Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 29 (2008) 1082–1093
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Towards the experimental evaluation of novel supervised fuzzy adaptive resonance theory for pattern classification

Alireza Akhbardeh [a,c,*], Nikhil [a,b], Perttu E. Koskinen [b], Olli Yli-Harja [a]

[a] Institute of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland
[b] Institute of Environmental Engineering and Biotechnology, Tampere University of Technology, P.O. Box 541, 33101 Tampere, Finland
[c] School of Biomedical Engineering, Science and Health Systems, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA

Available online 4 November 2007

Abstract

This paper presents a comparative analysis of novel supervised fuzzy adaptive resonance theory (SF-ART), multilayer perceptron (MLP) and competitive neural trees (CNeT) Networks over three pattern recognition problems. We have used two well-known patterns (IRIS and Vowel data) and a biological data (hydrogen data) to evaluate and check SF-ART stability, reliability, learning speed and computational load. The comparative tests with IRIS, Vowels and H_2 data indicate that the SF-ART is capable to perform with a high classification performance, high learning speed (elapsed time for learning around half second), and very low computational load compared to the well-known neural networks such as MLP and CNeT which need minutes and seconds respectively to learn the training material.

Keywords: Affine look-up table; Classification; Pre-classification; Post-classification; Supervised fuzzy adaptive resonance theory (SF-ART) network

1. Introduction

Supervised neural networks have remarkable performance as practical tools and are effective for a broad range of pattern discrimination and functional approximation applications. As with any type of pattern classifier, the performance of neural networks relies heavily on the availability of a reprehensive set of training examples. In many practical applications, however, the acquisition of such a reprehensive data set is expensive and time consuming. The development of a neural classifier, with higher performance and speed, using minimum training set and fewer learning cycles is difficult, but desirable.

During the past several years, a large number of artificial neural networks either supervised or unsupervised with new structures/learning algorithms have been developed. Some of them are adaptive resonance theory (ART), neural trees, modified Hopfield, and learner ++ (Polikar et al., 2001, 2002). Most of the existing methods perform remarkably well when optimized learning parameters, rich training material and enough learning cycles are used.

Methods that do not deal with such important issues may potentially give us misleading information. Another limitation of the existing techniques concerns with their degree of success in the case of validation test with new data. Other limitations are their ease of hardware and/or software implementation, their stability across different patterns (generalization ability), and their suitability for real-time applications as well as learning incrementally from new data.

To have a reliable neural network with a high classification performance, high learning speed, incremental learning ability and easy to implement characteristics, we primarily presented in (Akhbardeh, 2007) a two stage supervised neural network called supervised fuzzy adaptive resonance theory (SF-ART). In this paper, we used two well-known

* Corresponding author. Tel.: +1 215 895 2234; fax: +1 215 895 4983.
E-mail address: aa485@drexel.edu (A. Akhbardeh).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2007.10.017

patterns (IRIS and vowel data) and a biological data, hydrogen (H_2) data, to evaluate more and check its stability, reliability, learning speed and computational load.

It must be mentioned that some multi-stage neural classifiers with two or more stages have been developed recently to increase recognition performance and reliability, but their structure is completely different from SF-ART. Zhang (2006) in his doctoral thesis proposed a novel cascade ensemble classifier system with a high recognition performance on handwritten digits. Joao (2006) worked on the impact of fusion strategies on classification errors for large ensembles of classifiers. Although the training time in the existing multistage learning algorithms is high, multistage learning algorithms are able to improve the performance of the classification in special applications (Kam et al., 1994; Teredesai and Govindaraju, 2005; Xu et al., 1992). However, these kinds of classifiers try to classify input features based on processing and clustering data in two stages. In the existing multi-stage classifiers, all the stages use input features (and desired values in supervised learning) to make the final decision.

Carpenter and Grossberg (1987, 1991, 1992) presented a supervised real-time learning algorithms using ARTMAP but those structures do not use output desired values for learning under supervision. On the other hand, they are supervised only to learn incrementally from new data, while SF-ART uses output desired values in its structure to learn under supervision and its is able to learn incrementally because of ALT structure. Therefore, in this paper, we have not compared SF-ART and ARTMAP. We have only used well-known multilayer perceptron (MLP) (Haykin, 1998; Bishop, 2005) and a fast neural network called competitive neural trees (CNeT) (Behnke and Karayiannis, 1996, 1998) to evaluate SF-ART performance over three patter recognition problems (IRIS, Vowel and H_2 data).

2. Structure of the applied neural networks

2.1. Multilayer perceptrons (MLP)

The most commonly used form of multilayer perceptron (MLP) is a feed forward neural network trained with the back propagation algorithm (Haykin, 1998). It is a supervised neural network and therefore requires a desired response to be trained. It learns how to transform input data into a desired response, and it is widely used for pattern classification. With one or two hidden layers, it can approximate virtually any input–output map. It has been shown to approximate the performance of optimal statistical classifiers in difficult problems. Most neural network applications involve MLP.

2.2. Competitive neural trees (CNeT)

Competitive neural trees (CNeT), developed by Behnke and Karayiannis (1996, 1998), is one of the fast supervised neural networks with high performance (see Fig. 1). A set of similar nodes forms a tree as shown in Fig. 1a. Fig. 1b shows a node in detail. Each node contains 'm' slots and a counter which shows the node age and increases each time an input pattern is presented to the node. The nodes show different behavior when the counter age increases. Each slot stores a prototype, counter 'count', and a pointer. The prototypes represent clusters of the input patterns. The counter 'count' increases each time the prototype is updated to fit an input pattern. The pointer points to a child-node assigned to a corresponding slot. A slot without any child-node (empty pointer) is called 'terminal slot' or 'leaf'. The internal slots are slots with an assigned child-node.

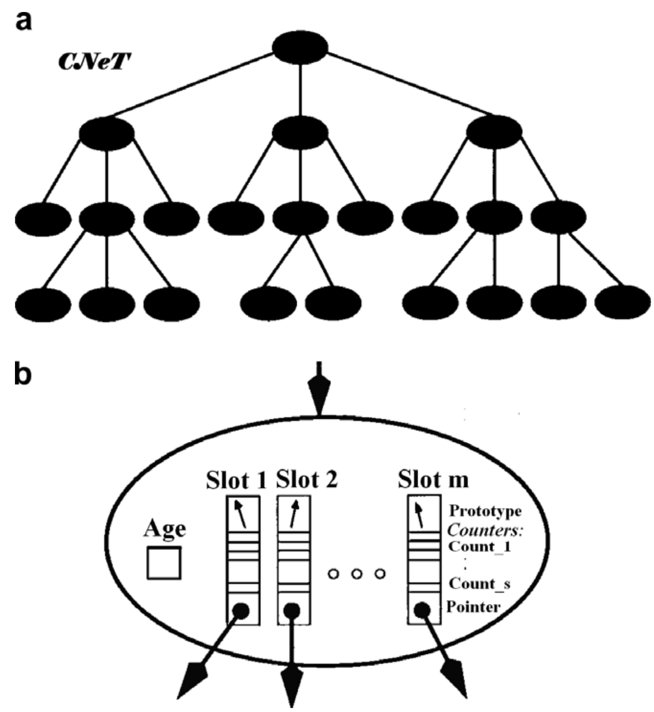

Fig. 1. Competitive neural tree (CNeT) structure and its nodes: (a) the tree with ellipsoids as its nodes; (b) A typical topology of a node. Each node contains prototypes, counters and pointer.

The growth of the CNeT is based on inheritance for initializing new nodes and can be controlled by forward pruning. CNeT applies unsupervised competitive learning in the node level and clusters the input feature vectors hierarchically. The prototype similar to the input pattern can be found by searching a part of the tree. Behnke and Karayiannis (1996, 1998) described different kinds of search methods applicable for training as well as for testing.

2.3. Supervised fuzzy adaptive resonance theory (SF-ART)

The SF-ART, presented in Fig. 2, classifies input samples at two levels. At the first level, the pre-classifier classifies the input samples primarily to M arbitrary classes. The second level, the post-classifier, is a special array called affine look-up table (ALT) with M elements. ALT stores

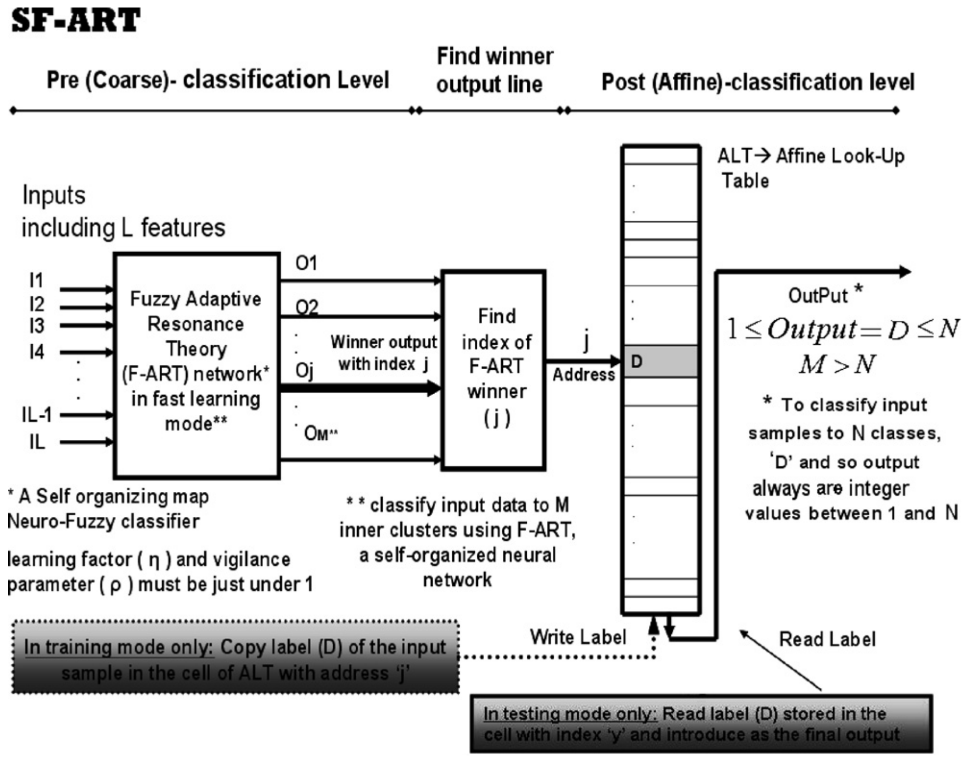

Fig. 2. Supervised fuzzy adaptive resonance theory (SF-ART) structure.

the label (D with a value between 1 and N) of the corresponding input sample in the address equal to the index of the pre-classifier winner (j). In the final step, SF-ART classifies the input dataset to N defined classes which should be less than M. If $N \geqslant M$, using another stage would be useless. In the testing mode, the content of an ALT cell with address equal to the index of the pre-classifier winner (j), which was saved during training mode, will be read. The read value ($y = D$) is the class label to which input data sample belongs.

2.3.1. Pre-classifier

This stage is used to classify the input samples to M inner classes. Output of the pre-classifier is $o_j = F_j(X)$ where F and j are function of the pre-classifier and index of the pre-classifier winner, respectively. As pre-classifier, any unsupervised/supervised classifiers such as neural network, statistical or fuzzy classifier can be used. In the case of supervised pre-classifier, the value of M must be set in advance. For an unsupervised pre-classifier such as Hopfield neural network, the pre-classifier will set the value of M, in a self-organized process.

2.3.2. Affine look-up table (ALT)

ALT is an array with M cells which stores NT labels of the corresponding input samples ($\{1 \leqslant D_i \leqslant N\}_{i=1}^{i=NT}$) in the address equal to the index of the pre-classifier winner (j) in the training mode. M, $N(\leqslant M)$ and NT are number of the inner classes, number of the final classes and number of the training samples, respectively. If $N \geqslant M$, the classifier will not work properly. In the testing mode, a cell with address equal to the index of the pre-classifier winner (j) will be called to pick a label (D_j) that was saved in ALT during training mode. The read value ($1 \leqslant y = \text{ALT}[j] \leqslant N$) is the final output declaring the class label to which input data belongs.

2.3.3. Adaptive resonance theory (ART) neural networks

As a pre-classifier, we can use the adaptive resonance theory (ART) neural network which is a popular self-organized classification method (Carpenter and Grossberg, 1987; Frank et al., 1998; Heins and Tauritz, 1995; Sapozhnikova and Lunin, 2000). ART uses single prototypes to internally represent and dynamically adapt clusters. It uses a minimum required similarity between patterns that are clustered within one cluster. The resulting number of clusters then depends on the similarity between all input patterns, presented to the network during the training cycles. Some interesting features of ART and its capabilities have led to its use in different applications in science and technology. As shown in Fig. 3, any kind of ART-network can be characterized into three levels: pre-processing, searching and adaptation. The pre-processing level tries to create an array with a constant number of elements using an input pattern. The format of this fixed size array depends on the kind of ART network used. When an input pattern is modified to a fixed format in the searching stage, it is compared to the stored templates located in the centre

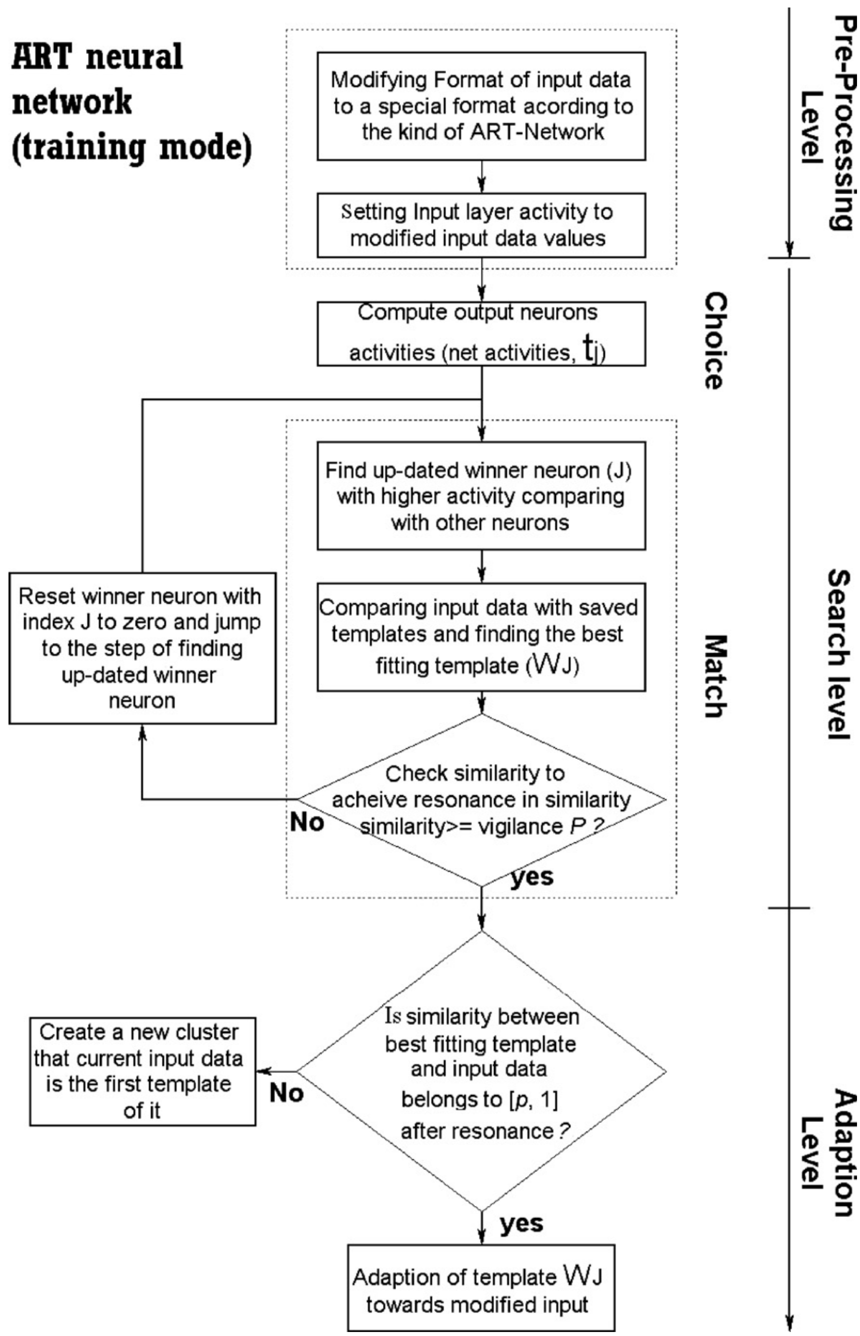

Fig. 3. Learning algorithm of an ART-network.

of existing clusters. In adaptation level, the similarity between best fitting template and the input data is checked. If the degree of similarity between the current input pattern and the best fitting template (J) is at least as high as vigilance ρ (typically limited to the range $[0,1]$), this template is chosen to represent the cluster containing the input. The template is then adapted by shifting the template's values towards the values of the input array. If similarity between input pattern and best fitting template does not fit into the vigilance interval $[\rho,1]$, a new cluster has to be installed where the current input is most commonly used as the first template or cluster centre. For detailed description of these levels, refer to (Akhbardeh, 2007).

The clustering performance of ART-networks is not well documented in the literature, but Frank et al. (1998) concentrated on the comparative analysis of the clustering properties and the performance of several variants of ART-networks. The performance of any kind of classifier depends not only on the network architecture and parameters, but also on the dimensionality and nature of the data to be classified. All kinds of ART-networks are very sensitive to the vigilance parameter rather than the nature and the dimension of the input data. This cause reduced reliability of this net and decreases its popularity.

2.3.4. Fuzzy adaptive resonance theory (F-ART) neural network

Among different kinds of ART-networks, the Fuzzy ART (F-ART) network has better performance compared to other kinds in the interpretation of a given dataset. It uses the 'fuzzy AND logic' in the second level to find similarity between input data sample and stored templates. The above mentioned three levels for F-ART are described in below:

2.3.4.1. Preprocessing step. Before doing any processing, the input data must be normalized into the range [0, 1]. One possible method is to use a Euclidean normalization. But, it loses any information stored in the vector length of an input pattern. To avoid this problem, a modified

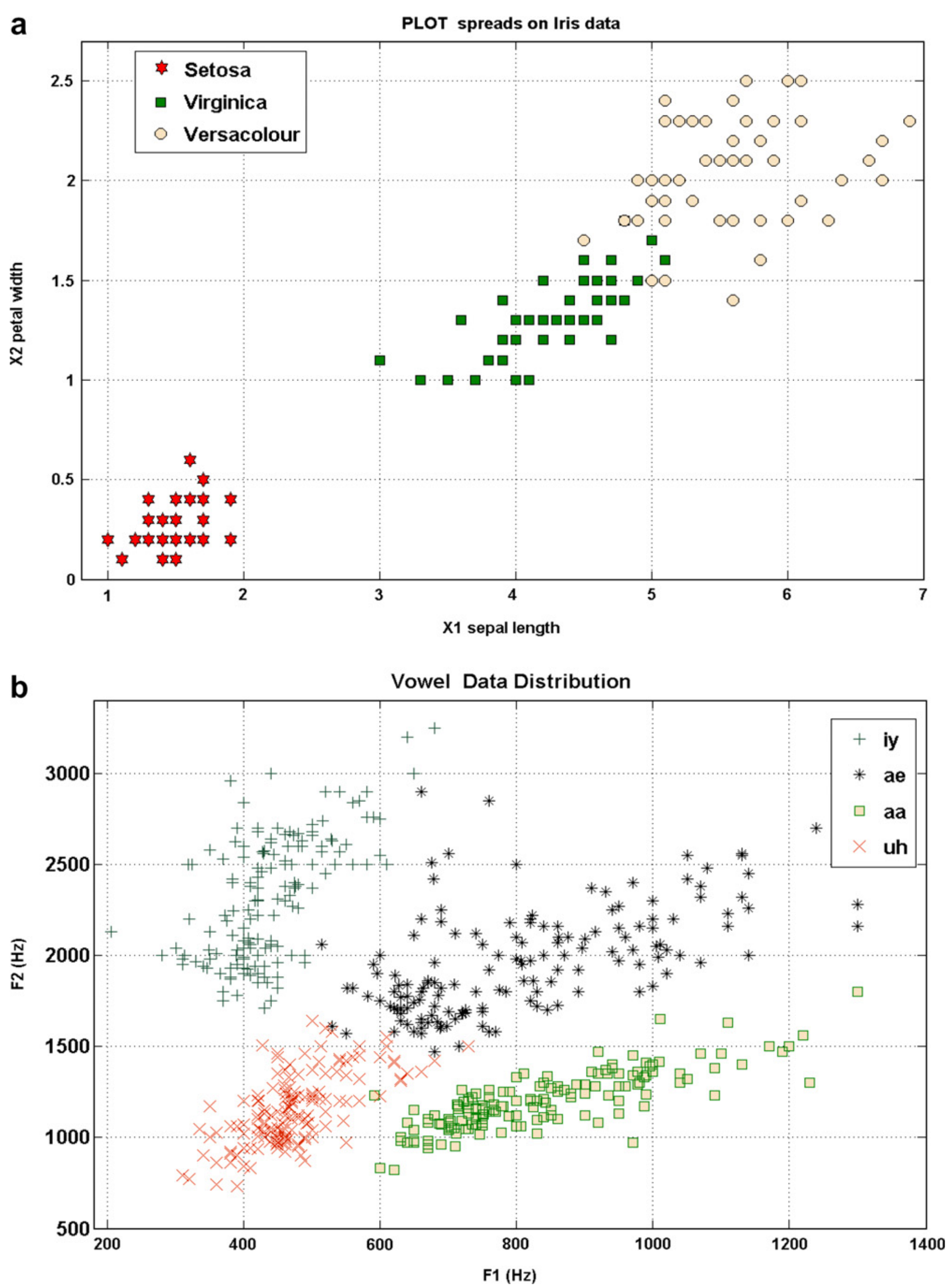

Fig. 4. Pattern recognition data sets used for evaluating SF-ART: (a) 2-D (petal length–petal width) IRIS Data and its distribution which includes three classes (Iris Setosa, Iris Versicolour, and Iris Virginica) and represent different IRIS subspecies and (b) Vowels Data and its distribution which includes four classes (iy, ae, aa, uh) and represent different Vowel subspecies.

normalization method called complement coding can be applied (Frank et al., 1998). In this method original vector $IN = (a_1, a_2, \ldots, a_{L/2})$ can be coded into an input pattern $I = (i_1, i_2, \ldots, i_L)$ by adding its complements to the original vector which doubles the dimension of input patterns: $I = (IN, IN^C) = (a_1, \ldots, a_{L/2}, 1 - a_1, \ldots, 1 - a_{L/2}),\ \ a_i \in [0,1]$ $\forall i$.

2.3.4.2. Search level. This level itself has two sublevels which are described below.

(A) *Choice step:* neurons activity t_j, can be computed using formula $t_j = \frac{|I \wedge W_j|}{\varepsilon + |W_j|}$, where ε is a small constant value to avoid having infinitive t_j when $|W_j| \to 0$. Symbol ($\wedge$) is fuzzy AND logic with definition as following:

$$P \wedge Q = (p_1 \wedge q_1, \ldots, p_K \wedge q_K),$$
$$p \wedge q = \min\{p, q\}.$$

(B) *Match step:* To find similarity between input and current winner template W_J, fuzzy AND logic can be applied. Resonance occurs when $\rho \leqslant \frac{|I \wedge W_J|}{I}$. If similarity between input pattern and best match template is not in the range of $[\rho, 1]$, a new cluster will be created and the current input will be located in its center as the first template. Else, if one of the previously established clusters fits with the input pattern, it is adapted by small shifting of the template's values toward the values of the input array in the next level.

2.3.4.3. Adaptation level. The winner template is adapted by moving its values toward the common MIN vector of I and W_J using formula $W_J = \eta \cdot (I \wedge W_J) + (1 - \eta) \cdot W_J$. where η is learning factor. At the first step of running F-ART, number of clusters is one (No = 1). The template W_1 is initialized with a constant value $w_{i1} = 1,\ 1 \leqslant i \leqslant L$. When new cluster is established (No = No + 1), the template of it will be set to all 1. It guarantees when best fitting template is W_{No}, similarity between input pattern and it does not belong to range $[\rho, 1]$, and the new cluster must be established (No = No + 1). If the best fitting template is not W_{No}, then the fitting template must be adapted by shifting the template's value toward the values of modified input array (I). In adaptation rule, the learning factor $\eta \in [0,1]$ defines the speed of convergence of templates to the common minimum of all input patterns assigned to the same cluster. With η closely near to 1 the network will work in fast learning mode (Frank et al., 1998) and convergence will be occurred after a few presentations of all training patterns. F-ART is too sensitive to the adjustment of similarity parameter (ρ) and learning factor (η) which are typically limited to the range $[0,1]$. By choosing the ρ closely near to 1 the higher number of clusters will be established than choosing ρ with small values. Controlling number of produced clusters is not easy using ρ. This problem caused reducing reliability of this unsupervised neural network and decreasing application of it in different science and technology fields.

SF-ART addresses to these problems. To store corresponding labels of the input patterns in the ALT of SF-ART with a high resolution, the best value for vigilance parameter (ρ) is very close to 1. To converge SF-ART after a few learning cycles (fast learning mode), the learning factor (η) must be set to 1. With these values SF-ART will be an automatic classifier and free from any adjustment of net parameters. The effects of ρ and η are discussed in the next section.

3. Experimental results of SF-ART

This section evaluates the performance of SF-ART as well as existing classifiers (MLP and CNeT) on a variety of pattern classification problems.

3.1. The IRIS data set

The SF-ART algorithm was tested using Anderson's IRIS data set (Anderson, 1939), which has been used exten-

Table 1
IRIS data classification to three classes using SF-ART, MLP, and CNeT classifiers

Classifier	Net parameters	% OP	NLC
SF-ART	$\eta = 1$ and $\rho \to 1$	96.67 ± 1.06	6
CNeT ([11] & [12])	–	94.67	60
MLP with learning rate of 0.001 for all layers	One hidden layer: Nh = 5	97.06 ± 1.65	1500
		97.06 ± 1.65	3000
	One hidden layer: Nh = 10	96.26 ± 1.06	1500
	Nh1 = 15, Nh2 = 10	97.06 ± 1.06	1000
	Nh1 = 20, Nh2 = 10	96.53 ± 2.47	1000

OP means 'overall performance (averaged)' after k-fold (five times) cross validation tests: '±' shows a 95% confidence interval on the average performance (mean). NLC means 'number of learning cycles for training'. η = learning factor, ρ = vigilance (similarity) parameter. Nh1,2 = Number of neurons for hidden layers 1 and 2.

Table 2
Vowel data classification to four classes using SF-ART, and MLP classifiers

Classifier	Net parameters	% OP	NLC
SF-ART	$\eta = 1$ and $\rho \to 1$	98.13 ± 0.45	5
MLP with learning rate of 0.001 for all layers	One hidden layer: Nh = 5	96.23 ± 1.94	1500
	One hidden layer:	91.23 ± 3.29	500
	Nh = 10	97.07 ± 1.33	1000
		97.01 ± 1.05	6000
	Nh1 = 15, Nh2 = 10	98.37 ± 0.58	2500
	Nh1 = 20, Nh2 = 10	98.11 ± 0.55	2500

OP means 'overall performance (averaged)' after k-fold (five times) cross validation tests: '±' shows a 95% confidence interval on the average performance (mean). NLC means 'number of learning cycles for training'. η = learning factor, ρ = vigilance (similarity) parameter. Nh1 and Nh2 = Number of neurons for hidden layers 1 and 2.

sively for evaluating the performance of pattern classification algorithms. This data set contains 150 samples of dimension four that are sepal width, sepal length, petal width, and petal length. These samples can be divided in three classes (Iris Setosa, Iris Versicolour, and Iris Virginica) representing different IRIS subspecies. Setosa class is far from the other two, which have overlap of their features. Fig. 4a shows IRIS dataset in terms of two dimensions (sepal length and petal width) out of four. The 150 samples were randomly split into two sets of 75 samples to obtain training and testing sets. SF-ART was trained with the training set and its ability was evaluated using the testing set.

As comparison, Table 1 shows the performance of SF-ART, MLP and CNeT. SF-ART performed well on the training as well as testing sets. All tests were done on a computer with a 3 GHz Pentium 4 microprocessor. The classification performance was averaged after k-fold

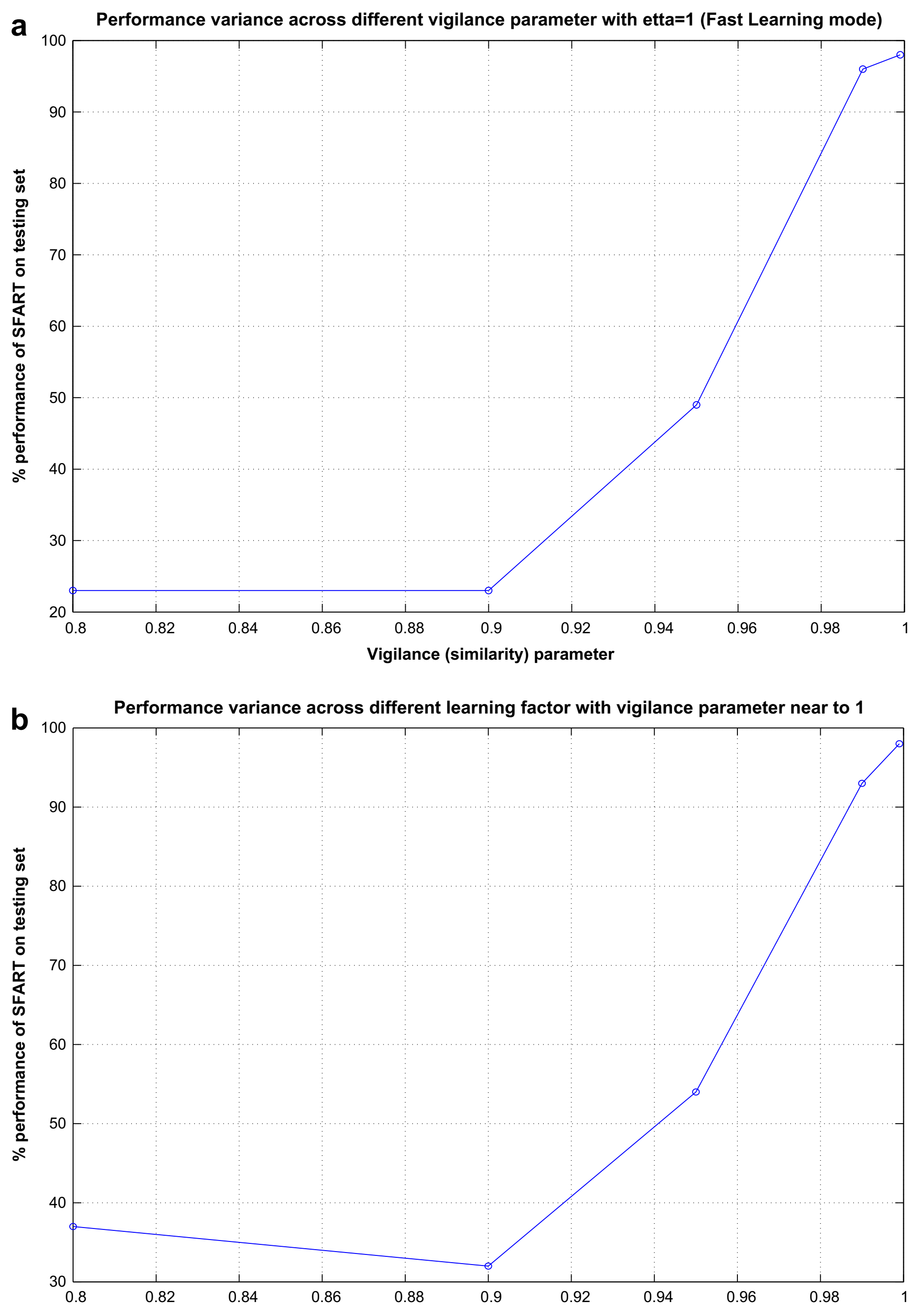

Fig. 5. SFART Performance variances over (a) different vigilance (ρ: similarity) parameter with constant learning factor (η) near to 1 and (b) different learning factor (η) with vigilance (ρ) near to 1. The examples for training and testing classifiers are selected randomly to learn the patterns correctly.

($k = 5$) cross validation tests. The '±' indicates 95% confidence interval on the average performance (mean). For MLP, we found that two hidden layers with 15 and 10 neurons and 1000 iterations of the training data set had the best performance (97.06 ± 1.06) with the testing set (elapsed time: 2.8 s). The SF-ART with $\eta = 1$ and vigilance $\rho = 0.999$ had a high performance (96.67 ± 1.06) with the testing set and convergence occurred after only six iterations (elapsed time: 0.03 s). Referring to the Table 1, maximum classification performance was 98.71 and 97.73, respectively for MLP (One hidden layer: Nh = 5) and SF-ART. The class 'Setosa' is linearly separable from other two classes. However, classes 'Virgincia' and 'Versacolour' in some kernel spaces seem to be separable.

Referring to the results obtained for CNet (Behnke and Karayiannis, 1996, 1998), after almost 60 adaptation cycles, the number of incorrect classification (above four features) with the training set remained almost constant with some fluctuations. The number of classification errors (above two features) with the training set reduced further as the tree kept growing. This is an indication of overtraining, which can be avoided by using the testing-set stopping criterion to terminate the training. Overall, the performance of CNeT in the best situations was above 94.67% on the testing set (Behnke and Karayiannis, 1996, 1998). It must be mentioned that IRIS data is designed in such a way that we cannot reduce training data size and half of the data set must be used for training and the rest of the data for testing the classifier.

3.2. The vowel data set

Another pattern recognition problem used to check the performance of SF-ART is a set of 2-D vowel data (608 samples). The samples belong to four classes which are: "IY" as in "eat", "AE" as in "at", "AA" as in "odd", and "UH" as in "two" (Behnke and Karayiannis, 1996, 1998). Fig. 4b shows vowel formant (a characteristic component of the quality of a speech sound) data for two repetitions of four vowels by 76 speakers. Formants correspond to resonant frequencies of the vocal tract. Vowel data set is a popular data to evaluate pattern classification methods while there are overlaps between three classes. The 608 samples were split into a training set (300 samples) and a testing set (308 samples).

Table 2 shows the comparative analyses and performance of SF-ART, and Multi layer perceptrons (MLP) applied for the classification of vowel data into four classes. The classification performance was averaged after k-fold ($k = 5$) cross validation tests. The '±' indicates

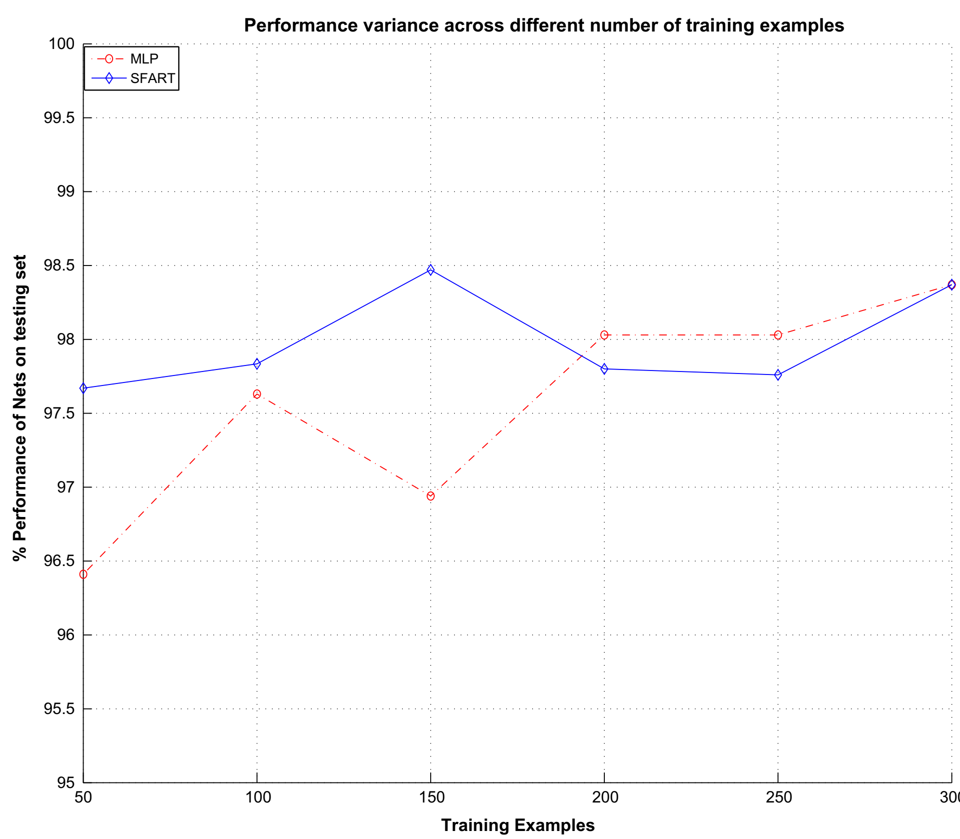

Fig. 6. MLP and SFART performance variances over a different number of training examples. For SFART learning factor (η) and vigilance (ρ: similarity) were chosen near to 1. Also, for MLP with two hidden layers, numbers of neurons in hidden layers were Nh1 = 15, Nh2 = 10. The examples for training and testing classifiers are selected randomly to learn the patterns correctly. The training and testing data sets were the same for MLP and SF-ART.

95% confidence interval on the average performance (mean). All tests were done on a 3 GHz Pentium 4 computer.

Using MLP having two hidden layers, the convergence during training occurred after 2500 iterations; and it had a performance of 98.37 ± 0.58 with testing dataset. Elapsed time for learning was 3.2, 5.4, 26.7 and 34.3 s, respectively, for the MLP with $Nh = 5$, $Nh = 10$, ($Nh1 = 15$ and $Nh2 = 10$), and ($Nh1 = 20$ and $Nh2 = 10$) structures.

However the SF-ART, with η and ρ close to 1, needed only five learning cycles (250 ms) to achieve similar performance (98.13 ± 0.45) with the test dataset. Fig. 5a indicates the performance variances of SF-ART across different values for vigilance parameter (ρ) when the learning factor is constant ($\eta = 1$). Fig. 5b indicates the performance variances of SF-ART for different values of learning factor (η) when vigilance parameter (ρ) is constant. Fig. 6 indicates the effects of training samples on the performance of SF-ART and MLP. As can be seen from Fig. 6, even using a small amount of the training samples, SF-ART is more stable than MLP across different volumes of training samples. However, in SF-ART the network capacity for generalization slightly decreased (small fluctuation between 97.8% and 98.5%) if more than 200 examples (30% of vowels data) were used to train SF-ART. MLP seems to be slightly more stable than SF-ART as can be seen from Tables 1 and 2. MLP has a lower confidence interval over mean classification performance.

Table 3
Ten classes' vowel data classification using CNeT, MLP, and k-NN classifiers (Behnke and Karayiannis, 1996, 1998)

Classifier	Net parameters	% OP
CNeT	–	82.04
k-NN	–	75.45
MLP	5 Hidden units	76.58
	10 Hidden units	80.18

OP means 'overall performance'.

Behnke and Karayiannis (1996, 1998) used their own 2-D vowel data with ten classes, instead of the four classes that we used, to evaluate CNeT performance. The ten vowel classes which they used came from: "head", "hid", "hod", "had", "hawed", "heard", "heed", "hud", "who'd", and "hood". The available 608 feature vectors were divided into a training set, containing 300 vectors, and a testing set, containing 308. They compared CNeT performance with other existing methods such as K-Nearest Neighbor (K-NN) classifier and MLP with five and ten hidden units trained using gradient descent. The classifica-

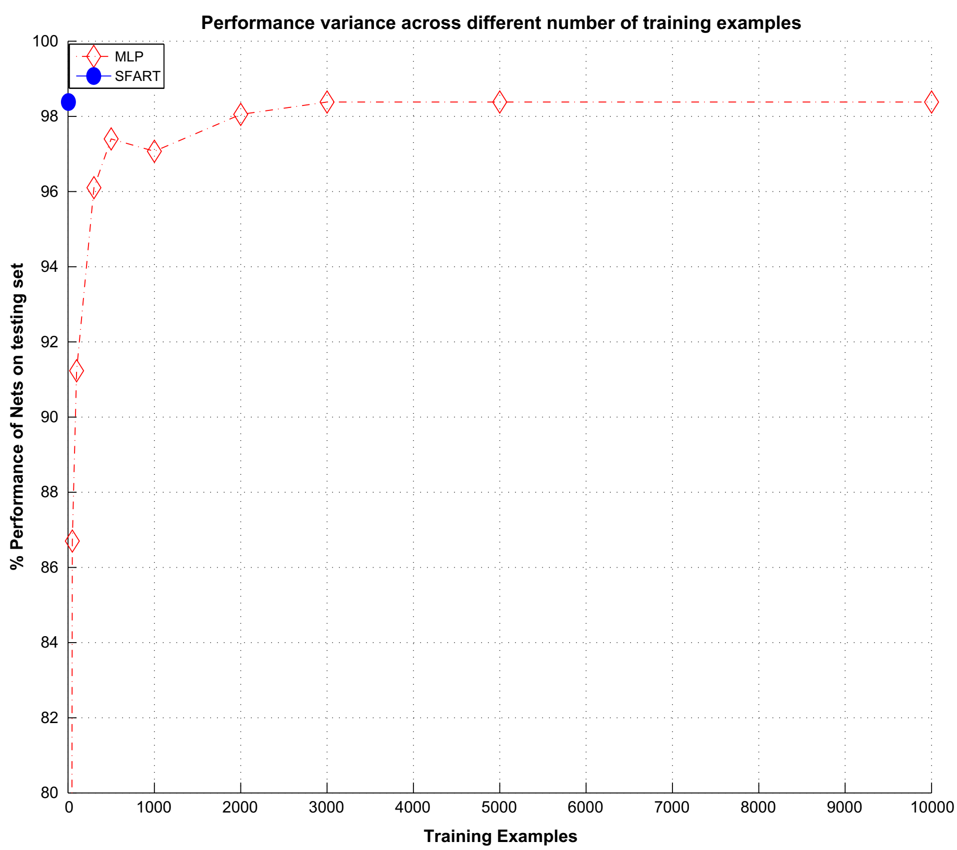

Fig. 7. MLP and SF-ART performance variances over a different number of learning cycles during training mode. For SF-ART learning factor (η) and vigilance (ρ: similarity) were chosen near to 1. As SF-ART stops automatically (after five adaptation cycles), we have only one point for SF-ART. Also, for MLP with two hidden layers, the numbers of neurons in hidden layers were $Nh1 = 15$, $Nh2 = 10$. The examples for training and testing classifiers were selected randomly to learn the patterns efficiently.

tion performance for CNeT, K-NN, and MLP with five hidden units and ten hidden units, were 82.04%, 75.45%, 76.58% and 80.18% for the testing set, respectively (Table 3). As can be seen the CNeT had better performance than MLP and K-NN to classify vowel data to ten classes and learnt faster. But it needed more than 60 learning cycles to converge on the training set with some fluctuations in the number of incorrect classifications. CNeT is a heavy algorithm and the time required for training is high compared to SF-ART.

In another test, we compared the number of learning cycles and its effects on the performance of existing

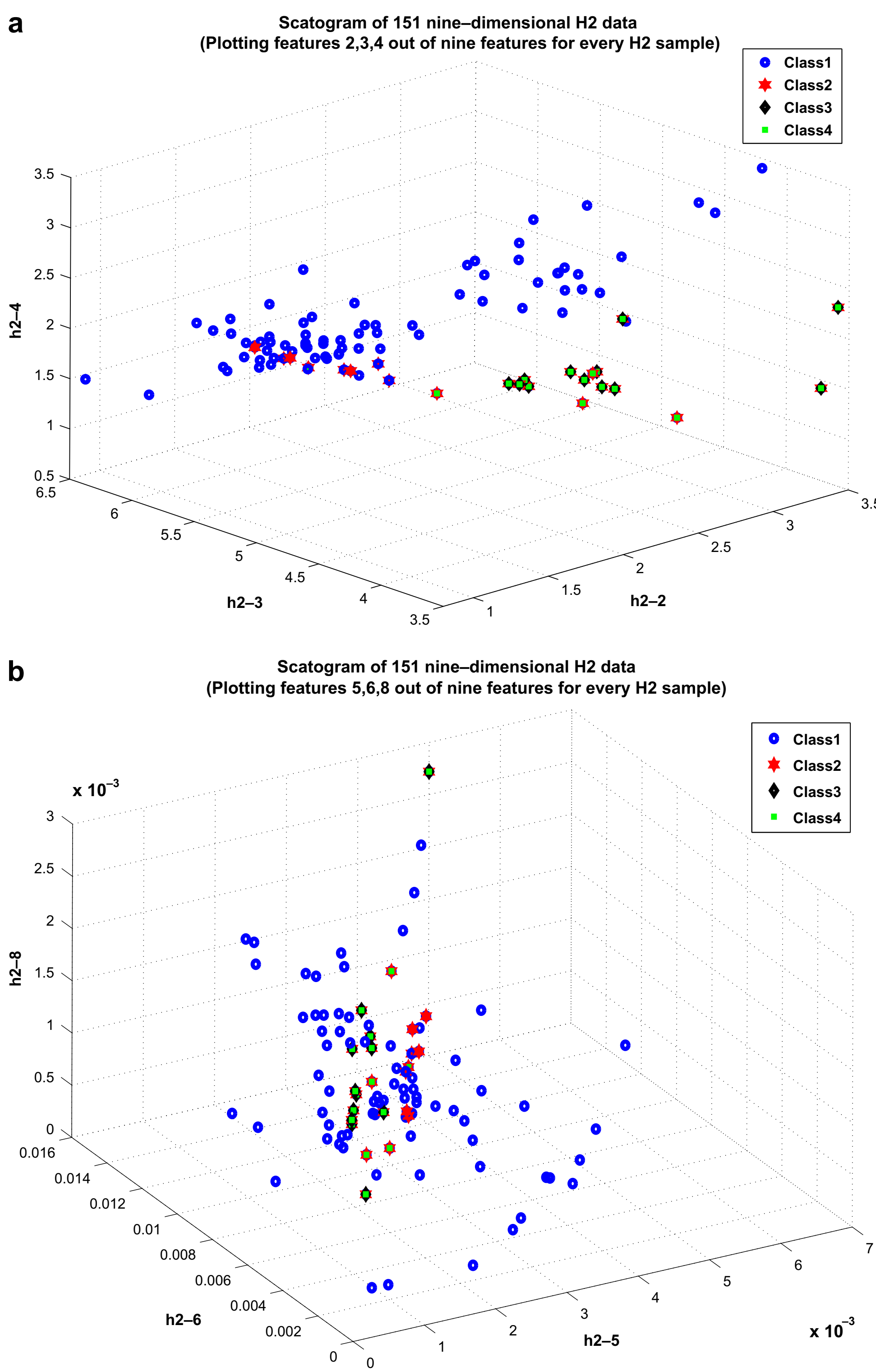

Fig. 8. 3-D Representation of three out of nine features of 150 H_2 samples: (a) features 2, 3, 4: H_2 production rate, CO_2 production rate and pH, respectively and (b) features 5, 6, 8: ethanol, acetate, butyrate, respectively.

Table 4
H_2 data classification to four classes using SF-ART, and MLP classifiers

Classifier	Net parameters	% OP	NLC
SF-ART	$\eta = 1$ and $\rho \to 1$	98.40 ± 1.50	5
MLP with learning rate of 0.001 for all layers	One hidden layer: Nh = 15	98.60 ± 1.20	1500
	One hidden layer: Nh = 10	96.80 ± 1.49	1500
	Nh1 = 15, Nh2 = 10	97.40 ± 2.35	2000
	Nh1 = 20, Nh2 = 10	96.20 ± 1.74	2000

OP means 'overall performance (averaged)' after *k*-fold (five times) cross validation tests: '±' shows a 95% confidence interval on the average performance (mean). NLC means 'number of learning cycles for training'. η = learning factor, ρ = vigilance (similarity) parameter. Nh1 and Nh2 = Number of neurons for hidden layers 1 and 2.

supervised neural networks such as MLP (see Fig. 7). In most cases, neural networks need the adjustment of a stopping criterion in the training mode using mean square error, a manual limit or other kinds of criteria. However, SF-ART is free from any adjustment of the stopping criterion because it stops automatically after a resonance occurrence in the training mode. The convergence was done very fast (after five learning cycles for the vowel data set) because SF-ART uses the fast learning mode of F-ART. Therefore, in Fig. 7, there is only one point with position (iteration number = 5, performance = 98.38%) in the figure for SF-ART. This ability helps SF-ART to be a fully automatic neural network based classifier.

3.3. Hydrogen (H_2) data set

The classification performance of above mentioned classifiers was analyzed and compared on a biological (H_2) dataset. In this study, our aim was the evaluation of the SF-ART from a methodological point of view of using H_2 data.

The H_2 dataset consisted of nine dimensions (Hydraulic retention time, H_2 production rates, CO_2 production rates, pH, acetate, ethanol, butyrate, valerate and propionate) and 150 samples. The dataset consists of four class labels. These class labels were given after analyzing the metabolic profile of each sample. The idea of using such dataset for comparative analyses of other computational methods was to find out whether other methods are also able to group data into similar clusters.

The dataset was obtained by monitoring the hydrogen producing bioreactor. An anaerobic, completely mixed bioreactor (total volume 0.8 l, height to diameter ratio 7.7) with a gas extraction module was used for H_2 production at 35 °C. Bioreactor was operated continuously for 156 days and reactor performance was determined by measuring gaseous and soluble end products, glucose degradation and biomass concentrations.

Fig. 8 shows H_2 dataset in terms of three dimensions out of nine. The 150 samples were randomly split into two sets of 75 samples to obtain training and testing sets.

As a comparison, Table 4 shows the performance of SF-ART and MLP. SF-ART performed well on the training as well as testing sets. All tests were done on the computer with a 3 GHz Pentium 4 microprocessor. The classification performance was averaged after *k*-fold ($k = 5$) cross validation tests. The '±' indicates 95% confidence interval on the average performance (mean). For MLP, we found that one hidden layer with 15 neurons and 1500 iterations of the training data set had the best performance (98.60 ± 1.20) with the testing set (elapsed time: 3 s). The SF-ART with learning parameter ($\eta = 1$) and vigilance parameter ($\rho = 0.999$) had a high performance (98.40 ± 1.50) with the testing set. The convergence occurred after only six iterations (elapsed time: 0.08 s). Referring to the Table 4, maximum classification performance was 100%, respectively for MLP (One hidden layer: Nh = 15) and SF-ART.

4. Conclusions

In this paper, we presented a new pattern recognition method (SF-ART) with a higher learning speed, a lower computational load, higher performance, and fewer parameters to adjust compared to classical methods. The results indicated that SF-ART learned patterns (of two well-known datasets, IRIS, Vowel) very fast, in less than 1 s for SF-ART. For comparison, MLP and CNeT needed seconds to train. We also tested SF-ART using H_2 dataset. Classifiers, except SF-ART, were sensitive to the volume of the training set as well as the number of adaptation cycles during training mode. They also suffer from trade-offs between learning speed and performance. Moreover, the results showed that SF-ART learnt patterns with a lower amount of training samples. When the shortest possible training time is important and slightly lower classification performance is allowed, SF-ART can be recommended. One interesting benefit of SF-ART can be incremental learning when new data become available (on-line learning). This capability has not yet been fully explored. Therefore, in future SF-ART should be applied in classification of more patterns, especially testing its reliability, stability and performance for on-line (incremental) learning.

Acknowledgements

This research was funded by the Academy of Finland (HYDROGENE Project, No. 107425), Nordic Energy Research (BIOHYDROGEN Project, No. 28-02) and Tampere University of Technology Graduate School (P.E.P. Koskinen). The work was also supported by the Academy of Finland (Application Number 213462, Finnish Programme for Centers of Excellence in Research 2006-2011).

References

Akhbardeh, A., Signal Classification Using Novel Pattern Recognition Methods and Wavelet Transforms, Ph.D. Thesis, Institute of Signal processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland, March 2007.

Anderson, E., 1939. The IRISes of the Gaspe peninsula. Bull. Amer. IRIS Soc. 59, 2–5.

Behnke, S., Karayiannis, N.B., 1996. CNeT: competitive neural trees for pattern classification. Proc. IEEE Internat. Conf. Neural Networks 3, 1439–1444.

Behnke, S., Karayiannis, N.B., 1998. CNeT: competitive neural trees for pattern classification. IEEE Trans. Neural Networks 9, 1352–1369.

Bishop, C.M., 2005. Neural Networks for Pattern Recognition. Oxford University Press, UK.

Carpenter, G.A., Grossberg, S., 1987. The ART of adaptive pattern recognition by a self-organizing neural network. IEEE Trans. Computer, 77–88.

Carpenter, G.A., Grossberg, S., 1991. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. Neural Networks 4, 565–588.

Carpenter, G.A., Grossberg, S., 1992. Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimentional maps. IEEE Trans. Neural Networks 3, 698–713.

Frank, T., Kraiss, K.F., Kuhlen, T., 1998. Comparative analysis of fuzzy ART and ART-2A network clustering performance. IEEE Trans. Neural Networks 9, 544–559.

Haykin, S., 1998. Neural Networks, a Comprehensive Foundation, second ed. Prentice Hall PTR, New Jersey, USA.

Heins, L.G., Tauritz, D.R., 1995. Adaptive Resonance Theory (ART): An Introduction, Technical Report, Leiden University, pp. 95–35.

Joao, B.D. Cabrera, 2006. Cabrera, On the impact of fusion strategies on classification errors for large ensembles of classifiers. Pattern Recognition 39 (11), 1963–1978.

Kam, H.T., Hull, J.J., Srihari, S.N., 1994. Decision combination in multiple classifier systems. IEEE Trans. Pattern Anal. Machine Intell. 16, 66–75.

Polikar, R., Udpa, L., Udpa, S.S., Honavar, V., 2001. Learn++: an incremental learning algorithm for supervised neural networks, IEEE Trans. Systems Man Cybernet., Part C, vol. 31, pp. 497–508.

Polikar, R., Byorick, J., Krause, S., Marino, A., Moreton, M., 2002. Learn++: a classifier independent incremental learning algorithm for supervised neural networks. In: Proc. 2002 Internat. Joint Conf. on Neural Networks (IJCNN'02), pp. 1742–1747.

Sapozhnikova, E.P., Lunin, V.P., 2000. A modified search procedure for the ART neural networks. In: Proc. 2000 Internat. Joint Conf. on Neural Networks (IJCNN'00), vol. 5, pp. 541–544.

Teredesai, A., Govindaraju, V., 2005. GP-based secondary classifiers. Elsevier J. Pattern Recognition 38, 505–512.

Xu, L., Krzyzak, A., Suen, C.Y., 1992. Methods of combining multiple classifiers and their applications to handwriting recognition. IEEE Trans. Systems Man Cybernet. 22, 418–435.

Zhang, Ping, 2006. Reliable Recognition of Handwritten Digits Using a Cascade Ensemble Classifier System and Hybrid Features, Doctoral Thesis, Computer Science Department, Concordia University, Montreal, April.