Application of the Fuzzy Min-Max Neural Networks to Medical Diagnosis

Anas Quteishat and Chee Peng Lim

School of Electrical & Electronic Engineering
University Science Malaysia, Malaysia
cplim@eng.usm.my

Abstract. In this paper, the Fuzzy Min-Max (FMM) neural network along with two modified FMM models are used for tackling medical diagnostic problems. The original FMM network establishes hyperboxes with fuzzy sets in its structure for classifying input patterns into different output categories. While the first modified FMM model uses the membership function and the Euclidian distance to classify the input patterns, the second modified FMM model employs only the Euclidian distance for the same process. Unlike the original FMM network, the two modified FMM models undergo a pruning process, after network training, to remove hyperboxes with low confidence factors. To assess the effectiveness of the three FMM networks in medical diagnosis, a set of real medical records from suspected Acute Coronary Syndrome (ACS) patients is collected and used for experimentation. The bootstrap method is used to analyze the results statistically. Implications of the experimental outcomes are discussed, and the potential of using the FMM networks a decision support tool for medical prognostic and diagnostic problems is demonstrated.

1 Introduction

In general, medical diagnosis is the process where physicians attempt to identify the disease a patient is suffering from based on the patient's physical symptoms, clinical tests data, and other related signs and information. In some cases, physicians may require additional opinion for disease diagnosis since some diseases have common symptoms and some are related to others. However, it is not always easy to obtain a second opinion or to statistically analyze the symptoms. This problem has inspired researchers to develop computerized medical diagnosis tools. Since medical diagnosis can be viewed as a pattern classification problem, artificial neural networks (ANN) have shown potentials as a tool to diagnose (classify) patients based on their symptoms. Indeed, ANNs have the ability to integrate data-based analytical techniques such as decision and classification theory, and to use knowledge-based approaches to provide useful information to support the decision making process [1]. From the literature, a lot of successful ANN applications to medical problems can be found. In [1], autonomously learning ANN models were used to classify real coronary care unit (CCU) patient records and trauma patient records into different categories. In [2] a pulmonary disease diagnostic system was proposed. The system used real clinical data to attempt to treat a whole category of distressed body organs. A model selection

* Corresponding author.

I. Lovrek, R.J. Howlett, and L.C. Jain (Eds.): KES 2008, Part III, LNAI 5179, pp. 548–555, 2008.

method that used the self-organizing map (SOM) for breast cancer diagnosis was demonstrated in [3]. In [4], the performances of a number of ANN models for breast cancer diagnosis were compared and analyzed. A neuromuscular disorder diagnosis system that employed two different ANN models (supervised and unsupervised) for analyzing features selected from electromyography (EMG) was presented in [5].

In this paper, use of ANN models to support the decision making process of suspected Acute Coronary Syndrome (ACS) patients is presented. A rapid and timely diagnosis of patients with suspected ACS is necessary as it could potentially lead to mortality if proper medical treatment is not provided to ACS patients as soon as possible. Specifically, the Fuzzy Min-Max (FMM) network [6], and two of its modified versions (MFMM1 and MFMM2) [7] were employed to classify in this work. The results obtained are analyzed statistically using the bootstrap method.

The paper is organized as follows. Section 2 describes the original FMM network. The modified versions of FMM are explained in section 3. The experiments using suspected ACS patient records along with the analysis for the results are presented section 4. A summary of this work is presented in section 5.

2 The Fuzzy Min-Max (FMM) Neural Network

FMM is an incremental learning system. It learns incrementally in a single pass through the data set. It refines the existing pattern classes as new information is received. It also has the ability to add new pattern classes online. In FMM, hyperboxes with fuzzy sets are formed to represent knowledge learned by the network. Learning in FMM comprises a series of expansion and contraction processes that fine-tune its hyperboxes to establish boundaries between classes. If overlapping hyperboxes of different classes occur in the input space, contraction is performed to eliminate the overlapping regions. The membership function is defined with respect to the minimum and maximum points of a hyperbox. It describes the degree to which a pattern fits in the hyperbox. The hyperboxes have a range from 0 to 1 along each dimension; hence the pattern space is an n-dimensional unit cube I^n. A pattern which is contained in the hyperbox has a unity membership function.

Mathematically, the definition of each hyperbox fuzzy set B_j [6] is defined by

$$B_j = \{X, V_j, W_j, f(X, V_j, W_j)\} \qquad \forall X \in I^n \tag{1}$$

where $X = (x_1, x_2, \ldots\ldots, x_n)$ is the input pattern, $V_j = (v_{j1}, v_{j2}, \ldots\ldots, v_{jn})$, $W_j = (w_{j1}, w_{j2}, \ldots\ldots, w_{jn})$ are the minimum and maximum points of B_j, respectively and $f(X, V_j, W_j)$ is the membership function. Figure 1 illustrates an example of the minimum and maximum points for a two-class problem where the universe of discourse is $\Re^3$.

Applying the definition of a hyperbox fuzzy set, the combined fuzzy set that classifies the K^{th} pattern class, C_k, is defined as

$$C_k = \bigcup_{j \in K} B_j \tag{2}$$

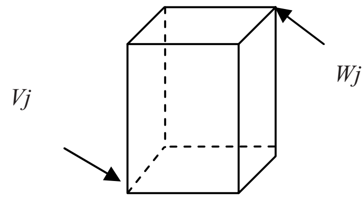

Fig. 1. A min-max hyperbox $B_j = \{V_j, W_j\}$ in $\Re^3$

where K is the index set of those hyperboxes associated with class k. One important property of this approach is that the majority of the processing is concerned with finding and fine-tuning the class boundaries. The learning algorithm of FMM allows overlapping of hyperboxes of the same class but eliminates overlapping of hyperboxes from different classes. The membership function for the j^{th} hyperbox $b_j(A_h)$, $0 \le b_j(A_j) \le 1$, measures the degree to which the h^{th} input pattern, A_h, falls outside hyperbox B_j [6]. As $b_j(A_h)$ approaches 1, the pattern is said to be more "contained" by the hyperbox. The resulting membership function is [6]:

$$b_j(A_h) = \frac{1}{2n}\sum_{i=1}^{n}\Big[max\big(0, 1 - max\big(0, \gamma\, min\big(1, a_{hi} - w_{ji}\big)\big)\big) + max\big(0, 1 - max\big(0, \gamma\, min\big(1, v_{ji} - a_{hi}\big)\big)\big)\Big] \tag{3}$$

where, $A_h = (a_{h1}, a_{h2}, ..., a_{hn}) \in I^n$ is the h^{th} input pattern, and γ is a sensitivity parameter that regulates how fast the membership value decreases as the distance between A_h and B_j increases.

The architecture of FMM consists of three layers of nodes, as shown in Figure 2. The first layer, F_A, is the input layer that contains input nodes equal in number to the number of dimensions of the input pattern. Layer F_C is the output layer. It contains

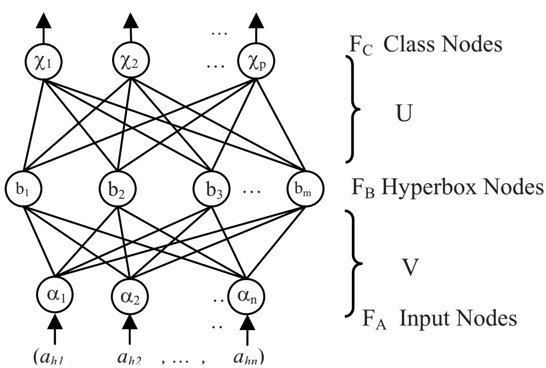

Fig. 2. A three layer FMM network

nodes equal in number to the number of classes. The middle or hidden layer, F_B, is called the hyperbox layer. Each F_B node represents a hyperbox fuzzy set, where F_A to F_B connections are the min-max points. The F_B transfer function is the hyperbox membership function defined by (3). The minimum and maximum points are stored in matrices V and W, respectively.

Note that the above provides an overview of the FMM network. Details of the FMM dynamics can be found in [6].

3 Modified Fuzzy Min-Max Neural Networks

In FMM, the size of a hyperbox is controlled by θ, which is varied between 0 and 1. When θ is small, more hyperboxes are created. When θ is large, the number of hyperboxes is small, and the size of hyperboxes is large. Therefore, when θ is large, a lot of hyperboxes may assume high membership function values for a new input pattern, in which the winner-take-all approach in determining the winning hyperbox may lead to inaccurate predictions by FMM.

To enhance the effectiveness of FMM, useful modifications are proposed in [7]. After the training stage, a pruning procedure is incorporated in FMM to reduce the number of hyperboxes, hence the network complexity. To assist pruning, a confidence factor is calculated using the method in [8]. A data set is divided into three: training set (for learning), prediction set (for pruning), and test set (for evaluation). The confidence factor for each hyperbox is based on its usage frequency and its predictive accuracy on the prediction set, as follows

$$CF_j = (1-\gamma)U_j + \gamma A_j \tag{4}$$

where U_j is the usage of hyperbox j, A_j is the accuracy of hyperbox j, and $\gamma \in [0,1]$ is a weighing factor. The value of U_j 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. On the other hand, the value of A_j is defined 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 confidence factor identifies hyperboxes that are frequently used and generally give high classification accuracy, as well as hyperboxes that are rarely used and, yet highly accurate. Hyperboxes with a confidence factor lower than a user-defined threshold is pruned.

Modification to the prediction process of FMM is also suggested. In original FMM, the input patterns are classified based on the membership function calculated using equation (3). The degree of membership for the input pattern is calculated against all hyperboxes created in the learning process. The pattern is classified to the class associated with the hyperbox that has the highest membership function value.

Based on the above modifications, two modified FMM (MFMM) models are proposed [7]. The first modified FMM (MFMM1) model uses both the membership function and the Euclidean distance to predict its target output. The second modified FMM (MFMM2) model uses the Euclidean distance alone to predict the target output.

For MFMM1, given a new input pattern, the membership function values of all hyperboxes are first calculated. A pool of hyperboxes that have high membership function values is then selected. The number of hyperboxes selected can be based on a user-defined threshold. After that, the Euclidean distances between the selected hyperboxes and the input pattern are computed. The hyperbox with the shortest Euclidean distance is chosen as the winner. Figure 3 shows the classification procedure of a two-dimensional input pattern using the proposed method. Suppose hyperboxes 1 and 2 are selected, and E_1 and E_2 are the distances between the input pattern and the centroids of hyperboxes 1 and 2, respectively. Since $E_2<E_1$, the input pattern is predicted to be in Class 2, even though it is contained in hyperbox 1. This modified FMM network, as shown in [7], is able to improve the classification results when the network has a small number of hyperboxes, and each with a large size (i.e. in situations where θ is large).

As for MFMM2, the Euclidian distance is calculated between the input pattern and all the hyperboxes. The hyperbox with the shortest distance is used to classify the input pattern. Details of the modifications can be found in [7].

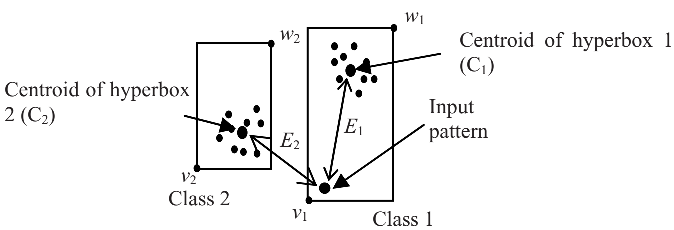

Fig. 3. The classification process of modified FMM models

4 Diagnosis of Acute Coronary Syndrome (ACS)

The ACS data set contained 118 real records of suspected heart attack patients admitted to Penang Hospital, Malaysia. After consulting with medical specialists, 16 features comprising clinical historical data, physical examination, ECG, and cardiac enzymes, and Troponin-T tests, were extracted from each record for this study. Out of the 16 features, 12 binary features were coded as 1, 0.25, and 0 for the presence, absence and missing values, respectively. Four real-valued features were age, degree of pain, duration of pain, and cardiac enzymes data. They were normalized between 0.25 and 1, and 0 was used to represent missing values. Among all the records, 103 patients were with ACS and 15 patients without ACS.

The experimental procedure was as follows. The data set was divided into three sub-sets: 60 samples for training, 23 samples for prediction, and 35 samples for test. Since the prediction data set was not needed for FMM, 83 samples were used for FMM training. The pruning threshold was set to 0.7 for both the modified FMM networks. Based on the 16 features, the FMM networks were employed to determine whether the patient suffered from ACS or otherwise. The experiments were repeated 10 times, each time the training data set was presented in a different sequence, and the average classification accuracy rates were computed. To further assess the results statistically, the bootstrap methods [9] was employed. The 95% confidence intervals of the results were estimated using 5000 bootstrap samples, and the *p*-value was used to quantify the performance statistically.

4.1 Results and Discussion

Figure 4 shows the average accuracy rates versus the hyperbox size, θ. The performance of the three networks was almost similar for small hyperbox sizes (θ < 0.1). However, as the hyperbox size increased, the FMM performance started to degrade. Compared with FMM, both modified FMM networks degraded in a slower manner. In general, MFMM1 yielded the best test accuracy rates, as shown in Figure 4. The highest result of MFMM1 was 85.43%.

The error bars in Figure 4 are the 95% confidence intervals of the average accuracy rates. Notice that the error bars of FMM did not overlap with those of MFMM1 and MFMM2, indicating that the FMM performances were statistically inferior to those of modified FMM networks.

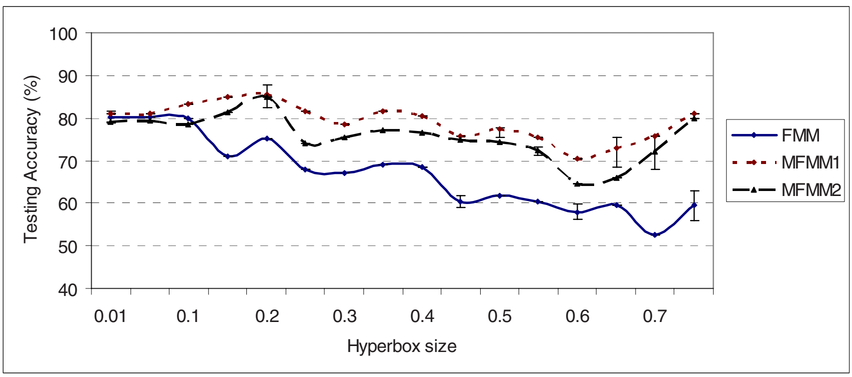

Fig. 4. Average test accuracy rate versus hyperbox size for the ACS data set

To further compare the performances of different FMM networks, the *p*-value hypothesis test was conducted. The null hypothesis stated that the average accuracy rates between FMM versus MFMM1, FMM versus MFMM2, and MFMM1 versus MFMM2 were equal (at the 95% confidence level). As shown in of Table 2 (column 2), the *p*-values of FMM versus MFMM1 are lower than the significance level ($\alpha = 0.05$) for $\theta \geq 0.15$. This means that the null hypothesis is rejected. In other words, there is a significant difference in terms of the test accuracy rates between FMM and MFMM1 for $\theta \geq 0.15$. The same trend (except for $0.55 \leq \theta \leq 0.65$) can be observed in Table 2 (column 3) for performance comparison between FMM and MFMM2. However, all the *p*-values between MFMM1 and MFMM2 are higher than 0.05, as shown in Table 2 (column 4), indicating that the performances of the two modified FMM networks are statistically equal at the 95% confidence level.

Table 1. The *p*-values for performance comparison between different FMM networks

Hyperbox size, θ	FMM vs. MFMM1	FMM vs. MFMM2	MFMM1 vs. MFMM2
0.05	0.8026	0.6016	0.562
0.1	0.0594	0.4322	0.0538
0.15	0	0.0004	0.266
0.2	0	0.0002	0.8034
0.25	0	0.0416	0.0278
0.3	0.004	0.011	0.43
0.35	0.0006	0.0256	0.2398
0.4	0.0138	0.0398	0.645
0.45	0.0094	0.005	0.9158
0.5	0.0082	0.0074	0.7446
0.55	0.0224	0.0962	0.5724
0.6	0.048	0.2962	0.4606
0.65	0.0298	0.3032	0.2598
0.7	0.001	0.0104	0.6104
0.75	0.027	0.0174	0.7416

5 Summary

This paper has presented an application of the FMM neural networks as a medical diagnosis tool. Two useful modifications have been incorporated into the FMM network: use of a pruning procedure to remove hyperboxes with low confidence factor from the network, and use of the Euclidian distance in the prediction process. Based on the modifications, two modified FMM networks have been introduced: MFMM1 uses both the membership value and the Euclidian distance for classifying the input pattern, while MFMM2 uses only the Euclidian distance for classification.

To assess the applicability of the FMM networks to medical diagnosis, a set of real patient records has been collected, and the FMM networks have been used to classify the patients into ACS and non-ACS categories. Stability of the results has been assessed and quantified using the bootstrap method. The results obtained have demonstrated that the modifications on the FMM network have improved the performances when the hyperbox size is high.

Future work will focus on using more data sets from different medical domains to further ascertain the capability and effectiveness of the modified FMM networks, so that they can be employed as a useful and usable decision support tool for medical prognostic and diagnostic tasks.

References

[1] Lim, C.P., Harrison, R.F., Kennedy, R.L.: Application of autonomous neural network systems to medical pattern classification tasks. Artificial Intelligence in Medicine 11, 215–239 (1997)

[2] Economou, G.P.K., Spiropoulos, C., Economopoulos, N.M., Charokopos, N., Lymberopoulos, D., Spiliopoulou, M., Haralambopulu, E., Goutis, C.E.: Medical diagnosis and artificial neural networks: a medical expert system applied to pulmonary diseases. In: Proceedings of the 1994 IEEE Workshop Neural Networks for Signal Processing, pp. 482–489 (1994)

[3] West, D., West, V.: Model selection for a Medical Diagnosis decision support system: A Breast Cancer Detection Case. Artificial Intelligence in Medicine 20, 183–204 (2000)
[4] Kıyan, T., Yıldırım, T.: Breast Cancer Diagnosis Using Statistical Neural Networks. In: International XII. Turkish Symposium on Artificial Intelligence and Neural Networks, pp. 754–760 (2003)
[5] Pattichis, C.S., Schizas, C.N., Middleton, L.T.: Neural network models in EMG diagnosis. IEEE Transactions on Biomedical Engineering 42, 486–496 (1995)
[6] Simpson, P.K.: Fuzzy Min-Max neural networks-Part 1: Classification. IEEE Transactions on Neural Networks 3, 776–786 (1992)
[7] Quteishat, A., Lim, C.P.: A modified fuzzy min–max neural network with rule extraction and its application to fault detection and classification. Applied Soft Computing 8, 985–995 (2008)
[8] Carpenter, G., Tan, A.: Rule extraction: From neural architecture to symbolic representation. Connection Science 7, 3–27 (1995)
[9] Efron, B.: Bootstrap methods: Another look at the jackknife. Annals of Statistics 7, 1–26 (1979)