

Typical Sample Selection and Redundancy Reduction for Min-Max Modular Network with GZC Function*

Jing Li¹, Baoliang Lu¹, and Michinori Ichikawa²

¹ Department of Computer Science and Engineering, Shanghai Jiao Tong University, 1954 Hua Shan Road, Shanghai 200030, China

jinglee@sjtu.edu.cn, blu@cs.sjtu.edu.cn

² Lab. for Brain-Operative Device, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama, 351-0198, Japan

Abstract. The min-max modular neural network with Gaussian zero-crossing function (M^3 -GZC) has locally tuned response characteristic and emergent incremental learning ability, but it suffers from high storage requirement. This paper presents a new algorithm, called Enhanced Threshold Incremental Check (ETIC), which can select representative samples from new training data set and can prune redundant modules in an already trained M^3 -GZC network. We perform experiments on an artificial problem and some real-world problems. The results show that our ETIC algorithm reduces the size of the network and the response time while maintaining the generalization performance.

1 Introduction

The min-max modular (M^3) neural network [1, 2] is an alternative modular neural network model for pattern classification. It has been used in real-world problems such as part-of-speech tagging [3] and single-trial EEG signal classification [4]. The fundamental idea of M^3 network is divide-and-conquer strategy: decomposition of a complex problem into easy subproblems; learning all the subproblems by using smaller network modules in parallel; and integration of the trained individual network modules into a M^3 network.

Using Gaussian zero-crossing (GZC) function [5] as a base network module, the M^3 network (M^3 -GZC) has locally tuned response characteristic and emergent incremental learning ability. But M^3 -GZC network remembers all the samples that have been presented to it. The space in memory and response time can not be satisfied when more and more samples become available, because the space complexity and time complexity of M^3 -GZC network are $O(n^2)$, here n is the total number of training samples.

In this paper, we introduce a novel method called Enhanced Threshold Incremental Check (ETIC) algorithm for selecting representative samples and pruning redundant modules for M^3 -GZC network. Based on ETIC, M^3 network adds

* This work was supported in part by the National Natural Science Foundation of China under the grants NSFC 60375022 and NSFC 60473040.

samples selectively. The size of the network will increase only when necessary and the corresponding response time can drop sharply.

2 ETIC Algorithm

2.1 Influence of Threshold Limits in M³-GZC Network

Gaussian zero-crossing discriminate function is defined by

$$f_{ij}(x) = \exp \left[- \left(\frac{\|x - c_i\|}{\sigma} \right)^2 \right] - \exp \left[- \left(\frac{\|x - c_j\|}{\sigma} \right)^2 \right], \tag{1}$$

where x is the input vector, c_i and c_j are the given training inputs belonging to class C_i and class C_j ($i \neq j$), respectively, $\sigma = \lambda \|c_i - c_j\|$, and λ is a user-defined constant.

The output of M³-GZC network is defined as following.

$$g_i(x) = \begin{cases} 1 & \text{if } y_i(x) > \theta^+ \\ \text{Unknown} & \text{if } \theta^- \leq y_i(x) \leq \theta^+ \\ -1 & \text{if } y_i(x) < \theta^- \end{cases} \tag{2}$$

where θ^+ and θ^- are the upper and lower threshold limits, and y_i denotes the transfer function of the M³ network for class C_i , which discriminates the pattern of the M³ network for class C_i from those of the rest of the classes.

From equation (2) we can see that the decision boundary can be easily controlled by adjusting θ^+ and θ^- . If a test sample is accepted by a M³-GZC network with high absolute values of θ^+ and θ^- , it will be accepted by the same network with lower absolute values of θ^+ and θ^- (as depicted in Fig.1). The threshold limit can be viewed as a degree of confidence of correct classification. When the M³-GZC net can classify a new sample with a high degree of confidence, it treats the sample as already learned, and will not change itself. While the network misclassifies or correctly classifies a new sample only in a low degree of confidence, it treats the sample as not learned or not learned well; and adds it. So the samples in accept domain and can be classified correctly will not be added to the net in future, and the size of the net can be guaranteed, will not expand if there is no new knowledge presented.

Depending on the important role of thresholds, ETIC chooses the representative samples from training data set. The algorithm can be used in two cases, one is that new training samples are available to the network; the other is that the network still has redundant samples in it, and need to be reduced. Inspired by Condensed Nearest Neighbor [6] and Reduced Nearest Neighbor [7], our algorithm in the two circumstance are called Condensed ETIC and Reduced ETIC, respectively.

2.2 Condensed ETIC Algorithm

When new training samples are available, Condensed ETIC algorithm stores samples misclassified by the current network. The network can be started from

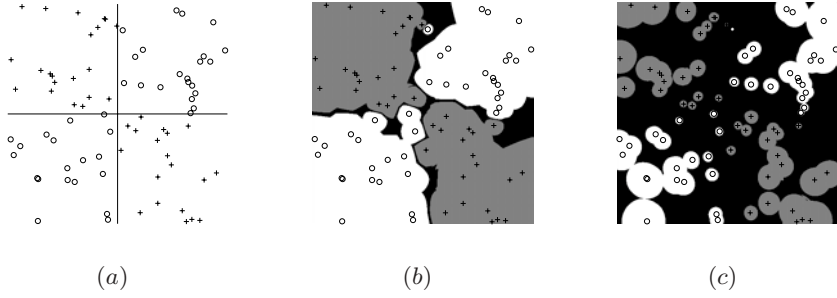


Fig. 1. Checkerboard problem and the corresponding decision boundary. The black denotes unknown decision regions. (a) A checkerboard problem; (b) decision boundary at $\theta^+ = 0.1$ and $\theta^- = -0.1$; (c) decision boundary at $\theta^+ = 0.8$ and $\theta^- = -0.8$.

scratch, or from a network that has been trained by previous training data. And the new training set can have only one sample or a batch of samples.

When a new training set S_{new} is presented to the network Net , the Condensed ETIC algorithm works as below:

1. Adjust θ^+ and θ^- to θ_e^+ and θ_e^- , ($|\theta_e^+| > |\theta^+|, |\theta_e^-| > |\theta^-|$).
2. $S = S_{new}$.
3. For each sample (x, d) in S :
 If (x, d) is misclassified by current network Net :
 $S = S \setminus (x, d)$, and add (x, d) to network Net .
4. If S is not exhausted and the network Net has been changed in step 3, go to step 3 for another loop. Or else restore θ^+ and θ^- to their previous values and return Net .

2.3 Reduced ETIC Algorithm

Although Condensed ETIC algorithm can filter out many samples during learning, there are still redundant samples in the network, especially after some groups of training samples are presented to the network. Then we use the Reduced ETIC to remove these samples. The fundamental idea of Reduced ETIC algorithm is to remove samples in the M^3 -GZC network if the removal does not cause any misclassification of other samples. Because M^3 -GZC network will not misclassify samples in the network, the algorithm only needs to check whether the removal causes any misclassification of all the removed samples.

When an already trained M^3 -GZC network has some redundant samples that need to be filtered out, the Reduced ETIC algorithm works as follows. Suppose $Net(S)$ denotes the network that has been set up based on training set S .

1. Adjust θ^+ and θ^- to θ_e^+ and θ_e^- , ($|\theta_e^+| > |\theta^+|, |\theta_e^-| > |\theta^-|$).
2. $S_{garbage} = \Phi$, $S_{net} =$ all samples in Net .

3. For each sample (x, d) in network Net :
 - If (x, d) is classified correctly by $Net(S_{net} \setminus (x, d))$, and all the samples in $S_{garbage}$ are also classified correctly by $Net(S_{net} \setminus (x, d))$, then $S_{garbage} = S_{garbage} \cup (x, d)$, $S_{net} = S_{net} \setminus (x, d)$.
4. Restore θ^+ and θ^- to their previous values and return Net .

2.4 Complexity Analysis

In worst cases, the Condensed ETIC algorithm will include only one sample in each loop, so the M³-GZC network will check n^2 times, and the time complexity is $O(n^2)$, where n denotes the number of samples in the new training set.

Suppose there are n samples in a trained network, and m samples will be filtered out by the Reduced ETIC Algorithm. So the network will check m samples at most to decide whether to remove one sample or not. The corresponding time complexity is $O(m \times n)$.

3 Experimental Results

In order to verify our method, we present three experiments. The first is an artificial problem and the other two are real-world problems. All the experiments were performed on a 2.8GHz Pentium 4 PC with 1GB RAM.

3.1 Checkerboard Problem

A checkerboard problem is depicted in Fig. 1 (a). The checkerboard divides a square into four quadrants. The points labelled by circle and plus are positive and negative samples, respectively. In this experiment, we randomly generate 400 samples as training data set, and another 400 samples as test data set. We use the Condensed ETIC to build M³-GZC network at different threshold. The results are listed in Table 1, and show that the lower the absolute value of thresholds, the smaller size and lower response time ratio, but the generalization performance can be guaranteed in a higher absolute value of thresholds.

Table 1. Results of checkerboard problem. The value in column threshold is θ_e^+ , while the corresponding $\theta_e^- = \theta_e^+$. ‘no check’ means using the traditional algorithm to build up M³-GZC network. The unit of ‘Time’ is *ms*.

Threshold	Accuracy	Unknown	False	Size	Time	Size Ratio	Time Ratio
no check	94.75%	1.00%	4.25%	400	13570	100%	100%
0.01	92.75%	4.50%	2.75%	77	1458	19.25%	10.74%
0.02	92.75%	4.50%	2.75%	80	1489	20.00%	10.97%
0.05	94.00%	3.00%	3.00%	85	1689	21.25%	12.45%
0.08	94.50%	2.00%	3.50%	95	2105	23.75%	15.51%
0.1	94.75%	2.25%	3.00%	97	2433	24.25%	17.93%

3.2 UCI Database

In this experiment, our algorithm is tested on five benchmark data sets from the Machine Learning Database Repository[8]. At start up, all the training samples are presented to an empty network, the Condensed ETIC algorithm and traditional algorithm are used respectively to build the network. The results are shown in Table 2. We can see that our Condensed ETIC algorithm will reduce the size of the network and speed up the response time of trained network, while the generalization ability is influenced only a little.

Table 2. Results on UCI database. Parameters of each net: $\lambda = 0.5$; $\theta^+ = 0.01$; $\theta^- = -0.01$; $\theta_e^+ = 0.1$; and $\theta_e^- = -0.1$.The unit of ‘Time’ is *ms*.

Data set	With ETIC	Accuracy	Unknown	False	Size	Time	Size Ratio	Time Ratio
balance	Y	92.0%	0.0%	8.0%	299	2510	59.8%	43.5%
	N	92.0%	0.0%	8.0%	500	5767		
car	Y	62.15%	34.14%	3.70%	436	36392	50.5%	33.7%
	N	57.87%	42.13%	0.0%	864	107878		
image	Y	82.0%	9.24%	8.76%	122	12035	58.1%	40.5%
	N	84.0%	7.33%	8.67%	210	29730		
Iris	Y	94.67%	1.33%	4.0%	36	125	48.0%	49.6%
	N	94.67%	1.33%	4.0%	75	252		
optdigits	Y	96.05%	2.62%	1.34%	1257	840613	32.9%	11.1%
	N	97.22%	1.45%	1.34%	3823	7548237		

3.3 Industry Image Classification

The database of this experiment comes from the images of glass in a product line. The images are converted into 4096 dimension vectors as training and test data. All of the data are divided into four groups; the number of data in each group is 1149, 1138, 1133 and 1197, respectively. We use the first to third data sets as the training data, and the fourth as the test data. At first, the first data set was presented to an empty network, and net_1 was built. Then the second data set was presented to net_1 , and net_2 was built. At last the third data set was presented to net_2 , and net_3 was built. We do this experiment in two ways, one is using our Condensed ETIC algorithm, and the other is using traditional way. We also use the Reduced ETIC algorithm to reduce the size of the final network. The results are listed in Table 3, and show that the generalization performance becomes better and better when new training data sets are available. And the Reduced ETIC algorithm can prune redundant samples in the network.

4 Conclusions

In this paper we have presented a novel algorithm, called ETIC (Enhanced Threshold Incremental Check), which can select representative samples from

Table 3. Results of industry images classification. Parameters of each net: $\lambda = 1$; $\theta^+ = 0.01$; $\theta^- = -0.01$; $\theta_e^+ = 0.5$; and $\theta_e^- = -0.5$. The net_4 denotes the network built up according to Reduced ETIC algorithm from net_3 . The unit of ‘Time’ is s.

Net	With ETIC	Accuracy	Unknown	False	Size	Time	Size Ratio	Time Ratio
net_1	Y	69.17%	0.08%	30.74%	33	20.3	2.87%	6.49%
	N	69.17%	0.00%	30.83%	1149	313.3		
net_2	Y	88.05%	0.0%	11.95%	550	119.4	24.05%	18.27%
	N	86.63%	0.33%	13.03%	2287	653.5		
net_3	Y	88.30%	0.25%	11.45%	1870	1148.6	54.68%	52.49%
	N	86.97%	0.50%	12.53%	3420	2188.2		
net_4	Y	88.30%	0.25%	11.45%	1764	1087.3	51.58%	49.69%
	N	86.97%	0.50%	12.53%	3420	2188.2		

new training data set and can prune redundant modules of an already trained M^3 -GZC network. Using ETIC, the M^3 -GZC network still has the locally tuned response characteristic and emergent incremental learning ability, and its size will not increase if there is no new knowledge presented. Several experimental results indicate that our ETIC algorithm can reduce the size of the network and the response time while maintaining the generalization performance.

References

1. Lu, B.L., Ito, M.: Task Decomposition Based on Class Relations: a Modular Neural Network Architecture for Pattern Classification. Lecture Notes in Computer Science. Vol. 1240. Springer (1997) 330-339
2. Lu, B.L., Ito, M.: Task Decomposition and Module Combination Based on Class Relations: A Modular Neural Network for Pattern Classification. IEEE Trans. Neural Networks, **10** (1999) 1244-1256
3. Lu, B.L., Ma, Q., Ichikawa, M., Isahara, H.: Efficient Part-of-Speech Tagging with a Min-Max Modular Neural Network Model. Applied Intelligence, **19** (2003) 65-81
4. Lu, B.L., Shin, J., Ichikawa, M.: Massively Parallel Classification of Single-Trial EEG Signals Using a Min-Max Modular Neural Network. IEEE Trans. Biomedical Engineering, **51** (2004) 551-558
5. Lu, B.L., Ichikawa, M.: Emergent On-Line Learning with a Gaussian Zero-Crossing Discriminant Function. IJCNN '02, **2** (2002) 1263-1268
6. Hart, P.E.: The Condensed Nearest Neighbor Rule. IEEE Trans. Information Theory, **14** (1968) 515-516
7. Gates, G.W.: The Reduced Nearest Neighbor Rule. IEEE Trans. Information Theory, **18** (1972) 431-433
8. Murphy, P.M., Aha, D.W.: UCI Repository of Machine Learning Database. Dept. of Information and Computer Science, Univ. of Calif., Irvine (1994)