ROBUST MODULAR ARTMAP FOR MULTI-CLASS SHAPE RECOGNITION

Chue Poh Tan, Chen Change Loy, Weng Kin Lai, Chee Peng Lim

Abstract— **This paper presents a Fuzzy ARTMAP (FAM) based modular architecture for multi-class pattern recognition known as Modular Adaptive Resonance Theory Map (MARTMAP). The prediction of class membership is made collectively by combining outputs from multiple novelty detectors. Distance-based familiarity discrimination is introduced to improve the robustness of MARTMAP in the presence of noise. The effectiveness of the proposed architecture is analyzed and compared with ARTMAP-FD network, FAM network, and One-Against-One Support Vector Machine (OAO-SVM). Experimental results show that MARTMAP is able to retain effective familiarity discrimination in noisy environment, and yet less sensitive to class imbalance problem as compared to its counterparts.**

I. INTRODUCTION

Binary classification refers to the categorization of data in the feature space into two regions. In contrast, multi-class pattern recognition is a task to classify the feature space into more than two regions using a set of discriminate functions, in which each region corresponds to a pattern class. More specifically, given n vectors of instances $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ drawn from feature space $\boldsymbol{\Omega}$, a multi-class classifier has to classify the inputs into k pre-defined classes $\boldsymbol{C} = (C_1, C_2, \ldots, C_k)$, where $C_a \neq C_b$ for $a \neq b$ and $k > 2$.

A number of studies have been carried out in the area of multi-class classification [1][2][3]. At present there are two main approaches that can be used to extend a binary classifier for multi-class problems. The first approach is by assembling several binary classifiers to form a recognition network. Examples of this type of implementation are one-against-all, one-against-one, one-against-higher-order, and P-against-Q, where P and Q are greater than one [1]. Apart from deploying multiple binary classifiers, another approach is to optimize a single classifier to recognize multiple classes. Note that it is computationally more expensive in solving a multi-class problem as compared with two-class problem by using the same amount of data, since multi-class classification involves an ensemble of several binary classifiers or more complex optimization. In other words, multi-class classification is not merely a trivial extension from binary classification because multi-class classification may involve more complex boundary formation and optimization. Typical differences in constructing a multi-class classifier and binary classifier are the network architectures, encoding schemes, and training methodologies [1], where each of them has a high influence to the network accuracy and computational speed.

This paper proposes a modular ARTMAP (MARTMAP), an extension of Fuzzy ARTMAP (FAM) [4] neural network for multi-class pattern recognition. It inherits the unique features from FAM, such as fast convergence and incremental learning capability. The proposed network is based on novelty detection approach, which differs from conventional approaches, such as one-against-one, one-against-all and strategies aforementioned. Basically MARTMAP is built up with multiple novelty detectors that are modeled independently by using only single class of information. For each novelty detector, a distance-based familiarity function is introduced to determine whether an unknown pattern is “familiar” to the respective pattern class or not. The outputs from all novelty detectors are then aggregated to form a collective decision on the class membership.

In this study, three data sets are employed in the performance evaluation. Apart from comparing with the original FAM, the performance of the proposed method is studied side-by-side with another similar ART-based neural network known as ARTMAP-FD [5]. The paper also compared the proposed method with the state-of-the-art technique, a multi-class SVM that implementing one-against-one strategy (OAO-SVM). It was chosen because a number of studies have reported that OAO-SVM gives good results as compared to other multi-class classification algorithms [3].

The organization of the paper is as follows. A detail explanation of the proposed method is provided in Section II. The data sets used in this study is described in Section III. The experimental results are reported and discussed in Section IV. Finally, Section V concludes the paper with some suggested future works.

Manuscript received December 14, 2007. This work was supported in part by the MIMOS Berhad, Malaysia.

C. P. Tan is with the MIMOS Berhad, Malaysia (corresponding author) phone: +6 03-89965000 ext.: 4251; mobile: +6 013 3322553; fax: +6 03-86579486; e-mail: chue.poh@mimos.my).

C. C. Loy was with MIMOS Berhad, Malaysia. He is now with the QueenMary University, United Kingdom. (e-mail: ccloy225@gmail.com).

W. K. Lai is with MIMOS Berhad, Malaysia (e-mail: lai@mimos.my)

C. P. Lim is with Faculty of Engineering, University of Science Malaysia, Malaysia (email : cplim@eng.usm.my).

978-1-4244-1821-3/08/$25.00©2008 IEEE

II. METHODOLOGY

A. Modular ARTMAP (MARTMAP)

This section will commence with an overview of the MARTMAP architecture and followed by more details on the underlying algorithms. As can be seen from Fig. 1, a MARTMAP is built up with multiple novelty detectors to form a recognition network. These novelty detectors are basically modified version of FAM to perform one class classification. Each novelty detector should be able to identify novel patterns that it is not aware of during training [6]. In the training stage, each novelty detector is trained separately by using data from single pattern class, i.e., i^{th} novelty detector is trained with all the instances from i^{th} class. Therefore, the number of novelty detector is proportional to the number of pattern class: for a *k*-class problem, *k* novelty detectors are required. In the prediction stage, the unknown pattern is fed into individual novelty detector. Each novelty detector then computes a affinity score to measure how familiar the unknown pattern to the class it recognized during training. The affinity score of each novelty detector serves as the inputs to a decision layer which make the final decision. Note that there is no direct connectivity between each novelty detector. Thus, new novelty detector can be added or the existing one can be removed from the MARTMAP classification module as the need arises, without affecting other trained novelty detectors.

The decision layer in MARTMAP plays an important role as it decides the final classification result based on the output from individual novelty detector. There are many different implementations that are suitable for the decision layer. It can be a simple rule-based classifier (e.g. max-win, min-win strategy) or neural networks depending on the complexity and the nature of the applications. For certain types of applications, the decision layer can make use of historical prediction results to increase the prediction hits. For instance, time series prediction such as object recognition in video stream, the decision layer can make use of past prediction results to form the final decision. In this case, the decision layer performs classification based on estimation from multiple hypotheses [7], whereby outputs from the novelty detectors are used to construct a decision histogram that records the classification hypotheses. Hypotheses are accumulated and averaged over a period of time and the final classification result is derived from the histogram. A threshold can be set to prevent the decision layer of making any meaningless guess before the estimation achieves certain confidence level. This method is able to reduce the generalization error by accumulating a classifier's predictions over time.

Similar to the architecture of FAM as shown in Fig. 2, each novelty detector in MARTMAP consists of two fuzzy Adaptive Resonance Theory (ART) modules designated as ART_a and ART_b, which create stable recognition categories in response to arbitrary sequences of input patterns [3]. Both ART modules are linked together by a map field module, F^{ab}, an associative learning network to establish an association between input patterns and target classes $\boldsymbol{C}$.

Similar to FAM, there are two key parameters that influence the performance of MARTMAP. The first parameter is base vigilance parameter $\bar{\rho}_a \in [0, 1]$ which determine the category formation of the network. Less categories are formed by using lower $\bar{\rho}_a$, which in turn leads to more generalized boundary. In the contrary, higher $\bar{\rho}_a$ will leads to firmer category formation and the close boundary is tighter. The learning parameter, $\beta_a \in [0, 1]$, determines the learning modes of the network. There are two learning modes: fast learning ($\beta_a = 1$ for all times) and fast-commit slow recode learning ($\beta_a = 1$ for an uncommitted node and $\beta_a < 1$ for a committed node).

The training stage of individual novelty detector in MARTMAP is identical to FAM. The following is a brief explanation on the typical operation in ART_a, which also occurs in ART_b. Initially, in the training stage, the original *M*-dimensional input vector $\boldsymbol{a}$ is complement-coded into a *2M*-dimensional vector $\boldsymbol{A}$:

$$\boldsymbol{A} = \left(\boldsymbol{a}, \boldsymbol{a}^c\right) \equiv \left(a^1, \ldots, a^M, 1 - a^1, \ldots, 1 - a^M\right) \tag{1}$$

$\boldsymbol{A}$ is propagated from the input layer F_1^a to the dynamic output layer F_2^a through a set of adaptive weights $\boldsymbol{w}^a$. Activation of the j^{th} F_2^a node is determined by the choice function $T_j(\boldsymbol{A})$ as defined in Equation (2), with $\boldsymbol{w}_j^a$ denoting the category weight vector of the j^{th} F_2^a node.

$$T_j(\boldsymbol{A}) = \frac{\left|\boldsymbol{A} \wedge \boldsymbol{w}_j^a\right|}{\alpha_a + \left|\boldsymbol{w}_j^a\right|} \tag{2}$$

According to the winner-take-all strategy, the node with the highest response value, denoted as node *J*, is selected as the winning node, while all other nodes $j \neq J$ are deactivated. The winning node *J* remains active if the match function of the chosen category meets the vigilance criterion:

$$\frac{\left|\boldsymbol{A} \wedge \boldsymbol{w}_J^a\right|}{\left|\boldsymbol{A}\right|} \geq \rho_a \tag{3}$$

Fig. 2 shows the architecture of Fuzzy ARTMAP (FAM) neural network. Each novelty detector in MARTMAP is equivalent to a modified FAM. If the vigilance test is satisfied, the network will proceed to the map field association. However, if the existing winning node fails to predict the output class, i.e., $c(J) \neq C$, a match tracking process is triggered until the best winning node that satisfies both the ART_a and map field vigilance test is found. Subsequently, learning takes place by updating the category weight vector of the winning node *J* in ART_a according to Equation (4). The process aforementioned is repeated until

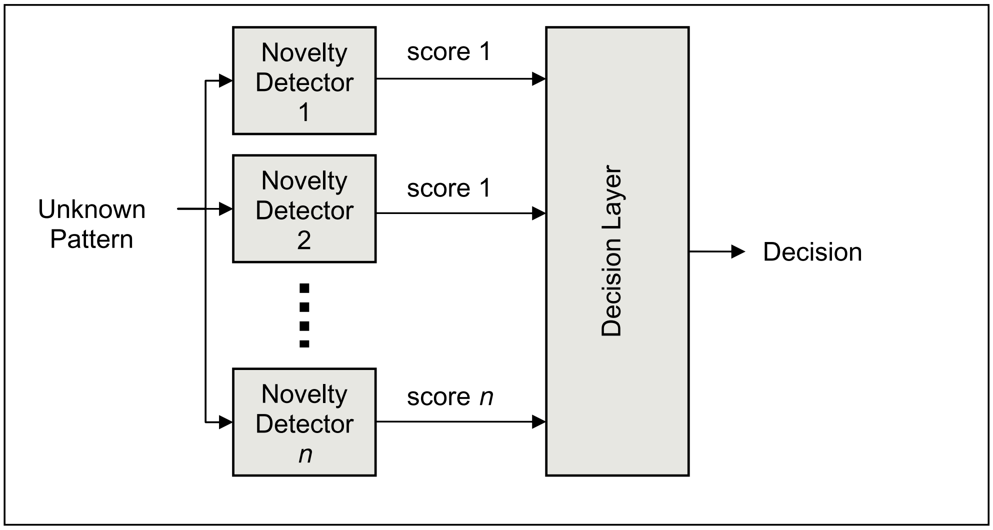


Fig. 1 – Architecture of modular ARTMAP (MARTMAP)

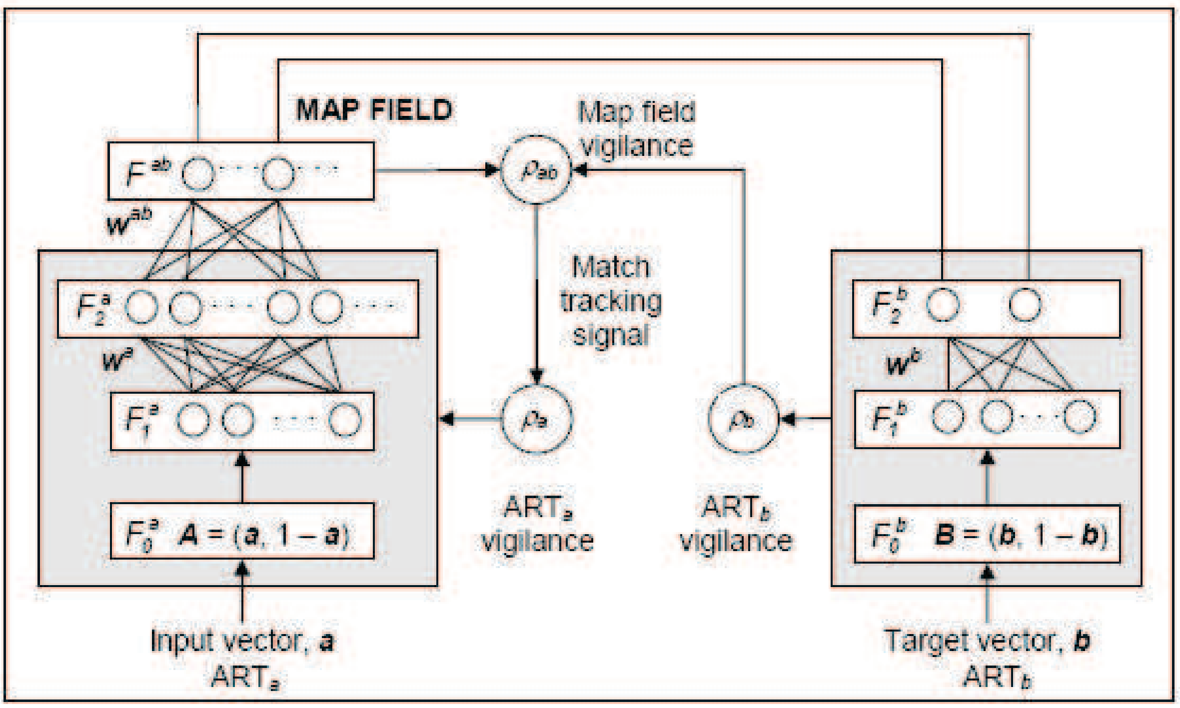


Fig. 2 – The figure shows the architecture of Fuzzy ARTMAP (FAM) neural network. Each novelty detector in MARTMAP is equivalent to a modified FAM

the novelty detector learns all the training instances assign to it. At the end of the training stage, each novelty detector would have at least one category that code all the instances of the respective class C. Sometimes, instances from a class may be coded by several categories. This is mainly due to the underlying distribution and internal structure of the training data space. But again, the number of category formed can be also controlled by the selection of ρ_a.

$$\boldsymbol{w}_J^{a(new)} = \beta_a\left(\boldsymbol{A} \wedge \boldsymbol{w}_J^{a(old)}\right) + \left(1-\beta_a\right)\boldsymbol{w}_J^{a(old)} \tag{4}$$

The prediction phase is divided into two stages: the first stage is the internal competition of categories within a novelty detector, whereas the second stage is the competition among novelty detectors. Initially, an unknown pattern is presented to every novelty detector. In the first stage, the response of each category to the unknown pattern is measured using the Equation (2). The node that has the highest response, denoted as node J, is selected as the winning node. All other nodes $j \neq J$ are deactivated in accordance with the winner-take-all competition. As a result, each novelty detector would have a winner category that can join the subsequent competition.

In the second stage, a familiarity function is used to measure the familiarity of a novelty detector to the new input pattern. The familiarity function is basically the Euclidean distance between the input pattern to the centroid of the winner categories. The resulting distances are then transmitted as the affinity scores to the decision layer. If the input pattern appears more "familiar" to the novelty detector, the distance value would be smaller. Therefore, "min-win" strategy is used in the decision layer, the i^{th} novelty detector with the smallest Euclidean distance will be selected as winner in the second stage, and the class label C_i will be assigned to the unknown input pattern. In the case where two novelty detectors give the same smallest value, the decision layer is forced to make a decision by selecting the novelty detector with lower index as the winner.

$$\boldsymbol{w}_j^{a\text{-}c} = (w_{j1}^{a\text{-}c}, w_{j2}^{a\text{-}c}, \ldots, w_{jM}^{a\text{-}c}) \tag{5}$$

Note that the dimension of the centre weight vectors covers only the original dimension of the input space. At the beginning, the centre weight vectors are initialised to zero. When learning takes place, the centre weight vectors of the J^{th} winning node are updated as follows, where N_{inputs} denotes the number of inputs that the category has coded.

$$\left(\boldsymbol{w}_J^{a\text{-}c}\right)^{new} = \left(\boldsymbol{w}_J^{a\text{-}c}\right)^{old} + \frac{1}{N_{inputs}}\left(\boldsymbol{a} - \left(\boldsymbol{w}_J^{a\text{-}c}\right)^{old}\right) \tag{6}$$

In order to compute the Euclidean distance, a new set of weight vectors is introduced in the ART_a module of MARTMAP called the centre weight vectors,

$$dist\left(\boldsymbol{a}, \boldsymbol{w}_J^{a\text{-}c}\right) = \sqrt{\sum_{i=1}^{M}\left(a_i - w_{Ji}^{a\text{-}c}\right)^2} \tag{7}$$

There is another ART-based neural network called ARTMAP-FD which is similar to the novelty detector in MARTMAP. ARTMAP-FD is an extension of FAM network with improvement in performing novelty detection. In contrast to the Euclidean distance-based method proposed in this paper, the familiarity function in ARTMAP-FD is the choice function expressed in Equation (2). Although novelty detector in MARTMAP and ARTMAP-FD are both trained with local knowledge, using choice function as familiarity measurement, however, may easier to prone to classification errors in the presence of noise and outliers, and may face classification uncertainty in overlapped boundaries.

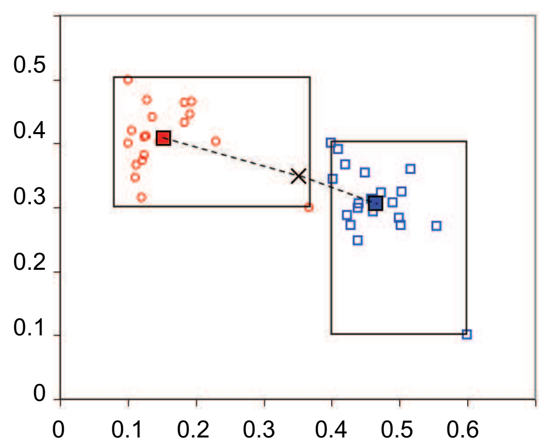
Fig. 3 gives an example of two decision boundaries generated from two novelty detectors. A decision boundary basically is a hyper-rectangular R_J formed in F_2^a category to enclose all the data points fall in that particular class region.

Fig. 3a shows two decision boundaries that are corrupted with noisy data located at the lower right corner of the hyper-rectangular. As can be seen from

Fig. 3a, although the unknown pattern X is nearer to the right hyper-rectangular labeled as C_b, it is however enclosed by the hyper-rectangular labeled as C_a due to unwanted noise in the training data. If choice function was used as familiarity discriminate function, X will be classified as C_a instead of C_b, which clearly an error in classification. Such error can be mitigated by measuring the Euclidean distance between X and the clusters' centroid as depicted in

Fig. 3a, where the distances are denoted as d_1 and d_2, respectively. As a result, pattern X is classified as C_b since $d_2 < d_1$.

Fig. 3b illustrates an unknown pattern X located inside two overlapped decision boundaries. In this case, more than one novelty detector will claim the unknown pattern as their pattern class, which causes classification ambiguity. Again, by using Euclidean distance, one can easily discriminate the point X from two overlapping clusters.



(a)

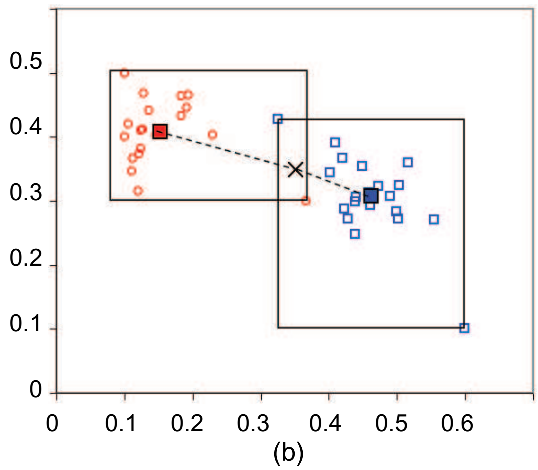


Fig. 3 – Classification ambiguities caused by outliers in training data

III. DATA USED

In this study, three data set were used, namely, Gaussian data set, Gaussian dataset corrupted with noise and shape dataset. The three data sets were used to evaluate the performance of the classifiers.

The synthetic data (hereafter called as Gaussian data set) as depicted in Fig. 4 was drawn from four overlapping Gaussian distributions centered at different mean but with the same standard deviation $\sigma = 0.2$, resulting a total of 250 data points for each distribution. The main purpose of generating the Gaussian data set was to simulate a data space with four classes, so as to examine the ability of a classifier in separating the classes during prediction. The second data set was derived from the original Gaussian data set with 15% of the training data corrupted with zero-mean Gaussian noise with a variance of 1.0.

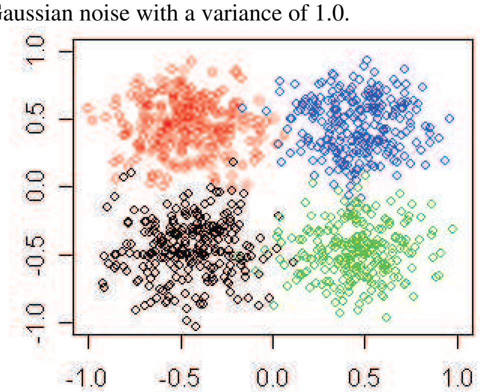


Fig. 4 – Synthetic data generated from four Gaussian distributions

The third data set consists of 193 shapes in eight categories. The dataset were selected from [8] and the MPEG-7 test database. Six simple features were extracted from the images in this study, namely dispersedness, compactness, axis ratio of a fitted ellipse, roughness, occupancy, and ratio of squared hull perimeter to hull area. The data set was characterized as imbalanced data set because some of the classes are represented by significantly more number of instances compared with other classes. For instance, the number of instances in class 1 (bird) was four times more than the number of instances in class 4 (car).

Fig. 5 – The figure shows some sample shapes in the shape data set

III. RESULTS AND DISCUSSION

A. *Comparison with FAM and ARTMAP-FD*

The objective of the experiments is to compare the performance of MARTMAP with ARTMAP-FD and FAM. Random sample cross-validation was employed in this study, whereby a hundred random partitions were generated by partitioning the original data set into 80%/20% training/testing sets randomly. Bootstrapping method was implemented to compute the confidence intervals (CI) for the performance metrics. The accuracies obtained in each sub-experiment were bootstrapped into 1000 samples. The average of the estimated accuracy along with the 95% CI was then reported. Fast learning approach (β_a = 1) was adopted throughout the experiments. Base vigilance parameter $\bar{\rho}_a$ was changed from low value ($\bar{\rho}_a$ = 0.00) to high value ($\bar{\rho}_a$ = 0.90) in order to examine the effect of this parameter. Note that $\bar{\rho}_a$ = 1.00 is not applicable in this study as this value causes the classifier to assume each input is essentially from different classes.

The results obtained by using the data set are shown in Fig. 6. As can be seen from Fig. 6, the performance of ARTMAP-FD and FAM remained stable across different values of $\bar{\rho}_a$. Although MARTMAP had a sudden drop in accuracy at $\bar{\rho}_a$ = 0.50 and degraded gradually when $\bar{\rho}_a$ was further increased to 0.9, MARTMAP generally is significantly superior than ARTMAP-FD and FAM at lower values of $\bar{\rho}_a$, with the highest accuracy at 97.99% [CI = 97.83, 98.16] for $\bar{\rho}_a$ = 0.10.

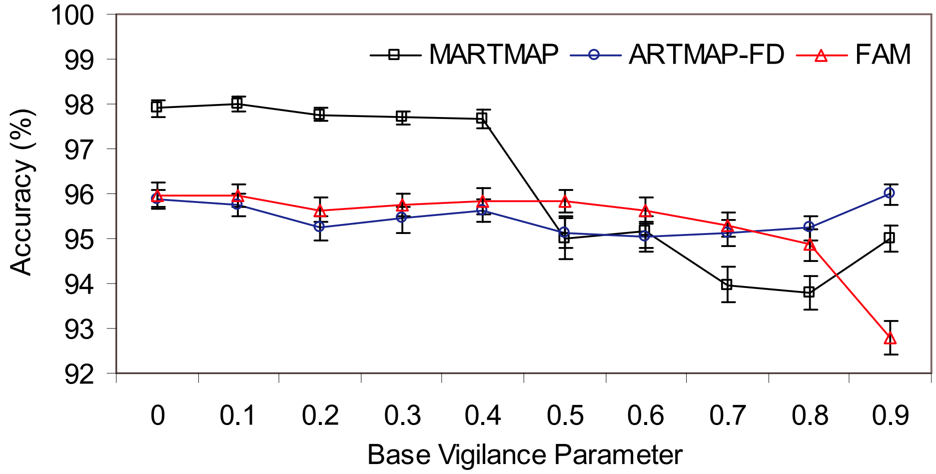


Fig. 6 – This figure shows the results averaged over 1000 bootstrap accuracies along with 95% confidence intervals.

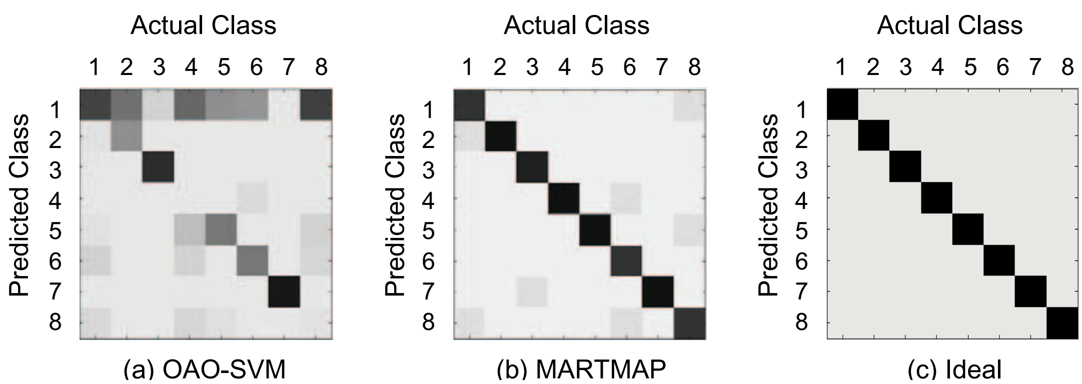


Fig. 7 – Confusion matrices for shape classification by using MARTMAP and OAO-SVM. Each row represents the probabilities of that class being confused with all the other class averaged over 100 runs

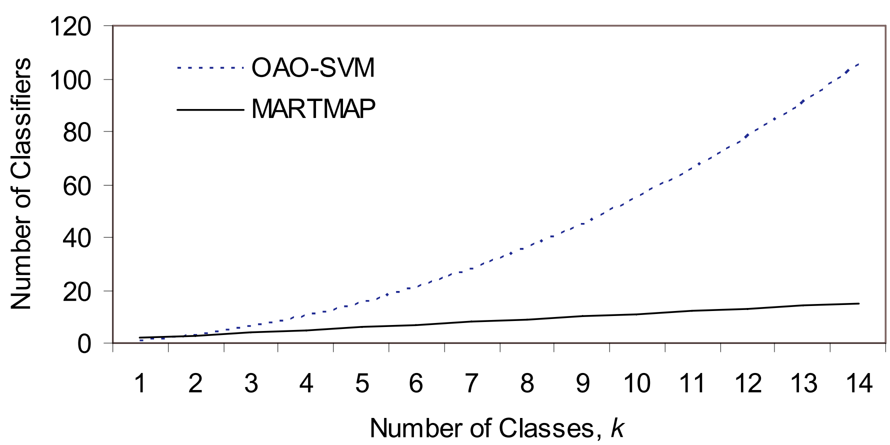


Fig. 8 – This figure illustrates the number of classifier needed by OAO-SVM and MARTMAP

B. Comparison with Multi-Class SVM

In this study, the approach used for multi-class SVM was "One-Against-One" (OAO-SVM) approach, with a Gaussian radial basis kernel. The classifier was implemented by using LIBSVM library [9]. In OAO-SVM, $k(k-1)/2$ binary classifiers were trained for a k-class problem. Each binary classifier is trained with data from two classes, C_a and C_b ($1 \leq a, b \leq k$). In the prediction stage, a binary classifier produces a vote indicating whether a feature vector $\boldsymbol{a}$ belongs to class C_a or C_b. A voting scheme is used to select the class with the most votes and assign it to the feature vector $\boldsymbol{a}$.

Since there are four classes in the Gaussian data set, the value k was set to 4. With the Gaussian data set, the result obtained using OAO-SVM was 97.63% [CI = 97.46, 97.81]. In the case where Gaussian data was corrupted with noise, OAO-SVM achieved an accuracy of 97.73% [CI = 97.52, 97.93]. In both cases, there were no obvious difference between the performance achieved by MARTMAP and OAO-SVM.

In the experiment based on shape data set, the best accuracy obtained by using OAO-SVM was 62.17% [CI = 61.10, 63.26], compared with 92.37% [CI = 91.64, 93.08] achieved by using MARTMAP. As can be seen from the confusion matrices depicted in Fig. 7a, the shapes were more likely to be misclassified by OAO-SVM as class 1. Clearly, the performance deterioration of OAO-SVM was caused by the imbalance training data. The amount of training data contributed by minority classes was far less than the amount in majority class (class 1). As a result, OAO-SVM had a tendency to produce outputs skewed to the majority class. In other words, it classified far more patterns as belonging to majority class than it should. Huang et al. [10] suggested that the undesirable biasing problem of OAO-SVM might be due to the equal error penalty of misclassification for all the classes. In contrast, the same problem did not occur in MARTMAP. As depicted in Fig. 7b, MARTMAP was less sensitive to uneven training class size as compared with OAO-SVM and it was able to classify patterns, that are from the minority classes.

Another observation in this study was the number of classifiers needed by both OAO-SVM and MARTMAP. As can be observed from Fig. 8, the number of classifiers needed in MARTMAP increases linearly with number of classes k. On the other hand, the number of binary classifiers required by OAO-SVM increases in quadratic with k. As a result, OAO-SVM may computationally be more expensive since large number of binary classifiers has to be trained to handle each binary sub-problem when k is large.

C. CONCLUSION AND FUTURE WORK

The paper has presented a new FAM-based modular architecture known as MARTMAP for multi-class pattern recognition. The dynamics of the proposed network have been described in detail to explain how multiple novelty detectors can be used to draw collective decision. In MARTMAP, individual novelty detector is employed to discover and learn the natural groupings of the pattern class assigned to it by forming hyper-rectangulars to enclose the pattern region. In the prediction stage, Euclidean distance was used to measure the familiarity between unknown patterns with the centroid of the clusters formed in each novelty detector. Familiarity scores are aggregated to a make a collective decision in several ways such as min-win strategy or multiple hypothesis estimation.

By using three data sets, the paper has demonstrated that the proposed architecture is capable of classifying multi-class patterns with higher accuracy as compared to ARTMAP-FD and FAM. In particular, the paper has shown the capability of MARTMAP in retaining its classification accuracy when the training data is corrupted with noise. Although individual novelty detectors in both ARTMAP-FD and MARTMAP are trained with local knowledge, but MARTMAP is able to resolve the classification uncertainty problem faced by ARTMAP-FD in overlapped boundaries and regions that are not enclosed by hyper boxes.

In the comparison study against OAO-SVM, MARTMAP was found to be less sensitive in dealing with class imbalance problem. In addition, it is better in terms of implementation simplicity and network complexity. The implementation simplicity arises from the flexibility in making changes to individual novelty detector including re-training without affecting the whole classification module. Besides, MARTMAP is comparatively simpler than OAO-SVM as it requires less classifiers in solving the same multi-class problem.

The paper has revealed the potential of MARTMAP as a multi-class classifier with good robustness to noisy and imbalance training data. However, there are still a number of areas that can be enhanced and pursued as further work. Firstly, some studies have shown that hyper-rectangular may not be a good geometrical representation for certain types of data. Therefore, it is worthwhile to investigate other geometrical representation such as ellipsoid or hyper-sphere, which may give better generalization and representation of category. Secondly, instead of using Euclidean distance as the discrimination function, effectiveness of other distance metrics such as Mahalanobis distance can be examined. Finally, the effectiveness of the MARTMAP has to be vindicated against more data sets.

REFERENCES

[1] G. Ou, Y. L. Murphey, "Multi-class pattern classification using neural networks," *Pattern Recognition*, Vol. 40, No. 1, 2007, pp. 4–18.

[2] F. Schwenker, "Solving multi-class pattern recognition problems with tree-structured support vector machines," *Proceedings of the 23rd DAGM-Symposium on Pattern Recognition*, 2001, pp. 283–290.

[3] J. Weston and C. Watkins, "Support vector machines for multiclass pattern recognition," *Proceeding of the 7th European Symposium On Artificial Neural Networks*, April 1999.

[4] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for

incremental supervised learning of analogue multidimensional maps," *IEEE Transactions on Neural Networks*, Vol. 3, 1992, pp. 698–712.
[5] G. A. Carpenter, M. A. Rubin, W. W. Streilein, "ARTMAP-FD: familiarity discrimination applied to radar target recognition," *Proceeding of the International Conference on Neural Networks*, 1997, pp. 1459–1464.
[6] M. Markou, S. Singh, "Novelty detection: a review–part 1: statistical approaches," *Signal Processing*, Vol. 83, No. 2, 2003, pp. 2481–2497.
[7] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," *Proceedings IEEE Workshop Applications of Computer Vision*, 1998, pp. 8–14.
[8] T. B. Sebastian, P. N. Klein, B. B. Kimia, "Shock-based indexing into large shape databases," *Proceedings of 7th European Conference on Computer Vision*, 2002, pp. 83–89.
[9] C. C. Chang and C. J. Lin, "LIBSVM : a library for support vector machines," Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm
[10] Y. M. Huang and S. X. Du, "Weighted support vector machine for classification with uneven training class sizes," *Proceedings of the 4th IEEE International. Conference on Machine Learning and Cybernetics*, 2005, pp. 4365–4369.