International Journal of Information Technology and Language Studies (IJITLS)

Fuzzy Min Max Neural Network for pattern classification: An overview of complexity problem

Osama Nayel Al Sayaydeh and **Abeer Shamaileh**
osamasayaydeh@hotmail.com; abeershamyleh@yahoo.com

Faculty of Computer Systems & Software Engineering, Universiti Malaysia Pahang, Malaysia

Abstract. Over the last years, the pattern classification is considered one of the most significant domains in artificial intelligence (AI), because it shapes a fundamental in many diverse real live applications where the artificial neural networks (ANNs) and fuzzy logic (FL) are most extensively utilized in pattern classification. In order to construct an effective and robust classifier, researchers have invented hybrid systems that combine both FL and ANNs. The Fuzzy Min Max (FMM) neural network has been proven to be a robust classifier for handling pattern classification issues. Although FMM has several features, it suffers from several limitations. Thus, researchers have introduced a lot of improvements to beat the shortcomings of FMM neural network. This paper focuses on a complete review of developments carried out on FMM neural network for addressing the complexity problem in order to help new researchers in identifying the recent strategies used to address the complexity problem.

Keywords: *FMM, Pattern Classification, online learning, neural network.*

1. Introduction

Pattern classification refers to the ability to classify patterns into classes when a set of features has been determined. It has been the most active application field of AI in recent years such as: industry (Salahshoor, Kordestani, & Khoshro, 2010), behavior speech recognition (Hinton et al., 2012), behavior analysis (Ji, Xu, Yang, & Yu, 2013), weather forecasting (Isa, Omar, Saad, Noor, & Osman, 2010), medicine (Amato et al., 2013). Over previous years, many techniques have been introduced for pattern classification. Among the many numerous techniques that are consecrated to handling pattern classification problems, i.e., ANN and FL are the most popular.

ANNs are common techniques in tackling pattern classification issues. An ANN possesses a number of characteristics that recognize it from other techniques, for example, ability to learn from processes, tackle noisy data that are obtained from real-world environments and dealing with the nonlinear problem. However, most of ANNs suffer from a number of limitations, for example, the number of neurons in hidden layers and number of hidden layers need to be found by repetitive experimentation, ANNs are unable to extract the rules from trained data. To overcome ANNs limitations, researchers integrated the ANNs with FL to form the fuzzy neural network (FNN). This integration led to the generation of more robust classifiers designs, i.e., Fuzzy Artmap (FAM) (Gail A. Carpenter, Grossberg, & Reynolds, 1991), ANFIS Fuzzy (Jang, 1993), and FMM neural networks (Simpson, 1992).

In FNN, the input signals and/or connection weights and/or the outputs are fuzzy subsets or set of membership values to fuzzy set (Simpson, 1992)(Gail A. Carpenter et al., 1991). Where the FNN is considered one of the most popular hybrid methods because it combines the main characteristics of ANNs and FL such as the ability to learn and the ability to deal with imprecise data in constructing classifiers. However, most of FNN use offline/batch learning process for training, which makes the FNN suffers from the "Catastrophic forgetting" issue, which it is also named as stability-plasticity dilemma (McCloskey & Cohen, 1989; Robins, 1993).

Stability plasticity dilemma means the system requires plasticity for the combination of new information, but also requires stability in order to deny forgetting of the previous information (Gail A Carpenter & Grossberg, 1987; Simpson, 1992). To beat the issue of the stability-plasticity dilemma, FAM was introduced by supporting the online training process. Although the FAM is capable to address the stability-plasticity dilemma, it suffers from permits overlapping between hyperboxes belongs to different classes (Simpson, 1992). Therefore, the FMM for pattern classification problems was introduced by Simpson.

FMM neural networks have many salient features that make it an effective classifier in tackling pattern classification issues (Mohammed & Lim, 2015; Sayaydeh, Mohammed, & Lim, 2018; Simpson, 1992), i.e., online learning, overlapping classes, supporting hard and soft decision making, and single one pass operation. Although these features make FMM robust classifier, the FMM training process still suffering from several shortcomings. Therefore, over previous years, a number of FMM variants have been suggested in order to improve the FMM training process (Sayaydeh et al., 2018).

This paper provides a detailed review on FMM variants that were introduced over past years in order to handle the complexity problem, because the complexity, because the complexity problem is considered one of the most significant dilemmas that negatively affected the performance of the neural network, whenever the size of network is increased, the calculation and high computation cost increased. The significance of this review is that provides a clear compassion between FMM variants. This paper presented in five sections. Starting with the introduction, the subsequent sections cover the FMM neural network, an overview of FMM variants, discussion, and concluding remarks.

2. Fuzzy Min Max Neural Network

The FMM network was introduced by Simpson in 1992 for pattern classification(Simpson, 1992). The structure of FMM composed of three layers, i.e., input (F_A), output (F_C), and hidden layers (F_B). Where the number of nodes in input layer equal to the number of features, while the number of nodes in hidden layers is equal to the number of hyperboxes (fuzzy rules) that are generating through the training process. The number of nodes in the output layer is equal to the number of classes (Sayaydeh et al., 2018), Figure.1 illustrates the FMM network topology. Each hyperbox in FMM network represents a fuzzy set, and a hyperbox is categorized through a pair of min and max points in an n-dimensional space, as shown in Figure 2.

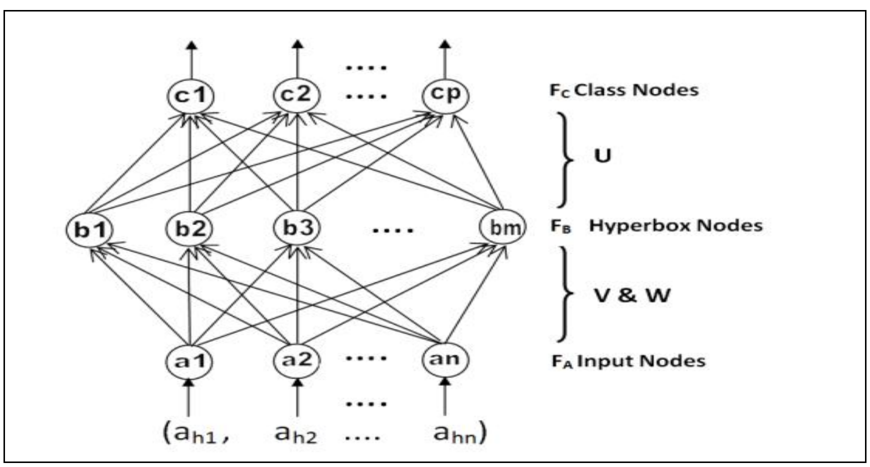

Figure 1. FMM structure.

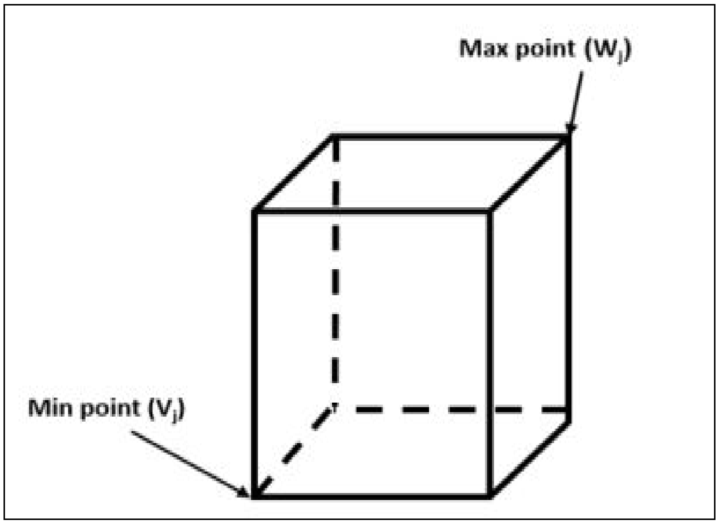

Figure 2. FMM hyperbox.

FMM utilizes the membership function to calculate the belongingness of an input sample with respect to jth hyperbox. The membership value ranged between 0 and 1. Where the input sample falls inside hyperbox has a membership value equal to 1. The belongingness degree of an input sample with respect to the jth hyperbox decreases whenever the distance between the jth hyperbox and input sample increases. Mathematically, the following equation is used to calculate the membership value:

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

Where $b_j(A_h)$ is the membership function; $A_h = (a_{h1}, a_{h2}, \ldots, a_{hn}) \in I^n$ is the h^{th} input sample; γ is the sensitivity parameter that controls how fast the membership value decreases as the distance between input pattern A_h and hyperbox B_j increases; and $V_j = (v_1, v_2, v_3, \ldots, v_n)$ and $W_j = (w_1\ \ w_2, w_3, \ldots, w_n)$ are the min and max points of the hyperbox, respectively.

In the FMM neural network, the hyperbox size is controlled using a user-defined parameter, named as expansion parameter (θ). A larger θ leads to yielding a larger hyperbox size, which as result, generates a smaller number of hyperboxes. In the FMM training process, the topology of a network is constructed by constructing hyperboxes in the hidden layer and adding the class node in the output layer. Generally, the FMM training process composed of three processes, i.e., expansion, overlap test, and contraction. Through the training stage, an input sample with its class is chosen. Next, a hyperbox from the same class that has the highest fitness value is chosen and expand to absorb the input sample. A new hyperbox is created, if the hyperbox fails to satisfy the expansion coefficient. In FMM, the overlap between hyperboxes belong to the same class is permitted, while overlapping between hyperboxes belong to various classes is prohibited. Then, the contraction process is activated when two hyperboxes from classes are overlapped. The FMM training stage can be explained as follows:

Expansion: prior to the expansion process starts, the membership degrees are computed to specify the winner hyperbox. The winner hyperbox is expanded to include the input sample if the expansion coefficient satisfies Eq. 2; otherwise, a new hyperbox is generated.

$$n\theta \geq \sum_{i=1}^{n} (\max(w_{ji}, a_{hi}) - \min(v_{ji}, a_{hi})) \quad (2)$$

Where θ is utilized to control the maximum size of the hyperbox. The hyperbox value range between 0 and 1. The min and max points of the winner hyperbox are modified through utilizing equations 3 and 4 if the Eq. 2 is met.

$$v_{ji}^{new} = \min(v_{ji}^{old}, a_{hi}) \tag{3}$$

$$w_{ji}^{new} = \max(w_{ji}^{old}, a_{hi}) \tag{4}$$

Overlap test: the overlap test is utilized to specify whether there is an overlap between hyperboxes that belong to various classes. FMM uses four overlap rules to determine the overlap area. The following overlap test rules:

Case 1:

$$V_{ji} < V_{ki} < W_{ji} < W_{ki}, \delta^{new} = \min(W_{ji} - V_{ki}, \delta^{old}) \tag{5}$$

Case 2:

$$V_{ki} < V_{ji} < W_{ki} < W_{ji}, \delta^{new} = \min(W_{ki} - V_{ji}, \delta^{old}) \tag{6}$$

Case 3:

$V_{ji} < V_{ki} < W_{ki} < W_{ji}$,

$$\delta^{new} = \min(\min(W_{ji} - V_{ki}, W_{ki} - V_{ji}), \delta^{old}) \tag{7}$$

Case 4:

$V_{ki} < V_{ji} < W_{ji} < W_{ki}$,

$$\delta^{new} = \min(\min(W_{ji} - V_{ki}, W_{ki} - V_{ji}), \delta^{old}) \tag{8}$$

Initially δ^{old} =1. If the difference between δ^{old} =1 and δnew is greater than 1, then $\delta^{new} = \delta^{old}$. The dimension has minimal overlap specified using Δ = i. The contraction process is not triggered, when Δ= –1, which means no overlapping issue.

Contraction: This process is initiated if all dimensions that belong to different classes are overlapped; hence the contraction process is utilized to eliminate the minimal overlap dimension. The contraction of rules of FMM neural network are as follows:

Case 1:

$$V_{j\Delta} < V_{k\Delta} < W_{j\Delta} < W_{k\Delta}, W_{j\Delta}^{new} = V_{k\Delta}^{new} = \frac{W_{j\Delta}^{old} + V_{k\Delta}^{old}}{2} \tag{9}$$

Case 2:

$$V_{k\Delta} < V_{j\Delta} < W_{k\Delta} < W_{j\Delta}, W_{k\Delta}^{new} = V_{j\Delta}^{new} = \frac{W_{k\Delta}^{old} + V_{j\Delta}^{old}}{2} \tag{10}$$

Case 3a: $V_{j\Delta} < V_{k\Delta} < W_{k\Delta} < W_{j\Delta}$ *and*

$$(W_{k\Delta} - V_{j\Delta}) < (W_{j\Delta} - V_{k\Delta}), V_{j\Delta}^{new} = W_{k\Delta}^{old} \tag{11}$$

Case 3b: $V_{j\Delta} < V_{k\Delta} < W_{k\Delta} < W_{j\Delta}$ *and*

$$(W_{k\Delta} - V_{j\Delta}) > (W_{j\Delta} - V_{k\Delta}), W_{j\Delta}^{new} = V_{k\Delta}^{old} \tag{12}$$

$Case\ 4a: V_{k\Delta} < V_{j\Delta} < W_{j\Delta} < W_{k\Delta}\ and$

$$(W_{k\Delta} - V_{j\Delta}) < (W_{j\Delta} - V_{k\Delta}), W_{k\Delta}^{new} = V_{j\Delta}^{old} \quad (13)$$

$Case\ 4b: V_{k\Delta} < V_{j\Delta} < W_{j\Delta} < W_{k\Delta}\ and$

$$(W_{k\Delta} - V_{j\Delta}) > (W_{j\Delta} - V_{k\Delta}), V_{k\Delta}^{new} = W_{j\Delta}^{old} \quad (14)$$

3. FMM variants

Even though the FMM neural network is considered an efficient online learning classifier, there is still a need to improve the FMM training stage (Mohammed & Lim, 2015). Researchers have introduced several FMM variants to improve the performance. But few researchers were interested in addressing the FMM learning limitations and the same time reducing the FMM complexity (number of hyperboxes). In this paper, we investigate the FMM variants that addressing the complexity problem, Figure.3 shows the strategies adopted by FMM variants to address the complexity problem. The FMM variants tackled the complexity problem are discussed in the following section.

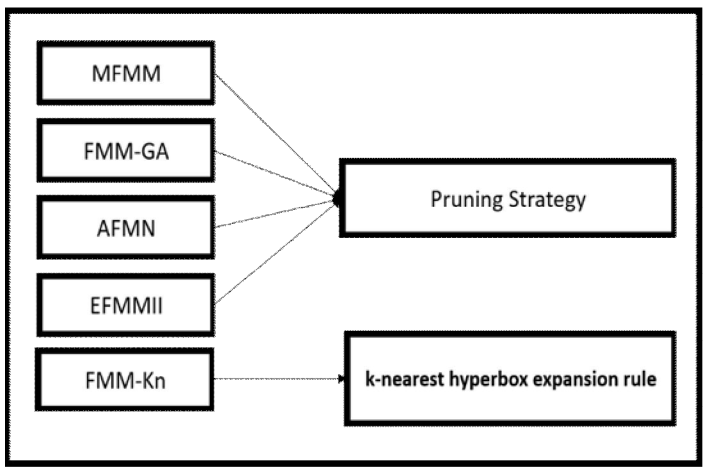

Figure 3. FMM variants.

Quteishat and Lim introduced a modified version of the original FMM, (MFMM) (A. Quteishat & Lim, 2008). MFMM improves the performance of original FMM when Θ is large. It utilizes the membership function and Euclidian distance to choose the winning hyperbox for predicting target output. MFMM also decreases the complexity (number of generated hyperboxes during the training stage) of FMM using a pruning strategy. After the FMM network is trained the number of hyperboxes generated during the training stage is reduced using the pruning strategy. Where the pruning strategy relies on confidence factor computed utilizing the following equation:

$$Cf_j = (1 - \gamma)U_j + \gamma A_j \quad (15)$$

Where A_j, U_j, γ are the accuracy of hyperbox j, the usage of hyperbox j, and the weighting factor, respectively. The value of Uj is defined as the number of patterns in the prediction set classified by any hyperbox j, divided by the maximum number of patterns in the prediction set classified by any hyperbox with the same classification class. While the A_j value is calculated as the number of correctly predicted set of patterns classified by any hyperbox j, divided by the maximum correctly classified patterns with the same classification class. The hyperbox with confidence factor lower than a user-defined threshold is pruned.

In 2010, Quteishat and Lim suggested a hybrid model that combines between FMM and Genetic Algorithm (FMM-GA) for pattern classification and rule extraction (Anas Quteishat, Lim, & Tan, 2010). The first phase is utilized to minimize the FMM complexity through using pruning strategy where the pruning strategy is used after the learning phase to reduce the number of generating hyperboxes during the learning phase by eliminating hyperboxes with low confidence factors. In the second phase, a "don't care" strategy is used by a GA rule extractor for minimizing the number of features in the extracted rules. FMM-GA was evaluated using UCI benchmark datasets and real medical dataset, the empirical results show that FMM-GA outperforms FMM and MFMM neural network.

In 2016, a Modified FMM neural network (MFMMN) was introduced for pattern classification(Shinde & Kulkarni, 2016). The main contribution of the MFMM network is the ability to deal with both discrete and continuous data at the same time. In MFMMN, the pruning strategy was utilized in order to reduce the network complexity.

Mohammed and Lim introduced a new enhancement on the original FMM classifier using a K-nearest (Kn) hyperbox expansion rule (Mohammed & Lim, 2017a). The Kn technique is utilized to decrease the FMM network complexity with marinating the classification accuracy. Kn introduced a new concept for choosing the winner hyperbox. Instead of choosing a single winner, the K-nearest hyperboxes capable of expanding and including the input sample without infringing equation (2) are chosen as the winners. Utilizing diverse UCI datasets, the introduced technique leads to reduce the FMM neural network complexity.

Mohammed and Lim further introduced an extension of EFMM, named as EFMM-II (Mohammed & Lim, 2017b). EFMM-II improved the performance of EFMM by utilizing two techniques, K-nearest hyperboxes and a pruning strategy. Where the Kn rule is utilized to minimize the number of constructed hyperboxes in the hidden layer. While the pruning strategy is utilized to minimize the impact of noise through deleting fragile hyperboxes that affected negatively on the performance of the EFMM. The empirical results show the effectiveness of the introduced techniques in enhancing the performance of EFMM in terms of network classification accuracy and complexity.

Although significant developments on the FMM classifier have been introduced in order to reduce the number of generated hyperboxes through the learning stage, FMM variants used the pruning strategy suffers from the using user-defined parameters which make the neural network less adaptive. Moreover, pruning strategy leads to remove significant parts of networks (knowledge) as they assume that have the small confidence factor, which may affect negatively on the classification rate (Augasta & Kathirvalavakumar, 2013).

4. Discussion

In general, the structure complexity term in FMM networks refers to the number of generated nodes (hyperboxes) in the hidden layer. As the number of hyperboxes increased the degree of computation cost increased. Hence, using a fewer hyperboxes number can reduce the computation cost (Ramos et al., 2008). Therefore, during the last decade, many models (MFMM, FMM-GA, MFMMN, EFMMII, FMM-Kn) have been presented to overcome the FMM network complexity problem. Most of those models, except FMM-Kn, focused on solving the complexity problem by using pruning techniques.

Using pruning techniques have been shown the ability to reduce the network structure by providing a comparable accuracy performance comparing to other non-pruned models. The dynamic process of pruning technique focuses on removing the created hyperboxes that have low knowledge structure, which identified as a hyperboxes that have confidence value below the user defined threshold factor. Despite the effectiveness of the technique to solve the problem complexity, it adversely affects the network structure by pruned a part of network knowledge, affect the learning process quality through using a part of learning sample for prediction process, and increase the number of user-defined parameters with random values initialization which make the network less adaptive. In order to solve these problems, Mohammed and Lim proposed the FMM-Kn model which is capable of overcoming the pruning problem (Mohammed & Lim, 2017a). FMM-Kn model uses a new technique to reduce the network structure by modifying the original expansion process and improving the way of specifying and selecting the winning hyperbox during the learning stage. That led to creating less number of more accurate hyperboxes, hence, improving the FMM classification accuracy.

5. Conclusion

In this paper, we review all the FMM variants that were introduced for addressing the complexity problem. We inferred that all the FMM variants used pruning strategy in order to reduce the network complexity suffer from two main limitations: make the neural network less adaptive by using tuning parameter, and the pruning may lead to remove significant knowledge from the neural network, consequently affect learning process quality through using a part of learning sample for prediction process. We conclude that the FMM-kn introduces a suitable strategy for overcoming the complexity problem, as well as maintaining the classification performance.

References

Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A., & Havel, J. (2013). Artificial neural networks in medical diagnosis. Journal of Applied Biomedicine, 11(2), 47-58. doi:https://doi.org/10.2478/v10136-012-0031-x

Augasta, M. G., & Kathirvalavakumar, T. (2013). Pruning algorithms of neural networks — a comparative study. Central European Journal of Computer Science, 3(3), 105-115. doi:10.2478/s13537-013-0109-x

Carpenter, G. A., & Grossberg, S. (1987). ART 2: Self-organization of stable category recognition codes for analog input patterns. Applied optics, 26(23), 4919-4930.

Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. Neural networks, 4(5), 565-588. doi:https://doi.org/10.1016/0893-6080(91)90012-T

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., . . . Sainath, T. N. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine, 29(6), 82-97.

Isa, I. S., Omar, S., Saad, Z., Noor, N. M., & Osman, M. K. (2010). Weather forecasting using photovoltaic system and neural network. Paper presented at the Computational Intelligence, Communication Systems and Networks (CICSyN), 2010 Second International Conference on.

Jang, J. S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man, and Cybernetics, 23(3), 665-685. doi:10.1109/21.256541

Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D convolutional neural networks for human action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(1), 221-231.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. In G. H. Bower (Ed.), Psychology of Learning and Motivation (Vol. 24, pp. 109-165): Academic Press.

Mohammed, M. F., & Lim, C. P. (2015). An enhanced fuzzy min-max neural network for pattern classification. IEEE Transactions on Neural Networks and Learning Systems, 26, 417-429. doi:10.1109/TNNLS.2014.2315214

Mohammed, M. F., & Lim, C. P. (2017a). Improving the Fuzzy Min-Max neural network with a K-nearest hyperbox expansion rule for pattern classification. Applied Soft Computing Journal, 52, 135-145. doi:10.1016/j.asoc.2016.12.001

Mohammed, M. F., & Lim, C. P. (2017b). A new hyperbox selection rule and a pruning strategy for the enhanced fuzzy min–max neural network. Neural networks, 86, 69-79. doi:10.1016/j.neunet.2016.10.012

Quteishat, A., & Lim, C. P. (2008). A modified fuzzy min-max neural network with rule extraction and its application to fault detection and classification. Applied Soft Computing, 8(2), 985-995. doi:10.1016/j.asoc.2007.07.013

Quteishat, A., Lim, C. P., & Tan, K. S. (2010). A modified fuzzy min–max neural network with a genetic-algorithm-based rule extractor for pattern classification. IEEE transactions on systems, man, and cybernetics-part a: systems and humans, 40(3), 641-650.

Robins, A. (1993). Catastrophic forgetting in neural networks: the role of rehearsal mechanisms. Paper presented at the Artificial Neural Networks and Expert Systems, 1993. Proceedings., First New Zealand International Two-Stream Conference on.

Salahshoor, K., Kordestani, M., & Khoshro, M. S. (2010). Fault detection and diagnosis of an industrial steam turbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro-fuzzy inference system) classifiers. Energy, 35(12), 5472-5482. doi:https://doi.org/10.1016/j.energy.2010.06.001

Sayaydeh, O. N., Mohammed, M. F., & Lim, C. P. (2018). A SURVEY OF FUZZY MIN MAX NEURAL NETWORKS FOR PATTERN CLASSIFICATION: VARIANTS AND APPLICATIONS. IEEE Transactions on Fuzzy systems, 1-1. doi:10.1109/TFUZZ.2018.2865950

Shinde, S., & Kulkarni, U. (2016). Extracting classification rules from modified fuzzy min–max neural network for data with mixed attributes. Applied Soft Computing, 40, 364-378. doi:https://doi.org/10.1016/j.asoc.2015.10.032

Simpson, P. K. (1992). Fuzzy min-max neural networks. I. Classification. IEEE Transactions on Neural Networks, 3(5), 776-786. doi:10.1109/72.159066