Expert Systems with Applications 37 (2010) 1628–1634

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A new computational intelligence technique based on human group formation

Arit Thammano [a,*], Jittraporn Moolwong [b]

[a] Computational Intelligence Laboratory, Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand
[b] Kiatnakin Bank Plc., Sukhumvit 21, Sukhumvit Road, Bangkok 10110, Thailand

ARTICLE INFO

Keywords:
Computational intelligence
Human group formation
Classification
Data mining

ABSTRACT

This paper proposes a novel computational intelligence technique, based on the sociological concept of human group formation, with the aim to acquire a better solution to classification problems. The key concept of the human group formation is about the behavior of in-group members that try to unite with their own group as much as possible, and at the same time maintain social distance from the out-group members. This study compares the performance of the proposed model with that of fuzzy ARTMAP, radial basis function network, and learning vector quantization. Experimental results demonstrate the potential of the proposed approach in offering an efficient and effective solution to the problem.

1. Introduction

Computational intelligence is a fast-moving, multidisciplinary field. It covers various disciplines of computer science such as granular computing, neuro-computing, evolutionary computing, and artificial life (Konar, 2005). In recent years computational intelligence has attracted more and more attention over the traditional artificial intelligence. Unlike traditional artificial intelligence, computational intelligence is tolerant of imprecise information, partial truth, and uncertainty (Andina & Pham, 2007). Traditional artificial intelligence is very good in inductive and analogy-based learning (Konar, 2005) but it is inefficient to solve problems with large input sizes, as in data mining. The incompetence of traditional artificial intelligence has opened up new avenues for the non-conventional techniques like computational intelligence.

Classification, one of the most common data mining tasks, is the process of finding a set of models that describe and distinguish data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown (Han & Kamber, 2001). A variety of computational intelligence techniques have been applied to deal with the classification problems such as artificial neural networks (Ham & Han, 1996; Huysmans, Baesens, Vanthienen, & Gestel, 2006; Lin, Xiao, & Micheli-Tzanakou, 1998; Loo & Rao, 2005; Mahanty & Dutta Gupta, 2004), fuzzy sets (Lee, Chen, Chen, & Jou, 2001; Li, Guo, & Kuo, 2005; Vernet & Kopp, 2002), and evolutionary algorithm (Chen & Hsu, 2006; Thammano & Trikeawcharoen, 2007; Wang & Wang, 2006; Zahiri & Seyedin, 2007). Among a number of computational intelligence methods in use, artificial neural networks are the most widely used approaches in solving classification problems. Many previous research works (Danaher et al., 1997; Goel et al., 2003; Lee et al., 1990; Quinlan, 1994; Russell & Norvig, 1995; Shavlik, Mooney, & Towell, 1991) show that neural network classifiers have better performance, lower classification error rate, and more robust to noise than other classification methods. Due to the superiority of neural networks in solving the classification problems, they were chosen to be reference methods in this research.

In this study, the new computational intelligence technique, which is inspired by the theory of human group formation, is proposed. The predictive performance of the proposed method is evaluated against three of the most powerful neural networks: the fuzzy ARTMAP neural network, the radial basis function network, and the learning vector quantization network.

This paper is divided into 6 sections. Following this introduction, Section 2 briefly introduces the three artificial neural networks which are used as a benchmark in this study. Section 3 presents the theoretical background of the human group formation. The proposed algorithm is described in Section 4. A brief description of the experimental data and the experimental results are given in Section 5. Finally, Section 6 is the conclusion.

2. Artificial neural networks

Artificial neural networks are the information processing systems that have been developed based on principles observed in human biological neural systems. Neural networks have many desirable characteristics such as resistance to noise, tolerance to distorted input patterns (ability to generalize), superior ability to recognize partially occluded or degraded images, ability to discriminate among overlapping pattern classes or classes with highly

* Corresponding author. Tel.: +66 2 723 4964; fax: +66 2 723 4910.
E-mail addresses: arit@it.kmitl.ac.th (A. Thammano), jittraporn@kiatnakin.co.th (J. Moolwong).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2009.06.046

nonlinear boundaries, and potential for parallel processing (Pal & Pal, 2001).

Various types of artificial neural networks have made their appearance over the years. In this study, however, only three of the most powerful neural networks that are designed particularly for classification are chosen and briefly described.

2.1. Fuzzy ARTMAP

Fuzzy ARTMAP (Carpenter & Grossberg, 1992; Carpenter, Grossberg, Markuzon, Reynolds, & Rosen, 1992), introduced by Carpenter et al in 1992, is a supervised neural network with very interesting properties, such as a very fast convergence, a capacity of incremental learning of recognition categories, and multidimensional mapping in response to input vectors. However, its major drawback is that the network performance is affected by the ordering of training sample presentation (Carpenter et al., 1992; Loo & Rao, 2005). Carpenter, Grossberg, and Iizuka (1992) compared the performance of fuzzy ARTMAP with that of learning vector quantization (LVQ) and Backpropagation neural network (BPNN). The findings show that fuzzy ARTMAP has the edge over LVQ and BPNN in terms of prediction accuracy, and is far superior in terms of speed of learning.

As shown in Fig. 1, Fuzzy ARTMAP is a supervised neural network composed of two fuzzy ART modules (ART_a and ART_b). The two fuzzy ARTs are linked together via the map field F^{ab}. Each fuzzy ART module has two layers: F_1^a and F_1^b are the input layers of each module, while F_2^a and F_2^b are the category layers. During the training phase, the input vector and its desired output vector are presented to ART_a and ART_b, respectively. Before the input vector and its desired output vector are transmitted to their corresponding input layers F_1^a and F_1^b, in order to avoid the category proliferation problem, they are flowed through the complement coder where their strings are stretched to double the size by adding their complements. The complement coded vectors, $A = (a, a^c)$ and $B = (b, b^c)$, are then applied to the corresponding input layers. The ART_a and ART_b modules classify the input and desired output vector into categories, then the map field uses a vigilance criterion to evaluate whether ART_a category corresponds to ART_b category. The criterion is as follows:

$$\frac{\left|y^b \wedge w_J^{ab}\right|}{|y^b|} \geqslant \rho_{ab} \tag{1}$$

where y^b is the output vector of ART_b, J is the index of the winning node in F_2^a, w_J^{ab} is the weights of the connections from the Jth node in F_2^a to the map field F^{ab}, and ρ_{ab} is the vigilance parameter of the map field. If the criterion is not respected, the vigilance parameter of ART_a, ρ_a, will be increased by a minimum value just enough to force ART_a module to search for another category. However, when the vigilance criterion is respected, the weight vector is updated according to the equation:

$$w_J^{ab} = \beta_{ab} x^{ab} + (1 - \beta_{ab}) w_J^{ab} \tag{2}$$

where β_{ab} is the learning rate, which is between 0 and 1. x^{ab} is the output vector of F^{ab}.

2.2. Radial basis function network

Radial basis function network (RBF) (Fig. 2) comprises 3 layers: the input layer, the hidden layer, and the output layer. The input layer simply transfers the input patterns to the nodes in the hidden layer. Each node in the hidden layer then calculates its output as follows:

$$y_j = e^{\left(\frac{-\|x - v_j\|^2}{\sigma_j^2}\right)} \tag{3}$$

where $j = 1, 2, 3, \ldots, C$. It denotes the jth node in the hidden layer.

C is the total number of nodes in the hidden layer; x is the input vector; v_j is the center of the jth cluster; σ_j is the width of the jth cluster.

The outputs of the network are defined as the linear combination of the basis functions in the hidden layer.

$$y_k = \sum_{j=1}^{C} w_{jk} y_j \tag{4}$$

There are two steps in the training of the RBF network. First, the basis function parameters, v_j and σ_j, need to be determined. This is usually done by unsupervised learning algorithms such as K-means algorithm, fuzzy C-means algorithm, SOM, or the successive approximation method. Second, the connection weights between the hidden layer and the output layer, w_{jk}, are trained using a method like gradient descent method or least squares method.

2.3. Learning vector quantization network

Learning vector quantization network (LVQ) (Kohonen, 1990), a special case of the Kohonen self-organizing maps, is a supervised competitive neural network model in which each output node represents a particular class (Fig. 3). During training, the output node whose weight vector most closely matches the input pattern is chosen as the winner. If the winning node has the correct class label, its weight vector is moved toward the input pattern. However, if it belongs to the wrong class, its weight vector is moved away from the input.

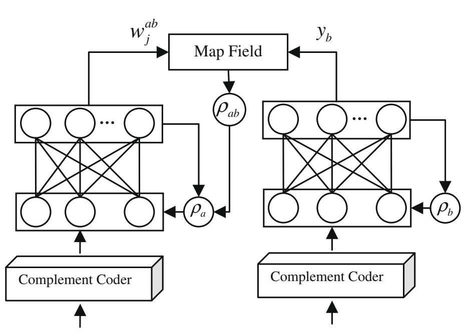

Fig. 1. The architecture of fuzzy ARTMAP.

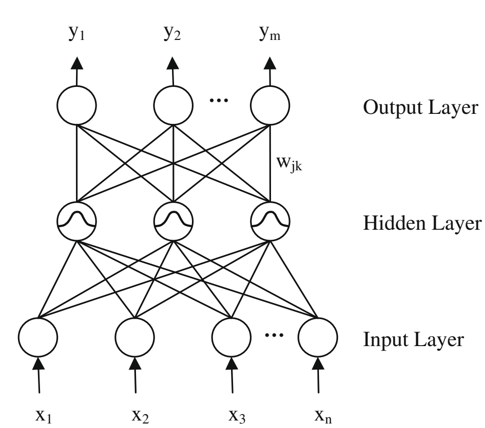

Fig. 2. The architecture of RBF network.

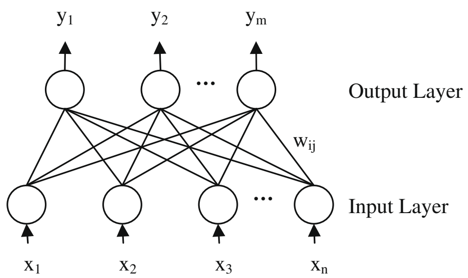

Fig. 3. The architecture of LVQ.

$$w_J^{new} = w_J^{old} + \Delta w_J \tag{5}$$

$$\Delta w_J = \begin{cases} \eta\left(x - w_J^{old}\right) & \text{Correct class} \\ -\eta\left(x - w_J^{old}\right) & \text{Wrong class} \end{cases} \tag{6}$$

The main disadvantage of LVQ is that it may yield unacceptable performance when the data clusters are multimodal or very complex (Principe, Euliano, & Lefebvre, 2000).

3. Human group formation

Sociologists have created the concept of "in-group" and "out-group" as a means to explain human social categorization. Members of the in-group are people who are accepted by the group as belonging to the group, while the out-group members are those whom in-group members consider not belonging to their group. Even though categorizing people into groups by identifying some common attributes reduces the complexity of the social world (Hamilton, 1981), there are also less desirable consequences that go along with this simplification strategy. Ethnocentrism, in-group bias, and prejudice are some of them (Tajfel, 1969; Tajfel, 1978).

When people identify themselves with a group, they perceive themselves and their group members as different from other groups (Schlabach, 1998). They also believe that their own group is superior to other groups. This leads in-group members to favor their own group over the out-group (Devine, 1989). Therefore in-group members, even when far away from the group, try to unite with their own group as much as possible, and at the same time maintain social distance from the non-members by creating negative attitude or taking a hostile stance against them. However, among the in-group members, each tends to have his own territory instead of staying tightly close to one another.

Fig. 4 is an example of how to put the above concept into practice. From this example, the data set contains 35 instances from 3 classes: "star," "circle," and "square." There are 12 instances of class "star," 10 instances of class "circle," and 13 instances of class "square." Small stars, circles, and squares represent the corresponding input instances while big stars, circles, and squares represent the corresponding cluster centers. By way of illustration let us look at the case of the 4th cluster center. The 4th cluster center belongs to class "star." According to the concept of the human group formation, the 4th cluster center must move toward its in-group members (the 1st, 2nd, and 3rd cluster centers) and away from the out-group members (the 5th, 6th, 7th, 8th, and 9th cluster centers). The velocity of this move depends on 3 factors:

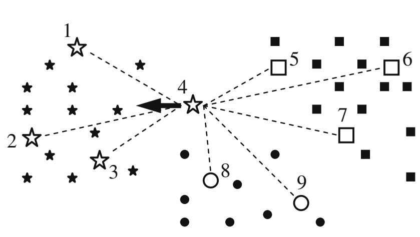

Fig. 4. Example of the human group formation concept.

- The distances between the 4th cluster center and the other cluster centers.
- The 4th cluster center's ability to move in a search space – If the 4th cluster center makes a move in the direction that does not improve a situation, the value of this factor will be reduced by a predefined amount. This action tells the 4th cluster center to proceed with more caution.
- The territorial boundary of each cluster center – This factor prevents clusters of the same class from being too close to one another.

4. Proposed algorithm

The proposed algorithm is developed according to the theory of human group formation discussed in the previous section. The detailed procedure of the proposed algorithm is as follows:

(A) Arbitrarily select one input pattern from the training data set of each class, and assign them to be the initial cluster centers. At first, there are a total of Q clusters, which is equal to the number of target output classes.

(B) Calculate the accuracy of the model as follows:

$$\text{Accuracy} = \frac{\sum_{i=1}^{P} A_i}{P} \tag{7}$$

$$A_i = \begin{cases} 1, & \text{if } J \in Y_i \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

$$J = \arg\min_j (d_j(X_i)) \tag{9}$$

$$d_j(X_i) = \|X_i - z_j\| \tag{10}$$

where P is the total number of patterns in the training data set. J is the index of a cluster whose reference pattern is the closest match to the incoming input pattern X_i.

Y_i is the target output of the ith input pattern.
z_j is the center of the jth cluster.
$d_j(X_i)$ is the Euclidean distance between the input pattern X_i and the center of the jth cluster.

(C) Based on the concept of human group formation that in-group members try to unite with their own group and maintain social distance from the non-members as much as possible, update the center value of each cluster (z_j) by using the following formula:

$$z_{jk}^{new} = z_{jk}^{old} + \Delta z_{jk} \tag{11}$$

$$\Delta z_{jk} = \sum_{m \in q} \eta_{jm}\beta_j\delta_{jm}(z_{mk} - z_{jk}) - \sum_{n \notin q} \eta_{jn}\beta_j\delta_{jn}(z_{nk} - z_{jk}) \tag{12}$$

where $k = 1, 2, 3, \ldots, K$. K is the number of features in the input pattern.
q is the class to which the jth cluster belongs.

$$\eta_{jm} = e^{-\left(\frac{z_{jk} - z_{mk}}{\sigma}\right)^2} \text{ and } \eta_{jn} = e^{-\left(\frac{z_{jk} - z_{nk}}{\sigma}\right)^2}.$$

η_{jm} and η_{jn} have values between 0 and 1. They determine the influence of mth and nth clusters on the jth cluster. They will have high values if mth and nth clusters are close to the jth cluster. However, if mth and nth clusters are far away from the jth cluster, the values of η_{jm} and η_{jn} will be low. The further apart mth and nth clusters are from the jth cluster, the lower the values of η_{jm} and η_{jn}.

β_j is the velocity of the jth cluster with respect to its own ability to move in the search space.

δ_{jm} is the velocity of the jth cluster with respect to 2 factors: (1) the distance between the jth cluster and the mth cluster, and (2) the territorial boundary of the clusters (T). This parameter prevents clusters of the same class from being too close to one another. If the distance between the jth cluster and the mth cluster is less than T, the value of δ_{jm} will be decreased by a predefined amount.

After each center is updated, recalculate the accuracy of the model according to Eqs. (7 to 10). If the accuracy is higher, save this new center value and then continue updating the next cluster center. If the accuracy of the model is lower, discard the new center value and return to the previous center. However, if the accuracy does not change, save the new center value and decrease the value of β_j by a predefined amount. As long as the accuracy remains unchanged, the values of β_j are continuously decreased. In doing this, the algorithm allows the cluster center to move around the search space, however with a slower and slower pace, even if its move does not do any good at the moment.

(D) This step is the cluster reduction step. The cluster which satisfies the following equation will be deleted.

$$-\frac{1}{2\log_2\left(\frac{n_j}{P}\right)}\left(\frac{n_j^q}{n_j}\right)\left(\frac{\sum_{\forall X_i^j\in q}\left\|X_i^j - z_j\right\|}{n_j}\right) < \rho \tag{13}$$

where n_j is the number of input patterns in the jth cluster; n_j^q is the number of input patterns in the jth cluster whose target outputs (Y) are q; q is the class to which the jth cluster belongs; P is the total number of patterns in the training data set; X_i^j is the ith input pattern in the jth cluster; ρ is the vigilance parameter.

(E) Recalculate the accuracy of the model according to Eqs. (7 to 10).

(F) For each remaining cluster, if the distance between the new center updated in step C and the previous center is less than 0.0001 $\left(\left\|z_{jk}^{new} - z_{jk}^{old}\right\| < 0.0001\right)$, randomly pick K small numbers between -0.1 and 0.1 and then add them to the center value of the cluster. The purpose of this step is to prevent the premature convergence of the proposed algorithm to sub-optimal solutions.

$$\begin{bmatrix} z_{j1}^{new} \\ z_{j2}^{new} \\ \vdots \\ z_{jK}^{new} \end{bmatrix} = \begin{bmatrix} z_{j1}^{old} \\ z_{j2}^{old} \\ \vdots \\ z_{jK}^{old} \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_K \end{bmatrix} \tag{14}$$

If the above operation makes the accuracy higher, save this new center value and then continue with the next cluster center. However, if the accuracy does not improve, discard the new center value and repeat this step until the accuracy improves or a predetermined number of iteration is reached.

(G) If a predetermined number of iteration is reached or the end condition is satisfied, stop the loop. If not, examine the following conditions:

(G.1) If the accuracy of the model improves over the previous iteration, randomly select one input pattern from the training data set of each target output class that still has error. Then go to step B.

(G.2) However, if the accuracy does not improve, randomly select the input patterns, a number equal to the number of clusters deleted in step D, from the training data set of each target output class. Then go to step B.

5. Experimental results

To test the performance, the proposed algorithm was benchmarked against the fuzzy ARTMAP neural network, the radial basis function network (RBF), and the learning vector quantization network (LVQ). The experiments were conducted on 4 artificial data sets and 12 real-life data sets.

The four artificial data sets (Figs. 5–8) are called Fan, Flower 1, Flower 2, and Sawtooth. The names imply the shape of the classes. Fan data has 4 classes, while the other three have 2 classes. In each of the four data sets, the 4000 data points in the database were randomly divided into a training set of 2000 data points and a testing set of 2000 data points.

The twelve real-life data sets are Iris; Wine Recognition; Haberman's Survival; Heart Disease; Ionosphere; Balance Scale Weight and Distance; Glass Identification; Vowel Recognition; Sonar, Mines vs. Rocks; Vehicle Silhouettes; Image; and Zoo.

1. The first data set is the well-known Iris data. The iris data (Fisher, 1936) has been widely used in the classification problem. The sepal length, sepal width, petal length, and petal width of 150 Iris flowers from 3 species (Iris-setosa, Iris-versicolor, and Iris-virginica) are measured in centimeters, and are used as the input of the problem. The training set contains 120 records, while the testing set contains 30 records.
2. The second data set is the Wine Recognition data. It was retrieved from the UCI machine learning database repository (Asuncion & Newman, 2007). This data is the result of a chemical analysis of wines grown in the same region but from three different cultivars. Thirteen continuous attributes are used to determine the type of wine (class 1, 2, or 3). There are 59 instances of class 1, 71 instances of class 2, and 48 instances of class 3. In this paper, the 178 instances in the database were randomly divided into a training set of 90 instances and a testing set of 88 instances.
3. The third data set is the Haberman's Survival data. It is publicly available from the UCI machine learning database repository (Asuncion & Newman, 2007). This data set contains 306 cases from a study that was conducted between 1958 and 1970 at the University of Chicago's Billings Hospital on the survival of patients who had undergone surgery for breast cancer. Three numerical attributes are used to predict the output class (1 or 2). The class value of 1 is corresponded to "the patient survived 5 years or longer" and the value of 2 is corresponded to "the patient died within 5 years." There is no standard split between the training and testing sets. In this paper, the data was divided into a training set of 214 examples and a testing set of 92 examples.
4. The fourth data set is the Heart Disease problem. This Statlog project data set was retrieved from the UCI machine learning database repository (Asuncion & Newman, 2007). The problem concerns the prediction of the absence (1) or presence (2) of heart disease given the results of various medical tests carried out on a patient. This data set contains 13 attributes and 270 records. There are 150 records of class 1 and 120 records of class 2. In this paper, the 270 records in the database were randomly divided into a training set of 135 records and a testing set of 135 records.
5. The fifth data set is the Ionosphere data created by the Space Physics Group at Johns Hopkins University. The data was taken from the UCI machine learning database repository (Asuncion & Newman, 2007). It contains the radar data collected by a system in Goose Bay, Labrador. This data set has

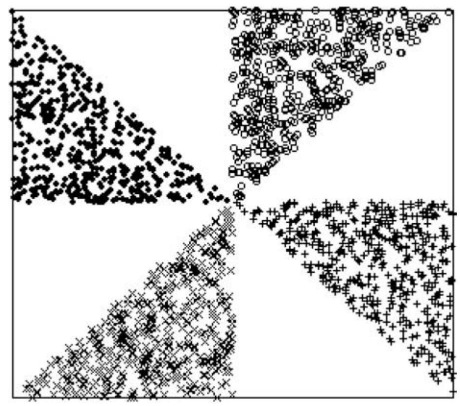

Fig. 5. Fan.

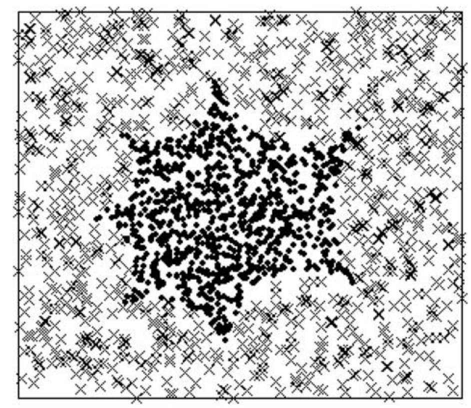

Fig. 7. Flower 2.

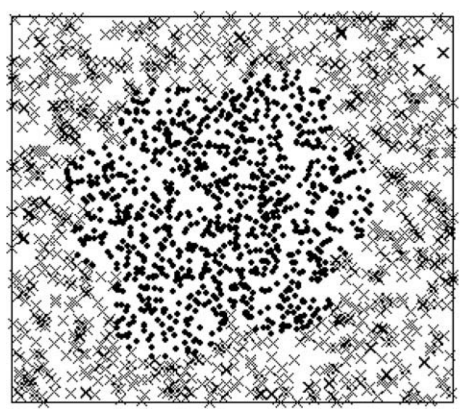

Fig. 6. Flower 1.

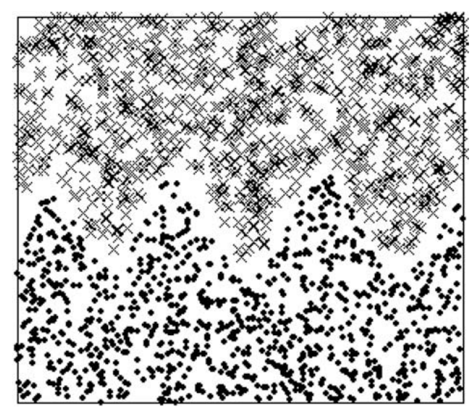

Fig. 8. Sawtooth.

Table 1
Experimental results on four artificial data sets.

Data set	Proposed algorithm		Fuzzy ARTMAP		RBF		LVQ	
	Accuracy	Cluster	Accuracy	Cluster	Accuracy	Hidden node	Accuracy	Cluster
1. Fan	**100(0)**	**4**	**100(0)**	**4**	99.95(1)	24	**100(0)**	**4**
2. Flower 1	**98.75(25)**	**32**	98.2(36)	151	96.55(69)	77	98.15(37)	298
3. Flower 2	**98.7(26)**	**43**	98.35(33)	142	98.4(32)	78	98.15(37)	334
4. Sawtooth	99.75(5)	17	99.9(2)	138	98.45(31)	44	**100(0)**	**246**

Table 2
Experimental results on twelve real-life data sets.

Data set	Proposed algorithm		Fuzzy ARTMAP		RBF		LVQ	
	Accuracy	Cluster	Accuracy	Cluster	Accuracy	Hidden node	Accuracy	Cluster
1. Iris	**100(0)**	**3**	**100(0)**	**4**	**100(0)**	**6**	**100(0)**	**3**
2. Wine	**100(0)**	**18**	97.73(2)	11	**100(0)**	**41**	95.45(4)	44
3. Haberman	**77.17(21)**	**2**	**77.17(21)**	**12**	**77.17(21)**	**52**	76.09(22)	2
4. Heart	**88.15(16)**	**15**	85.19(20)	46	**88.15(16)**	**37**	84.44(21)	2
5. Ionosphere	98.01(3)	53	98.68(2)	80	**99.34(1)**	**30**	95.36(7)	16
6. Balance Scale	**88.14(37)**	**27**	87.82(38)	155	86.22(43)	3	85.9 (44)	3
7. Glass	66.67(35)	23	**70.48(31)**	**23**	65.71(36)	6	62.86(39)	54
8. Liver Disorders	**72.83(47)**	**31**	67.63(56)	71	69.36(53)	83	68.79(54)	36
9. Sonar	**77.22(18)**	**13**	73.42(21)	39	72.15(22)	28	68.35(25)	52
10. Vehicle	67.06(139)	8	**72.27(117)**	**134**	65.4(146)	20	68.48(133)	160
11. Diabetes	**81.51(71)**	**6**	74.22(99)	128	79.95(77)	50	79.95(77)	8
12. Zoo	**97.96(1)**	**8**	95.92(2)	7	93.88 (3)	15	93.88(3)	7

200 instances in the training set and 151 instances in the testing set. Each instance is described by 34 continuous attributes and belongs to one of two classes ("good" or "bad"). "Good" radar returns are those showing evidence of some type of structure in the ionosphere. "Bad" radar returns are those that do not.

6. The sixth data set is the Balance Scale Weight and Distance database. The data was taken from the UCI machine learning database repository (Asuncion & Newman, 2007). This data set was generated to model psychological experiments. Each example is classified as having the balance scale tipping to the right, tipping to the left, or being balanced. The input data consists of 4 numerical attributes: Left-weight, Left-distance, Right-weight, and Right-distance. In this paper, the training set contains 313 examples, while the testing set contains 312 examples.
7. The seventh data set is the Glass Identification database (Asuncion & Newman, 2007). This data set contains 214 instances. Nine continuous attributes are used to predict the output class ("build wind float," "build wind non-float," "vehic wind float," "containers," "tableware," or "headlamps"). In this paper, the 214 instances in the database were randomly divided into a training set of 109 instances and a testing set of 105 instances.
8. The eighth is the BUPA Liver Disorders data set (Asuncion & Newman, 2007). Six numerical attributes are used to predict whether or not an unmarried man has a liver disorder. The first five attributes are the results of blood test while the last attribute is daily alcohol consumption. In this paper, the 345 patterns in the database were randomly divided into a training set of 172 patterns and a testing set of 173 patterns.
9. The ninth data set is the Sonar, Mines vs. Rocks (Asuncion & Newman, 2007). The task is to discriminate between sonar signals bounced off a metal cylinder and those bounced off a roughly cylindrical rock. This data set contains 111 patterns obtained by bouncing sonar signals off a metal cylinder at various angles and under various conditions and 97 patterns obtained from rocks under similar conditions. Sixty numerical attributes in the range 0.0–1.0 are used to predict the output class (mine or rock). In this paper, the 208 patterns in the database were randomly divided into a training set of 129 patterns and a testing set of 79 patterns.
10. The tenth is the Vehicle Silhouettes data set (Asuncion & Newman, 2007). The purpose is to classify a given silhouette as one of four types of vehicle (a double-decker bus, Chevrolet van, Saab 9000, and Opel Manta 400). Eighteen features were extracted from each of the silhouette by using the Hierarchical Image Processing System (HIPS). There are a total of 846 patterns in the data set. There is no standard split between the training and testing sets. In this paper, the data was randomly divided into a training set of 424 examples and a testing set of 422 examples.
11. The eleventh data set is the Pima Indians diabetes database. It is publicly available from the UCI machine learning database repository (Asuncion & Newman, 2007). The problem is to predict whether a patient would test positive (1) or negative (0) for diabetes according to World Health Organization criteria. All patients represented in the data set are females at least 21 years old of Pima Indian heritage living near Phoenix, Arizona, USA. This database contains 768 examples. Each example is described by 8 numerical attributes and belongs to one of two classes (0 or 1). There are 500 examples of class 0 and 268 examples of class 1. In this paper, the 768 examples in the database were randomly divided into a training set of 384 examples and a testing set of 384 examples.
12. The twelfth is the Zoo data set. It was retrieved from the UCI machine learning database repository (Asuncion & Newman, 2007). Sixteen Boolean-valued attributes are used to determine the type of animals (type 1–7). There is no standard split between the training and testing sets. In this paper, the 101 instances in the database were randomly divided into a training set of 52 instances and a testing set of 49 instances.

The performance is measured in terms of the classification accuracy and the size of the model. It has been documented in the literature that the performance of the artificial neural network depends highly on proper selection of network parameters. Since there is still no foolproof method for identifying optimum value of these parameters, the parameters of all four models – the vigilance parameter and the learning rate in the fuzzy ARTMAP, the number of basis functions in RBF, the number of output nodes in LVQ, and the vigilance parameter (ρ) and the territorial boundary of the clusters (T) in the proposed model – are varied throughout their ranges in order to get the best out of the four models. For each parameter setting, moreover, five experimental repetitions were performed with different orderings of training sample presentation. Results of the experiments are shown in Tables 1 and 2. The figures which are bold represent the best results among the four methods, while the figures in brackets indicate the number of misclassified patterns.

In view of the classification accuracy, the following are summaries of what we observe from the experimental results:

(A) Experimental results on four artificial data sets
- The performance of the proposed model is better than the performance of the fuzzy ARTMAP by 0.55, 0.35, 0.0, and −0.15% for Flower 1, Flower 2, Fan, and Sawtooth, respectively.
- The performance of the proposed model is better than the performance of RBF by 2.2, 1.3, 0.3, and 0.05% for Flower 1, Sawtooth, Flower 2, and Fan, respectively.
- The performance of the proposed model is better than the performance of LVQ by 0.6, 0.55, 0, and −0.25% for Flower 1, Flower 2, Fan, and Sawtooth, respectively.

(B) Experimental results on twelve real-life data sets
- The performance of the proposed model is better than the performance of the fuzzy ARTMAP by 7.29, 5.2, 3.8, 2.96, 2.27, 2.04, 0.32, 0.0, 0.0, −0.67, −3.81, and −5.21% for Diabetes, Liver Disorders, Sonar, Heart, Wine, Zoo, Balance Scale, Iris, Haberman, Ionosphere, Glass, and Vehicle, respectively.
- The performance of the proposed model is better than the performance of RBF by 5.07, 4.08, 3.47, 1.92, 1.66, 1.56, 0.96, 0.0, 0.0, 0.0, 0.0, and −1.33% for Sonar, Zoo, Liver Disorders, Balance Scale, Vehicle, Diabetes, Glass, Iris, Wine, Haberman, Heart, and Ionosphere, respectively.
- The performance of the proposed model is better than the performance of LVQ by 8.87, 4.55, 4.08, 4.04, 3.81, 3.71, 2.65, 2.24, 1.56, 1.08, 0.0, and −1.42% for Sonar, Wine, Zoo, Liver Disorders, Glass, Heart, Ionosphere, Balance Scale, Diabetes, Haberman, Iris, and Vehicle, respectively.

Next, we will examine the performance of all four methods concerning the model size. In this study, the size of the model is evaluated in terms of the number of clusters created during the training process. For most of the databases (12 out of 16 databases), the size of the proposed model is either smaller or approximately equal to the sizes of the other three methods. These results demonstrate the ability of the cluster reduction step in limiting the number of clusters created during the training process while still retaining a high degree of accuracy.

The above performance comparisons with other techniques support the validity of the proposed approach. It performs very well in solving all 16 classification problems.

6. Conclusions

The sociological theory of human group formation is employed in this research as the main concept to develop the proposed approach. In addition to the theory of human group formation, the proposed approach also contains the cluster reduction step and the local minimum escaping step.

The experiments show that the proposed approach can successfully classify all sixteen benchmark problems. In comparison to three well established neural network models, the time used for training the proposed approach is larger. However, the proposed approach is found to be superior to the three referenced models both in terms of the classification accuracy and the size of the model.

References

Andina, D., & Pham, D. T. (2007). *Computational intelligence for engineering and manufacturing*. Springer.

Asuncion, A. & Newman, D. J. (2007). UCI machine learning repository. <http://www.ics.uci.edu/~mlearn/MLRepository.html>, University of California, School of Information and Computer Science.

Carpenter, G. A., & Grossberg, S. (1992). A self-organizing neural network for supervised learning, recognition, and prediction. *IEEE Communications Magazine*, 38–49.

Carpenter, G., Grossberg, S., & Iizuka, K. (1992). Comparative performance measures of fuzzy ARTMAP, learned vector quantization, and back propagation for handwritten character recognition. In *Proceedings of the international joint conference on neural networks* (pp. 794–799).

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., & Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks, 3*(5), 698–713.

Chen, T. C., & Hsu, T. C. (2006). A GAs based approach for mining breast cancer pattern. *Expert Systems with Applications, 30*(4), 674–681.

Danaher, S. et al. (1997). *A comparison of the characterization of agricultural land using singular value decomposition and neural networks. Neurocomputing in Remote Sensing Data Analysis*. Springer.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annual Eugenics, 7*(II), 179–188.

Goel, P. K., Prasher, S. O., Patel, R. M., Landry, J. A., Bonnell, R. B., & Viau, A. A. (2003). Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. *Computers and Electronics in Agriculture, 39*, 67–93.

Ham, F. M., & Han, S. (1996). Classification of cardiac arrhythmias using fuzzy ARTMAP. *IEEE Transactions on Biomedical Engineering, 43*(4), 425–430.

Hamilton, D. L. (1981). *Cognitive processes in stereotyping and intergroup behavior*. Hillsdale.

Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques*. Academic Press.

Huysmans, J., Baesens, B., Vanthienen, J., & Gestel, T. V. (2006). Failure prediction with self-organizing maps. *Expert Systems with Applications, 30*, 479–487.

Kohonen, T. (1990). The self-organizing maps. In *Proceedings of the IEEE* (vol. 78(9), pp. 1464–1480).

Konar, A. (2005). *Computational intelligence: Principles techniques and applications*. Springer.

Lee, H. M., Chen, C. M., Chen, J. M., & Jou, Y. L. (2001). An efficient fuzzy classifier with feature selection based on fuzzy entropy. *IEEE Transactions on Systems Man and Cybernetics – Part B, 31*(3), 426–432.

Lee, J. et al. (1990). A neural network approach to cloud classification. *IEEE Transactions on Geoscience and Remote Sensing of Environment, 66*, 179–191.

Li, T. H. S., Guo, N. R., & Kuo, C. L. (2005). Design of adaptive fuzzy model for classification problem. *Engineering Applications of Artificial Intelligence, 18*, 297–306.

Lin, W., Xiao, J., & Micheli-Tzanakou, E. (1998). A computational intelligence system for cell classification. In *Proceedings of the 1998 IEEE international conference on information technology applications in biomedicine* (pp. 105–109).

Loo, C. K., & Rao, M. V. C. (2005). Accurate and reliable diagnosis and classification using probabilistic ensemble simplified fuzzy ARTMAP. *IEEE Transactions on Knowledge and Data Engineering, 17*(11), 1589–1593.

Mahanty, R. N., & Dutta Gupta, P. B. (2004). Applications of RBF neural network to fault classification and location in transmission lines. In *IEE proceedings generation, transmission and distribution* (vol. 151(2), pp. 201–212).

Pal, A., & Pal, S. K. (2001). *Pattern recognition: evolution of methodologies and data mining. Pattern Recognition: From Classical to Modern Approaches*. World Scientific.

Principe, J. C., Euliano, N. R., & Lefebvre, W. C. (2000). *Neural and adaptive systems: Fundamentals through simulations*. John Wiley and Sons.

Quinlan, J. R. (1994). Comparing connectionist and symbolic learning methods computational learning theory and natural learning systems. *The MIT Press, 1*, 445–456.

Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Springer.

Schlabach, J. (1998). In-group, out-group bias. <http://www.units.muohio.edu/psybersite/fans/inoutbias.shtml>.

Shavlik, J. W., Mooney, R. J., & Towell, G. G. (1991). Symbolic and neural learning algorithms: An experimental comparison. *Machine Learning, 6*(2), 111–143.

Tajfel, H. (1969). Cognitive aspects of prejudice. Journal of Social Issues (vol. 25, pp. 79–97). Blackwell Publishing.

Tajfel, H. (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations*. Academic Press.

Thammano, A., & Trikeawcharoen, S. (2007). Isan Dhamma character recognition using genetic algorithm. In *Proceedings of the third international conference on natural computation* (pp. 566–570).

Vernet, A., & Kopp, G. A. (2002). Classification of turbulent flow patterns with fuzzy clustering. *Engineering Applications of Artificial Intelligence, 15*, 315–326.

Wang, X., & Wang, H. (2006). Classification by evolutionary ensembles. *Pattern Recognition, 39*(4), 595–607.

Zahiri, S. H., & Seyedin, S. A. (2007). Swarm intelligence based classifiers. *Journal of the Franklin Institute, 344*(5), 362–376.