > 1 \iff w_{kj} - v_{ij} > 1/\gamma_j \) (because of \( \gamma_j > 0 \)). In this case: \( \hat{s}_{ik}^j = 0 \). The newly aggregated hyperbox meets the maximum hyperbox box size condition if and only if:

\[
w_{j}^{\text{new}} - v_{j}^{\text{new}} \leq \theta \iff w_{kj} - v_{ij} \leq \theta \iff (w_{kj} - v_{ij}) + (v_{ij} - v_{kj}) \leq \theta \iff w_{kj} - v_{kj} \leq \theta \quad \text{(because of } v_{ij} - v_{kj} \geq 0) \]

\[
\iff 1/\gamma_j < \theta \iff 1 - \theta \cdot \gamma_j < 0 = \hat{s}_{ik}^j
\]

The constraint holds in this case.

**Case 5.1b**: \((w_{kj} - v_{ij}) \cdot \gamma_j \leq 1 \Rightarrow \hat{s}_{ik}^j = 1 - (w_{kj} - v_{ij}) \cdot \gamma_j \iff w_{kj} - v_{ij} = (1 - \hat{s}_{ik}^j)/\gamma_j \). The newly aggregated hyperbox meets the maximum hyperbox box size condition if and only if:

\[
w_{j}^{\text{new}} - v_{j}^{\text{new}} \leq \theta \iff w_{kj} - v_{ij} \leq \theta \iff (w_{kj} - v_{ij}) + (v_{ij} - v_{kj}) \leq \theta \iff w_{kj} - v_{kj} \leq \theta \quad \text{(because of } v_{ij} - v_{kj} \geq 0) \]

\[
\iff (1 - \hat{s}_{ik}^j)/\gamma_j \leq \theta \iff 1 - \theta \cdot \gamma_j \leq \hat{s}_{ik}^j
\]

The constraint is also correct in this case.

**Case 5.2**: \( \hat{s}_{ik}^j = 1 - \min((w_{ij} - v_{kj}) \cdot \gamma_j, 1) \)
**Case 5.2a:** \((w_{ij} - v_{kj}) \cdot \gamma_j > 1 \iff w_{ij} - v_{kj} > 1/\gamma_j\) (because of \(\gamma_j > 0\)). In this case: \(\hat{s}_{ik}^j = 0\). The newly aggregated hyperbox meets the maximum hyperbox box size condition if and only if:

\[
\begin{align*}
\hat{s}_{ik}^j &= 0. \\
\text{The constraint holds in this case.}
\end{align*}
\]

**Case 5.2b:** \((w_{ij} - v_{kj}) \cdot \gamma_j \leq 1 \Rightarrow \hat{s}_{ik}^j = 1 - (w_{ij} - v_{kj}) \cdot \gamma_j \iff w_{ij} - v_{kj} = (1 - \hat{s}_{ik}^j)/\gamma_j\). The newly aggregated hyperbox meets the maximum hyperbox box size condition if and only if:

\[
\begin{align*}
\hat{s}_{ik}^j &= 1 - \theta \cdot \gamma_j < 0 = \hat{s}_{ik}^j \\
\text{The constraint holds in this case.}
\end{align*}
\]

**Case 6:** \(v_{ij} \leq v_{kj} \leq w_{kj} \leq w_{ij}\). This case is proved similarly to case 5.

From the above proofs, we can see that the hyperbox aggregated from two hyperboxes \(B_i\) and \(B_k\) satisfies the maximum hyperbox size condition if and only if the similarity value between these two hyperboxes at the \(j\)th dimension meets the condition: \(s_{ik}^j \geq 1 - \theta \cdot \gamma_j; \forall j \in [1, n]\). We also have \(0 \leq \theta \cdot \gamma_j \leq \theta \cdot \gamma_{\max} \Rightarrow 1 - \theta \cdot \gamma_j \geq 1 - \theta \cdot \gamma_{\max}; \forall j \in [1, n]\). Therefore, the overall similarity value between two hyperboxes: \(s_{ik} = \min_{j=1}^{n} (s_{ik}^j) \geq 1 - \theta \cdot \gamma_{\max}\). In addition, in the agglomerative learning, two hyperboxes \(B_i\) and \(B_k\) are aggregated if their similarity value \(s_{ik} \geq \sigma\), where \(\sigma\) is a given minimum similarity threshold. From these two conditions, two aggregated hyperboxes must meet the condition: \(s_{ik} \geq \max(\sigma, 1 - \theta \cdot \gamma_{\max})\). The lemma 2 is proved.

\[\Box\]

**Appendix C. Training time of algorithms**

This appendix subsection shows the training time of online and agglomerative learning algorithms from the experiments in this paper. Table C.7 shows the training time of the IOL-GFMM and original online learning algorithms. Table C.8 presents the training time of the AGGLO-2 algorithm, while Table C.9 describes the learning time of the AGGLO-SM algorithm.

**References**

[1] B. Gabrys. Agglomerative learning algorithms for general fuzzy min-max neural network. *Journal of VLSI signal processing systems for signal, image and video technology*, 32(1):67–82, 2002.

[2] B. Gabrys and A. Bargiela. General fuzzy min-max neural network for clustering and classification. *IEEE Transactions on Neural Networks*, 11(3):769–783, 2000.
| Dataset                  | IOL-GFMM w/t. lemma | IOL-GFMM w. lemma | Onln-GFMM w/t. lemma | Onln-GFMM w. lemma |
|--------------------------|---------------------|-------------------|----------------------|--------------------|
| balance scale            | 0.1465              | 0.0256            | 0.1479               | 0.0274             |
| banknote authentication  | 0.0984              | 0.0756            | 0.471                | 0.4459             |
| blood transfusion        | 0.0497              | 0.0398            | 0.1322               | 0.1251             |
| breast cancer wisconsin  | 0.1205              | 0.0311            | 0.121                | 0.0331             |
| breast cancer coimbra    | 0.0099              | 0.0042            | 0.0101               | 0.0051             |
| climate model crashes    | 0.2093              | 0.0335            | 0.2168               | 0.0358             |
| connectionist bench sonar| 0.03                | 0.0106            | 0.029                | 0.0113             |
| glass                    | 0.0141              | 0.0107            | 0.0382               | 0.0347             |
| haberman                 | 0.0212              | 0.0158            | 0.0469               | 0.0418             |
| heart                    | 0.0402              | 0.0113            | 0.0407               | 0.0131             |
| ionosphere               | 0.0584              | 0.0211            | 0.091                | 0.0563             |
| movement libras          | 0.0271              | 0.0174            | 0.0578               | 0.0526             |
| optical digit            | 4.9332              | 1.0854            | 3.7477               | 1.2472             |
| page blocks              | 0.8802              | 0.6761            | 4.8526               | 4.64               |
| pendigits                | 14.4605             | 3.3734            | 179.4647             | 168.4146           |
| pima diabetes            | 0.406               | 0.0821            | 0.6995               | 0.3686             |
| plant species leaves margin | 0.2702          | 0.1327            | 0.1768               | 0.1357             |
| plant species leaves texture | 4.3093        | 0.5983            | 4.5103               | 0.6081             |
| ringnorm                 | 38.1094             | 2.9893            | 54.7914              | 18.037             |
| seeds                    | 0.0315              | 0.0183            | 0.0714               | 0.0595             |
| image segmentation       | 0.7204              | 0.3984            | 10.2108              | 9.8831             |
| spambase                 | 5.567               | 1.6777            | 17.4908              | 13.4731            |
| spectf heart             | 0.1039              | 0.027             | 0.1033               | 0.0275             |
| landsat satellite        | 6.4317              | 2.3303            | 58.5758              | 54.5749            |
| **Average**              | **3.210358**        | **0.570238**      | **14.00407**         | **11.34798**       |

[3] T.T. Khuat, F. Chen, and B. Gabrys. An improved online learning algorithm for general fuzzy min-max neural network. arXiv e-prints, page arXiv:2001.02391, 2020.

[4] T.T. Khuat, D. Ruta, and B. Gabrys. Hyperbox based machine learning algorithms: A comprehensive survey. arXiv e-prints, page arXiv:1901.11303, 2019.

[5] C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1:206–215, 2019.

[6] P.K. Simpson. Fuzzy min-max neural networks. i. classification. IEEE Transactions on Neural Networks, 3(5):776–786, 1992.

[7] P.K. Simpson. Fuzzy min-max neural networks - part 2: Clustering. IEEE Transactions on Fuzzy Systems, 1(1):32, 1993.
Table C.8: Training time of the AGGLO-2 algorithm

| Dataset                          | Longest distance |        | Shortest distance |        | Mid-max distance |        | Mid-min distance |        |
|----------------------------------|------------------|--------|-------------------|--------|------------------|--------|------------------|--------|
|                                  | w/t. lemma       | w. lemma | w/t. lemma       | w. lemma | w/t. lemma       | w. lemma | w/t. lemma       | w. lemma |
| balance scale                    | 0.883            | 0.0278  | 0.8819            | 0.0276  | 0.8949           | 0.0423  | 0.8928           | 0.0421  |
| banknote authentication          | 0.6161           | 0.174   | 0.6153            | 0.2315  | 0.6686           | 0.2542  | 0.6675           | 0.243   |
| blood                            | 0.4673           | 0.1078  | 0.3892            | 0.1421  | 0.417            | 0.154   | 0.4842           | 0.1499  |
| breast cancer wisconsin          | 2.74             | 0.1284  | 2.7136            | 0.1282  | 2.8078           | 0.191   | 2.7892           | 0.1914  |
| breast cancer coimbra            | 0.0712           | 0.008   | 0.0713            | 0.0079  | 0.0764           | 0.0126  | 0.0761           | 0.0126  |
| climate model crashes            | 1.4255           | 0.0438  | 1.4104            | 0.0438  | 1.4442           | 0.069   | 1.4435           | 0.0691  |
| connectionist                    | 0.1423           | 0.0117  | 0.1421            | 0.0118  | 0.1511           | 0.0193  | 0.1599           | 0.0196  |
| glass                            | 0.0859           | 0.0205  | 0.0849            | 0.0231  | 0.0942           | 0.0318  | 0.0946           | 0.0299  |
| haberman                         | 0.1477           | 0.0291  | 0.1249            | 0.0356  | 0.1351           | 0.0429  | 0.1562           | 0.0414  |
| heart                            | 0.3037           | 0.0173  | 0.3014            | 0.0172  | 0.3149           | 0.0275  | 0.3127           | 0.0275  |
| ionosphere                       | 0.832            | 0.0572  | 0.6791            | 0.0524  | 0.6898           | 0.0771  | 0.8597           | 0.0887  |
| movement                         | 0.1728           | 0.045   | 0.142             | 0.0406  | 0.1612           | 0.0575  | 0.1963           | 0.0661  |
| optical digit                    | 323.1978         | 1.184   | 324.0414          | 1.2508  | 324.5103         | 1.8357  | 331.7654         | 1.8095  |
| page blocks                      | 14.5953          | 2.6135  | 10.8555           | 3.1821  | 13.2088          | 3.7672  | 15.3054          | 3.7015  |
| pendigits                        | 597.5753         | 7.973   | 596.2356          | 10.0636 | 599.6309         | 11.7277 | 599.684          | 10.5186 |
| pima diabetes                    | 6.3228           | 0.1788  | 4.9712            | 0.1767  | 5.0455           | 0.2371  | 6.3992           | 0.269   |
| plant species leaves margin      | 0.6798           | 0.117   | 0.6807            | 0.1189  | 0.7446           | 0.1805  | 0.7451           | 0.1806  |
| plant species leaves texture     | 119.9449         | 1.9998  | 120.5406          | 2.0027  | 122.1148         | 3.2466  | 121.6378         | 3.2535  |
| ringnorm                         | 1350.8695        | 8.7521  | 1352.427          | 12.2458 | 1357.874         | 15.1354 | 1354.012         | 13.4766 |
| seeds                            | 0.174            | 0.0279  | 0.1727            | 0.0331  | 0.1849           | 0.0442  | 0.187            | 0.0406  |
| image segmentation               | 11.882           | 0.8685  | 10.3135           | 1.0592  | 10.6223          | 1.2304  | 12.1874          | 1.204   |
| spambase                         | 441.0555         | 8.4443  | 396.2303          | 13.5323 | 400.5978         | 13.9315 | 446.2259         | 12.7619 |
| spectf heart                     | 0.6437           | 0.0311  | 0.6436            | 0.0311  | 0.6632           | 0.0502  | 0.6646           | 0.0502  |
| landsat satellite                | 255.4462         | 5.5045  | 242.8039          | 15.0677 | 243.0923         | 11.9613 | 307.8871         | 9.4819  |
| Average                          | 130.4281         | 1.591225 | 127.811          | 2.480242 | 128.5894        | 2.680292 | 133.5344        | 2.405383 |
Table C.9: Training time of the AGGLO-2 algorithm

| Dataset                             | Longest distance | Shortest distance | Mid-max distance | Mid-min distance |
|-------------------------------------|------------------|-------------------|-----------------|------------------|
|                                     | w/t. lemma | w. lemma | w/t. lemma | w. lemma | w/t. lemma | w. lemma | w/t. lemma | w. lemma | w/t. lemma | w. lemma |
| balance scale                       | 0.3724     | 0.0207   | 0.3702     | 0.0206   | 0.3704     | 0.0207   | 0.3702     | 0.0207   |
| banknote authentication              | 20.4138    | 18.0293  | 20.9532    | 18.1422  | 20.5455    | 17.8608  | 20.6175    | 17.722   |
| blood transfusion                    | 0.5674     | 0.4925   | 0.8059     | 0.7081   | 0.7212     | 0.6478   | 0.6057     | 0.5409   |
| breast cancer wisconsin              | 0.3999     | 0.1324   | 0.4103     | 0.1335   | 0.3982     | 0.1358   | 0.3998     | 0.1319   |
| breast cancer coimbra                | 0.0233     | 0.0087   | 0.0234     | 0.0087   | 0.0233     | 0.0087   | 0.0233     | 0.0087   |
| climate model crashes                | 0.6912     | 0.0319   | 0.6951     | 0.0324   | 0.6944     | 0.032   | 0.6914      | 0.0323   |
| connectionist bench sonar            | 0.0678     | 0.0091   | 0.0676     | 0.0091   | 0.0678     | 0.0091   | 0.0679     | 0.0091   |
| glass                               | 0.1718     | 0.1578   | 0.1733     | 0.1623   | 0.1686     | 0.1585   | 0.1738     | 0.1613   |
| haberman                            | 0.1577     | 0.1375   | 0.1893     | 0.1689   | 0.1893     | 0.1674   | 0.1598     | 0.1387   |
| heart                               | 0.0212     | 0.0123   | 0.021      | 0.0123   | 0.0233     | 0.0139   | 0.0208     | 0.0123   |
| ionsphere                           | 0.4828     | 0.3937   | 0.5086     | 0.3863   | 0.4907     | 0.3896   | 0.4833     | 0.3781   |
| movement libras                     | 0.2215     | 0.2043   | 0.225      | 0.2049   | 0.2218     | 0.2046   | 0.2208     | 0.2036   |
| optical digit                       | 105.8615   | 1.2225   | 103.9479   | 1.2399   | 98.2648    | 1.2936   | 104.8599   | 1.249    |
| page blocks                         | 494.1239   | 422.8161 | 1409.174   | 1206.817 | 1156.35    | 1015.346 | 511.347    | 439.3924 |
| pendigits                           | 130.6623   | 38.7196  | 267.1413   | 158.6442 | 250.7539   | 157.6118 | 140.968    | 46.2601  |
| pima diabetes                       | 3.7218     | 2.3679   | 3.9888     | 2.6727   | 3.932      | 2.5885   | 3.7252     | 2.3731   |
| plant species leaves margin         | 0.3564     | 0.0967   | 0.3577     | 0.0966   | 0.3562     | 0.0966   | 0.3562     | 0.0967   |
| plant species leaves texture        | 40.4664    | 12.0927  | 40.5373    | 12.1132  | 40.6338    | 12.1003  | 40.4593    | 12.0958  |
| ringnorm                            | 1327.968   | 872.3743 | 1442.132   | 986.0059 | 1382.036   | 929.1679 | 1332.649   | 874.2278 |
| seeds                               | 0.3235     | 0.279    | 0.3286     | 0.2895   | 0.3259     | 0.2859   | 0.3197     | 0.279    |
| image segmentation                  | 62.3063    | 58.843   | 72.7162    | 69.1137  | 71.0587    | 67.3851  | 63.6872    | 60.2109  |
| spambase                            | 1506.906   | 1412.081 | 1642.779   | 1543.002 | 1621.651   | 1521.519 | 1516.833   | 1427.837 |
| spectf heart                        | 0.3101     | 0.0246   | 0.3102     | 0.0246   | 0.3101     | 0.0247   | 0.3102     | 0.0247   |
| landsat satellite                   | 237.7967   | 184.5571 | 965.5685   | 884.7285 | 908.0397   | 826.8151 | 291.1204   | 236.4152 |
| **Average**                         | **163.9331** | **126.046** | **248.8927** | **203.5307** | **231.5678** | **189.7451** | **167.8987** | **129.9926** |