On Tree Types of Competitive Learning Algorithms with Their Comparisons and Applications to MRI Segmentation

Miin-Shen Yang,1,* Kuo-Lung Wu,2 Karen Chia-Ren Lin,3 Hsiu-Chih Liu,4 Jiing-Feng Lirng5
1Department of Applied Mathematics, Chung Yuan Christian University, Chung-Li 32023, Taiwan
2Department of Information Management, Kun Shan University of Technology, Tainan 71023, Taiwan
3Department of Management Information System, Nanya Institute of Technology, Chung-Li 32023, Taiwan
4Neurological Institute, National Yang-Ming University and Taipei Veterans General Hospital, Taipei, Taiwan
5Department of Radiology, National Yang-Ming University and Taipei Veterans General Hospital, Taipei, Taiwan

This paper considers competitive learning networks using three types of hard, soft, and fuzzy learning schemes. The hard competitive learning algorithm is with the winner-take-all. The soft competition learning algorithm is with a stochastic relaxation technique using the Gibbs distribution as a dynamic neighborhood function. The fuzzy competition learning algorithm is with a fuzzy relaxation technique using fuzzy membership functions as kernel type neighborhood interaction functions. Some numerical examples are made for these three types of competitive learning schemes. The numerical results show that the fuzzy learning has better performance than hard and soft learning under the normal mixture data. We then present an application to magnetic resonance image segmentation. A real case of ophthalmology recommended by a neurologist with MR image data is examined in this paper. These competitive learning algorithms are used in segmenting the ophthalmological MRI data for reducing medical image noise effects with a learning mechanism. Based on the segmentation results, the fuzzy learning gives better performance than hard and soft learning so that the fuzzy competitive learning algorithm is recommended for use in MRI segmentation as an aid for support diagnoses. © 2010 Wiley Periodicals, Inc.

*Author to whom all correspondence should be addressed: e-mail: msyang@math.cycu.edu.tw.
1. INTRODUCTION

Artificial neural networks have been studied for many years and widely applied in various areas such as image processing, signal processing, pattern recognition, and vector quantization. Lippmann and Kohonen gave tutorial reviews on this neural computing. Lippmann reviewed six important neural net models that can be used for pattern classification. In these interesting neural net models, the competitive learning networks are important and will be discussed in this paper.

The competitive learning networks are motivated by both anatomical and physiological evidence of the lateral interaction between neurons in the mammalian nervous system. The simplest computing method in a competitive learning network is winner-take-all (WTA). However, there is both anatomical and physiological evidence from mammalian brains that indicates the lateral neural interaction is like a “Mexican-hat” function. Approximation of the lateral neural interaction can be found in Kohonen’s self-organizing map (SOM). Kohonen’s SOM uses the neighborhood interaction set to approximate lateral neural interaction and discover the topological structure hidden in the data. To produce a more precise approximation, Ritter and Schulten and Lo and Bavarian introduced the Gaussian function as the neighborhood interaction function.

Yair et al. proposed a more powerful learning scheme, called the soft competition scheme (SCS), by incorporating the stochastic relaxation technique principles. They used the Gibbs distribution as a dynamic neighborhood function and then developed a deterministic algorithm that updates all of the codevectors in the vector quantizer simultaneously, rather than one at a time. Overall, the competitive learning network is an approach to unsupervised learning and is used most in vector quantizer design and clustering.

The k-means (or called hard c-means) clustering is a batch algorithm for designing a vector quantizer, which is a mapping of input vectors to one of c predetermined codevectors (also called codebooks). Fuzzy c-means (FCM) clustering is a fuzzy extension of hard c-means clustering. The FCM and its varieties have been widely studied and applied in various areas. Fuzzy membership functions from FCM algorithms give good cluster membership interpretations. Incorporating these fuzzy membership functions into neural networks becomes a reasonable way to make fuzzy types of neural computing. Keller and Hunt first incorporated fuzzy membership functions into the perceptron algorithm and created a so-called fuzzy perceptron. For a batch-type learning vector quantization (LVQ), Wu and Yang proposed a fuzzy-soft LVQ (FS-LVQ) based on the SCS concept and fuzzy membership functions. In this paper, we modify the batch-type FS-LVQ to be a sequential type. We then create the fuzzy-soft competitive learning algorithm (FS-CLA) by incorporating a fuzzy relaxation technique using fuzzy membership functions as a kernel type of neighborhood interaction function in Kohonen’s SOM.

The remainder of the paper is organized as follows. In Section 2, we first review the two types of competitive learning networks with WTA and SCS. We then modify Wu and Yang’s (batch-type) FSLVQ to be a sequential type and create the FS-CLA competitive learning algorithm. We also analyze the differences between WTA, SCS, and FS-CLA. In Section 3, some mixture model numerical examples
are presented. A criterion for monitoring the vector quantization is proposed. We give numerical comparisons to show the performance based on the mean squared error (MSE) criterion when using the three learning schemes. In Section 4, we apply them to the magnetic resonance imaging (MRI) segmentation. A real case recommended by a neurologist is examined. The patient suffered bilateral internuclear ophthalmoplegia, which indicates lesions in the bilateral medial longitudinal fasciculus in the tegmentum of the dorsal pons. The brain MRI revealed lacunar infarctions in the bilateral lentiform nuclei, thalami, and corona radiate, but not in the pons. For the purpose of enhancing the lesions from the MRI noise, these competitive learning based MRI segmentation techniques are used to reduce medical image noise effects with learning mechanisms. The results from segmenting the MRIs of ophthalmology show that FS-CLA has better performance than WTA and SCS. Finally, our conclusions are made in Section 5.

2. THREE TYPES OF COMPETITIVE LEARNING ALGORITHMS

Neural network models can roughly be divided into three categories: feedforward networks (e.g., multilayer perceptron), feedback network (e.g., Hopfield network) and competitive learning network (e.g., SOM). Both feedforward and feedback networks are supervised. The competitive learning network is unsupervised. Competitive learning is motivated by the anatomical and physiological evidence of lateral interaction between neurons in mammalian nervous systems. The lateral neural interaction of competitive learning can be approximated using the well-known WTA principle. Only the winner in neural learning is allowed while other neurons are inhibited, as shown in the sixth row in Figure 1. Suppose that the s-dimensional feature vector \( x_j, j = 1, \ldots, n \), is the input data at time \( t \), the Euclidean winner among the \( c \) neurons \( Z_i, i = 1, \ldots, c \) is then produced upon the nearest neighbor condition. That is, neuron \( i \) is the winner at time \( t \) when \( x_j \) is input if

\[
\| x_j - Z_i(t - 1) \| = \min_{1 \leq k \leq c} \| x_j - Z_k(t - 1) \|. \tag{1}
\]

The learning can then be described according to the learning formula

\[
\begin{align*}
Z_i(t) &= Z_i(t - 1) + \alpha(t)(x_j - Z_i(t - 1)) \quad \text{if } Z_i(t - 1) \text{ is the winner,} \\
Z_k(t) &= Z_k(t - 1) \quad \text{if } k \neq i,
\end{align*}
\tag{2}
\]

where \( \alpha(t) \) is called the learning rate and is confined to decrease monotonically with time \( t \). This is a very important factor that determines the gain for the winner neuron and the network performance. The winner after competitive learning will be updated toward the input vector with a step size \( \alpha(t) \). Other neurons are unchanged. The learning will stop when the network is stable. The WTA learning algorithm as summarized as follows:

International Journal of Intelligent Systems DOI 10.1002/int
**WTA algorithm**

Initialize $Z_1(0), \ldots, Z_c(0)$ and $\alpha$;

For $t = 1, \ldots, T$; $\alpha(t) = \alpha/t$;

For $j = 1, \ldots, n$;

Update $Z_i(t)$ using (1) and (2);

Next $j$;

Next $t$.

This WTA principle, based on Equations (1) and (2), will ignore global information about the geometric structure of the data represented in the remaining $(c - 1)$ neurons. There is both anatomical and physiological evidence from the mammalian nervous system that the lateral neural interaction is like a “Mexican-hat” function. This neural network model is called the self-organizing map (SOM) proposed by Kohonen.\textsuperscript{7,8}

The SOM is a two-layer feedforward competitive learning neural network that can find the topological structure hidden in the data. Different from the WTA principle, SOM uses “share bliss” competitive learning to realize the lateral neural interaction. Neurons that surround the excited neuron will also have different degrees of excitation or inhibition. Kohonen used the neighborhood interaction set to approximate the lateral neural interaction phenomenon. Suppose neuron $i$ is the winner at time $t$ and $N_i(t)$ denotes the neighborhood interaction set of neuron $i$ at time $t$ according to rule (1), the SOM learning can be described according to the

---

**Figure 1.** Approximated lateral interactions using different methods.
Learning formula

\[ Z_k(t) = Z_k(t-1) + \alpha_i(t)h_{i,k}(t)(x_j - Z_k(t-1)), \quad i = 1, \ldots, c, \quad j = 1, \ldots, n, \]

(3)

where \( h_{i,k}(t) \) denotes the degree of excitation of the neurons. Kohonen\(^7\) used the simple function \( h_{i,k}(t) \) with

\[ h_{i,k}(t) = \begin{cases} 1 & \text{if } Z_k(t-1) \in N_i(t) \\ 0 & \text{if } Z_k(t-1) \notin N_i(t). \end{cases} \]

(4)

This is shown in the fifth row in Figure 1. The neighborhood set \( N_i(t) \) is requested to decrease for accomplishing the WTA principle as time \( t \) increases. However, the excited signals in the lateral neural interaction are not constant. This either increases or decreases as the neural distance increases. For a more precise approximation, Kohonen\(^8\), Ritter and Schulten\(^9\), and Lo and Bavarian\(^10\) used the Gaussian function to model the lateral neural interaction. That is,

\[ h_{i,k}(t) = \begin{cases} \exp\left(-\frac{\|Z_i(t-1) - Z_k(t-1)\|^2}{2\sigma^2(t)}\right) & \text{if } Z_k(t-1) \in N_i(t) \\ 0 & \text{if } Z_k(t-1) \notin N_i(t). \end{cases} \]

(5)

The \( \sigma(t) \) corresponds to the kernel width of the neighborhood neurons and determines the degree of excitation of the neurons, which also decreases monotonically with time \( t \). The approximation is shown in the fourth row in Figure 1. We mention that if we take \( h_{i,k}(t) \) to be the Gaussian function and consider a level-cut of \( h_{i,k}(t) \) as

\[ h_{i,k}^\alpha(t) = \begin{cases} 1 & \text{if } h_{i,k}(t) \geq L, \\ 0 & \text{if } h_{i,k}(t) < L, \end{cases} \]

(6)

where \( L \) denotes a level value, the above definition will then reduce to a simple function (4).

Although the Gaussian interaction function (5) can improve the convergence rate of SOM\(^9,10\), the neighborhood concept will be restricted and not relative to the data. Yair et al.\(^11\) proposed another learning method called the soft competition scheme (SCS). SCS used the Gibbs distribution to describe the lateral interaction function \( h_{i,j}(t) \) with

\[ h_{i,j}(t) = \frac{e^{-\beta(t)\|x_j - Z_i(t-1)\|^2}}{\sum_{k=1}^{c} e^{-\beta(t)\|x_j - Z_k(t-1)\|^2}}, \]

(7)

where \( \beta(t) \) corresponds to the temperature with

\[ \beta(t) = \frac{\hat{r}^{1/c}}{T_0} \]

(8)
\[ \lim_{t \to \infty} \beta(t) = \infty. \] 

(9)

This is somehow analogous to simulated annealing. The neurons are then updated according to the learning formula

\[ Z_i(t) = Z_i(t-1) + \alpha_{ij}(t)h_{ij}(t)(x_j - Z_i(t-1)). \] 

(10)

They noted that the performance of this system was not heavily dependent upon the choice of \( \hat{r}^{1/c} \) where \( \hat{r}^{1/c} = 1.05 \) works well for all of the sources. The initial temperature was chosen to be \( T_0 = 1/dc \sigma^2_x \), where \( \sigma^2_x \) is the variance in the input vector \( x \) and \( d \) is a constant to be determined experimentally. The excitation approximation is shown in the first row in Figure 1. Comparing \( h_{ij}(t) \) in (7) with \( h_{i,k}(t) \) in (4) and (5), we see that the SCS presented three different points of view from Kohonen’s neighborhood interaction functions \( h_{i,k}(t) \) as follows (see also Yair et al.\(^{11}\)):

(a) The lateral interaction function \( h_{ij}(t) \) is a dynamic function that weighs the codevectors using their distance from the data in the input space, rather than their distance on the lattice presented in the neighborhood function \( h_{i,k}(t) \).

(b) All of the codevectors are simultaneously updated at each time step such that the \( i \)th codevector is shifted to the input data \( x_j \).

(c) There is no ultimate winner, but rather each codevector is updated using a certain proportion towards the data. This saves time in determining the winner in each loop of the algorithm.

The learning rate \( \alpha_{ij}(t) \) in the SCS is computed using

\[ \alpha_{ij}(t) = 1/n_i(t), \] 

(11)

where \( n_i(t) = 1 \), if \( t \) is a perfect square and \( n_i(t) = n_i(t-1) + h_{ij}(t) \), otherwise. They noted that the re-initialization technique for the perfect square time \( t \) is useful in practice to speed up the convergence. We then summarize the SCS learning algorithm as follows:

**SCS algorithm**

1. Initialize \( Z_1(0), \ldots, Z_c(0), \hat{r} \) and \( T_0 \);
2. For \( t = 1, \ldots, T \); \( \beta(t) = \hat{r}^{1/c}/T_0 \);
   1. For \( j = 1, \ldots, n \);
      1. Estimate \( h_{ij}(t) \) using (7);
      2. Update \( \alpha_{ij}(t) \) using (11);
      3. Update \( Z_i(t) \) using (10);
   3. Next \( j \);
3. Next \( t \).

We mention that the \( k \)-means (or hard \( c \)-means) clustering can be thought of as a batch version of the WTA algorithm. The fuzzy \( c \)-means (FCM) clustering is
a fuzzy extension of hard \( c \)-means clustering. The FCM and its varieties have been widely studied and applied in various areas. The fuzzy membership functions from FCM give a good interpretation of cluster memberships for most real cases. Let \( X \) be a subset of an \( s \)-dimensional Euclidean space \( \mathbb{R}^s \). A fuzzy \( c \)-partition \( \{\mu_1, \ldots, \mu_c\} \) of \( X \) into \( c \) clusters consists of the fuzzy sets \( \mu_i \) assuming values in the interval \([0, 1]\) such that \( \sum_{i=1}^{c} \mu_i(x) = 1 \) for all \( x \) in \( X \). The FCM is a clustering algorithm to find optimal fuzzy \( c \)-partitions and optimal \( c \)-means for a data set \( X = \{x_1, \ldots, x_n\} \). Tsao et al., Bezdek and Pal constructed the so-called fuzzy Kohonen clustering networks (FKCN) and fuzzy learning vector quantization (FLVQ). FKCN and FLVQ were proposed to integrate the FCM clustering with the Kohonen network. FKCN and FLVQ can be seen as a (batch) Kohonen type of learning FCM clustering using a decreasing fuzzifier (fuzziness) \( m_t \) of the FCM objective function to 1 (i.e., \( m_t \to 1 \) as \( t = 1, 2, \ldots, t_{\text{max}} \)) so that they can converge faster than the FCM clustering. In Wu and Yang, a batch-type LVQ, called a fuzzy-soft LVQ (FSLVQ), based on the SCS concept and fuzzy membership functions which has a different learning mechanism from FKCN and FLVQ was proposed. Wu and Yang also provided numerical comparisons of FSLVQ to FLVQ and FCM.

Next, we will modify the FSLVQ to be a sequential type. We call it a fuzzy-soft competitive learning algorithm (FS-CLA). The FS-CLA can be seen as a sequential type of FSLVQ. The neural lateral interaction and learning rates are approximated using fuzzy membership functions and all neurons are updated according to the learning formula (10). FS-CLA is constructed as a parameter-free scheme. We define \( \mu_i(x_j) \) as the fuzzy membership of \( x_j \) that \( Z_i(t-1) \) wins. In this paper, we adopted the well-known FCM membership functions

\[
\mu_i(x_j) = \left( \sum_{k=1}^{c} \frac{\|x_j - Z_i(t-1)\|^2/(m-1)}{\|x_j - Z_k(t-1)\|^2/(m-1)} / \sum_{k=1}^{c} \frac{\|x_j - Z_k(t-1)\|^2/(m-1)}{\sum_{k=1}^{c} \|x_j - Z_k(t-1)\|^2/(m-1)} \right)^{-1}
\]  

and \( m = 2 \) is typically used. We then approximate the neural lateral interaction using

\[
h_{ij}(t) = \left[ \frac{\mu_i(x_j)}{\max_{1 \leq i \leq c} \{\mu_i(x_j)\}} \right]^{(1 + \frac{t_{\text{max}}}{t})}, \quad i = 1, \ldots, c,
\]

where \( f(t) \) is a positive strict monotone increasing function of \( t \) that controls the degree of excitation. Typically, \( f(t) = t \) or \( f(t) = \sqrt{t} \). We used the function \( f(t) = \sqrt{t} \) in this paper. The approximated excitations are shown in the second row in Figure 1.
In the fuzzy-soft competitive learning, \( f(t) \) is confined as strictly increasing with

\[
\lim_{t \to \infty} f(t) = \infty
\]  

(14)

and the competitive learning is soft with

\[
0 \leq h_{ij}(t) \leq 1.
\]  

(15)

It is easy to show that when \( t \) tends to infinity, the learning rule in the fuzzy-soft network will tend toward WTA. That is,

\[
\lim_{t \to \infty} h_{ij}(t) = \begin{cases} 
1, & \text{if } Z_i(t - 1) \text{ satisfies the nearest neighbor condition (1)} \\
0, & \text{otherwise.}
\end{cases}
\]

When \( t \) tends to infinity, only the neuron that is the closest to the input data will be excited and other neurons will be inhibited. The function \( f(t) \) can determine the decreasing rate from fuzzy-soft competitive learning to hard competitive learning (i.e., WTA). For example, using \( f(t) = t^2 \) will have a more strenuous inhibition (small excited states) than using \( f(t) = t \) and hence the use of \( f(t) = t^2 \) will have a faster decreasing rate in \( h_{ij}(t) \) than the use of \( f(t) = t \). The neuron number \( c \) is also an important factor as we control the decreasing rate. The rate should be relatively slow as \( c \) is large and relatively fast as \( c \) is small. Therefore, we use \( f(t)/c \) as an overall consideration. We may also choose any other positive strictly increasing functions for \( f(t) \) to control the excitation states.

Using the normalization term \( \max_{1 \leq i \leq c} \{\mu_i(x_j)\} \) in equation (13) is important. The membership functions \( \mu_i(x_j) \) resulting from FCM have the restriction with \( \sum_{i=1}^{c} \mu_i(x_j) = 1 \). Suppose the network is trained with a large codebook and bad initial weights, the \( \mu_i(x_j) \) for each \( i \) will then be very close to a small positive number \( 1/c \) when \( c \) is large. Our normalization technique can ensure that the excited state of the closest neuron is one and the learning will have relative significance at each step. In the Gibbs distribution used in SCS, the excited states for each neuron will be very small when \( c \) is large and the temperature parameter \( \beta(t) \) is small. The learning is not significant if we take a too large or too small initial temperature \( T_0 \). Therefore, Yair et al.\textsuperscript{11} used the re-initialization technique to speed up the training. Although this is somewhat similar to simulated annealing, it will make the network unstable as \( t \) is a perfect square. These phenomena will be shown in the next section. The competitive SCS and FS-CLA schemes are shown in Figure 2. The limiting competitive learning behavior of SCS and FS-CLA will become WTA. As \( t \) increases in FS-CLA, the excitation state of the neuron that is the closest to the input vector will maintain \( h_{ij}(t) = 1 \) and other neurons will be inhibited. As \( t \) increases in SCS, the excitation state of the neuron that is the closest to the input vector will be strengthened to 1 and other neurons will be inhibited.
Suppose that we consider an $\alpha$-cut of the membership function that

$$
\mu^\alpha_i(x_j) = \begin{cases} 
1, & \text{if } \mu_i(x_j) \geq \alpha, \\
0, & \text{if } \mu_i(x_j) < \alpha,
\end{cases}
$$

the approximation will reduce to a simple approximation, which is similar to the approximation function (4) of Kohonen’s SOM. The neighborhood of the winner neuron is determined by the $\alpha$-cut. The approximate excitations are also shown in the third row in Figure 1.

After considering the neural excitation function $h_{ij}(t)$, we shall discuss the adaptation gain for each neuron. In WTA, the learning rates $\alpha_i(t)$ for neurons at time $t$ are all equal with $\alpha_i(t) = \alpha(t) = \alpha/t$. Note that $\alpha_i(t-1) = \alpha/(t-1)$. We may rewrite $\alpha_i(t)$ as follows:

$$
\alpha_i(t) = \frac{\alpha}{t} = \frac{\alpha}{\left[\frac{\alpha}{\alpha/(t-1)}\right] + 1} = \frac{\alpha}{\alpha/(t-1)} + 1.
$$

The decreasing gaps are controlled by adding one to the denominator and are all equal regardless if the neuron wins or not. In the fuzzy-soft competitive network, the decreasing gaps $\alpha_{ij}(t)$ will be controlled by the excited states as

$$
\alpha_{ij}(t) = \frac{\alpha_0}{\alpha_{ij}(t-1)} + h_{ij}(t).
$$

In each time $t$, neurons with small excited states will have small decreasing gaps and will then maintain their competitive learning potential. The decreasing gaps are dependent on their individual excited levels. This update schedule efficiently builds...
a fair competitive learning environment and can be also written as
\[
\alpha_{ij}(t) = \frac{\alpha_0}{\left[ \frac{\alpha_0}{\alpha_{i}(0)} \right] + \sum_{l=1}^{t} \sum_{k=1}^{l} h_{ik}(l)}.
\] (19)

Since \(\alpha_0\) and \(\alpha_{i}(0)\) are positive constants, the learning rate (19) in the fuzzy-soft network shall decrease monotonically to zero as the time \(t\) increases. The decreasing gap is controlled by adding the new excited state \(h_{ij}(t)\) to the previous total excited values to the denominator. Note that \(\alpha_0\) can be used to control the decreasing rate of learning. A large value for \(\alpha_0\) means a small decrease in the learning rate. Typically, \(\alpha_0 = \alpha_{i}(0) = 1\). The learning rate update schedule in SCS can also be rewritten as
\[
\alpha_{ij}(t) = \frac{1}{n_i(t)} = \frac{1}{n_i(t-1) + h_{ij}(t)} = \frac{1}{\alpha_{ij}(t-1)} + h_{ij}(t),
\] (20)

which is equivalent to the update schedule (19) in FS-CLA.

In FS-CLA, when the data point \(x_j\) is input, all of the codevectors are simultaneously updated towards \(x_j\) and the learning size for each of them is \((\alpha_{ij}(t)h_{ij}(t))\) with equations (13) and (18). If the learning rate \(\alpha_{ij}(t-1)\) are all equal to a fixed constant \(b\) at the time \((t-1)\), the learning size can then be described as
\[
\alpha_{ij}(t)h_{ij}(t) = \frac{\alpha_0h_{ij}(t)}{\left[ \frac{\alpha_0}{b} \right] + h_{ij}(t)} = \frac{\alpha_0}{\left[ \frac{\alpha_{ij}(t)}{b} \right]} + 1,
\] (21)

which is a monotone increasing function of \(h_{ij}(t)\). Thus, FS-CLA can ensure that the learning size of the neuron with a large excited state will be relatively large under the same starting learning conditions. We create the FS-CLA algorithm as follows:

**FS-CLA algorithm**

1. Initialize \(Z_1(0), \ldots, Z_c(0)\);
2. For \(t = 1, \ldots, T\);
   - For \(j = 1, \ldots, n\);
     - Estimate \(h_{ij}(t)\) using (12) and (13);
     - Update \(\alpha_{ij}(t)\) using (18);
     - Update \(Z_i(t)\) using (10);
   - Next \(j\);
3. Next \(t\).

In Bezdek and Pal\(^{23}\) under the conditions that all classes are equally likely and all classes have common covariance \(\sigma^2\), they interpreted \(h_{ij}(t)\) in SCS as the posterior probability that \(x_j\) is from class \(i\) where the class \(i\) is modelled as an \(s\)-variate normal distribution. Alpaydin\(^{24}\) pointed out that SCS learning can be seen as an approximation to the well-known EM algorithm. However, the first problem in using SCS is that \(T_0\) is usually determined by experiments and the SCS results are very sensitive to the parameter initials of \(\beta(t)\). Bezdek and Pal\(^{23}\)
provided numerical illustrations about the difficulty in choosing good parameter initials for SCS. The second problem is that the re-initialization technique will make the training process unstable as $\beta(t)$ is large. The reasons are that when $\beta(t)$ is large, the soft competition will become WTA. In this case, the neuron that satisfies the nearest neighbor condition will be the excited state. It will then have a large learning size when the iterative count $t$ is a large perfect square number. Regardless if $x_j$ is far away from any codevector, the winner codevector (neuron) will be updated to be close to $x_j$ when the update time $t$ is a large perfect square number. Therefore, the re-initialization process should retrain the network as $t$ is a large perfect square number so that many convergence conditions proven by the author will be violated when using this re-initialization technique. In fact, FS-CLA with the fuzzy-soft competition by incorporating fuzzy membership functions as kernel type neighborhood interaction functions can avoid the difficulty in choosing good parameter initials and unstable re-initialization in SCS.

3. NUMERICAL COMPARISONS

The objective of vector quantization is to find a set of reference vectors (or cluster centers) that can represent the data set as well as possible. WTA, SCS, and FS-CLA can be thought of as a vector quantization design to find a set of good reference vectors. Thus, we will compare the performance of these three algorithms at each time step $t$. The performance is measured using the mean squared error (MSE) which is calculated by

$$MSE(t) = \frac{1}{c} \sum_{i=1}^{c} \| Z_i(t) - a_i \|^2 / c,$$

(22)

where $a_i$ is the true population mean of the cluster $i$. We use two numerical examples in the comparison of the proposed FS-CLA to WTA and SCS.

**Example 1.** We give a two-cluster data set shown in Figure 3(a). The data set is randomly generated from the bivariate normal distributions with $a_1 = (0, 0)$ and $a_2 = (5, 0)$, respectively. We implement WTA, SCS and FS-CLA for the data set with the initial values of $Z_1^{(0)} = a_1$ and $Z_2^{(0)} = a_2$. The MSE values of $t = 1, \ldots, 25$ for each algorithm are shown in Table I. Two cases of parameter specifications with $\alpha = 0.5$ and 0.1 are used in WTA. The initial temperatures in SCS are $T_0 = 5, 10,$ and 15. FS-CLA is a parameter-free scheme. Because the learning is not significant in SCS when $T_0$ is small and the true population means are used as the initial values, SCS produces the best MSE values when $T_0 = 5$ in $t = 1, \ldots, 10$. When the learning becomes stable, say $t > 10$, FS-CLA gives good vector quantization performance as shown in Table I. Since SCS uses the re-initialization technique, we mark the MSE values in Table I as $t$ is a perfect square. When $T_0 = 10$ and 15 in SCS, the re-initialization may reduce the MSE values. However, re-initialization does not always guarantee this good property. Figure 3(b) shows the MSE behavior for each
Table I. The MSE values of the algorithms.

<table>
<thead>
<tr>
<th>Iterative time $t$ increasing</th>
</tr>
</thead>
<tbody>
<tr>
<td>excited level</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

FS-CLA

MSE values for $t = 36$ to 100. Notations 1 and 2 represent the MSE values of the WTA with $\alpha = 0.5$ and 0.1, respectively. Notations 3, 4, and 5 represent the MSE values of the SCS with $T_0 = 15$, 10 and 5, respectively. The MSEs of SCS with different $T_0$ converge to the same values shown in Figure 3(a). The re-initialization for large $t$ may increase MSE so that the MSE curves are not smooth. Thus, re-initialization makes the learning unstable. When $t$ is large, FS-CLA produces smaller MSE values than WTA and SCS.

Example 2. In this example, we use three groups of one-dimensional normal mixture distributions. The data set used to compare the performance of the algorithms
The histograms for these three data sets are shown in Figures 4(a), 4(b) and 4(c), respectively. In the first mixture model, the MSE values implemented by WTA and FS-CLA are shown in Figure 5(a). The MSE values implemented by SCS and FS-CLA are shown in Figure 5(b). FS-CLA gives good MSE values when $t > 15$. In the second mixture model, FS-CLA produces good performance as well as in the third mixture model shown in Figures 6 and 7, respectively. In general, FS-CLA produces a good result when the learning is stable. Although re-initialization makes SCS unstable, the MSE values of SCS seem to converge to the same values. This is a good property that makes SCS insensitive to the initial values, including the initial vectors and temperatures. Note that, SCS may produce better results with other initial temperatures. However, this good initial temperature should be determined experimentally making this good initial temperature difficult to find in most simulations. In the next section, we apply FS-CLA in the MRI segmentation problem.

4. APPLICATIONS TO MRI SEGMENTATION

A real ophthalmology case recommended by a neurologist with MR image data is examined in this section. A male patient, age 79, was admitted to the hospital for a sudden onset of double vision that lasted for ten days. Over the past several years, he suffered from hypertension, ischemic heart failure, and chronic atrial fibrillation. In October 1999, he received cardioversion. A neurological examination upon admission revealed only bilateral internuclear ophthalmoplegia, indicating lesions in the bilateral medial longitudinal fasciculus in the dorsal pons tegmentum. There was no weakness in his extremities. A standard spin-echo (SE) sequence with $TR = 2650$ ms with $TE = 82.9$ produced MR Images with a field of view (FOV) =

![Histograms of three one-dimensional normal mixture data sets.](image)
The MSE values of the algorithms of the data set shown in Fig. 4(a).

The MSE values of the algorithms of the data set shown in Fig. 4(b).

The MSE values of the algorithms of the data set shown in Fig. 4(c).

24 * 18 mm, Slice thickness = 5.0 mm without gap and a 256 * 256 pixel matrix, that is, contiguous slices were acquitted. This patient’s brain MRI revealed lacunar infarctions in the bilateral lentiform nuclei, thalami and corona radiate, but not in the
dorsal pons. The diffusion-weighted images did not show recent or acute infarctions. Because the clinical findings were consistent with dorsal pontine lesions, he was discharged two days later under a bilateral internuclear-opthalmoplegia diagnosis probably due to the bilateral dorsal pontine infarct. His double vision disappeared in about a month. Thus, the proposed FS-CLA and the other two competitive learning algorithms WTA and SCS are used in segmenting the patient’s MRI data. The segmentation results are introduced as a support for clinical diagnosis.

The above actual case was implemented to compare the performance of FS-CLA to WTA and SCS. The MR image data set is from the 79-year-old male patient. He suffered a sudden onset of double vision for 10 days. Neurological examination revealed only bilateral internuclear opthalmoplegia, indicating lesions in the dorsal pons. However, the MR image shown in Figure 8 did not reveal the lacunar infarctions in the pons area.

MRI may be helpful in the clinical diagnosis of brain lesions. However, it has limitations in detecting lesions and tumors as small as 0.01 mm³ and digital noise from sensor resolution. The competition learning segmentation techniques are useful in detecting small lesions. They are more helpful in outlying the edge between tissues from digital image noise. To enhance the lesions, under the recommendation of specialists, gray scale window segmentation around the dorsal pons area is selected (see Figure 8) from the original MR image. The window image is grouped into five tissue classes: edema, gray matter, nerve tissue, white matter and cerebrospinal fluid. The gray scale histogram chooses from the 168 * 168 pixels window segmentation, providing a set of starting points. We analyzed the accuracy and the computational efficiency of the lesions detected from the window MRI data set against the WTA, SCS, and FS-CLA algorithms.

The computational efficiency is calculated by the number of iterations (NI). All algorithms were processed with the same specifying initial vectors \(Z(0)\), stopping criterion \(\epsilon = 0.01\) and maximum NI = 50. The histogram gray scale ranges from
0 to 255 starting from dark to bright. In this MRI data set, the histogram suggested that the lesion peak is around 140, which is in the middle of five initials $Z(0)$. Therefore, we simulated the test by changing the third initials from 130 to 160 to examine the accuracy and computation efficiency. Furthermore, we tried to compare the learning ability for WTA, SCS, and FS-CLA by pulling the initial vector away from the lesion peak. According to the histogram, a good starting initial vector set is $\{20, 80, 140, 160, 180\}$. We then set the initial vector set as $\{20, 80, 160, 220, 240\}$, which pulls the 4th and 5th initial vectors toward the bright side of the gray scale to test the performance of these competitive learning algorithms. The pictures from Figure 8 were processed at 168*168 pixels and clustered into five tissue classes. From the red circle on the two dimensional MR images, two node shaped detailed lesions displayed at the dorsal pons of the brain was pointed out by the neurologist and radiologist.

The MR image data set shown in Figure 8 was simulated using FS-CLA, SCS,
Figure 9. FS-CLA segmentation results for selected window MR Image. Refine Accuracy from FS-CLA. Good Accuracy from FS-CLA. Bad Accuracy from FS-CLA.
Figure 10. SCS segmentation results for selected window MR Image. Refine Accuracy from SCS. Good Accuracy from SCS. Bad Accuracy from SCS.
We compared these three competitive learning algorithms by moving the third initial vector toward both sides of the gray scale. Furthermore, we tried to compare the learning ability of all three algorithms by simulating a set of initial vectors that were based on the histogram peaks. The results are listed in Table II. The data varies in NI and accuracy when the gap between the 3rd initial vector and 4th initial vector are closer. However, unequal quantities of MRI data set gray scale pixels to these three algorithms may affect the performance and the accuracy of R, G, and B where R denotes the refined accuracy; G denotes good accuracy and B denotes bad accuracy. Therefore, the set of starting initial vectors suggested from the histogram peak is more essential.

With a set of good initial vectors referenced from the MR image data set histogram, FS-CLA produced the most refined and accurate image (see Table II and
Figure 9). In the red circle, it shows two detailed lesion nodes at the pons area. The FS-CLA converged faster than SCS and WTA. According to our experiments, we found that SCS is very sensitive to the parameter changes. Because the proposed FS-CLA is a parameter-free scheme, it does not have the parameter selection problem. The results point out the initial temperature $T_0 = 4.86$ in SCS produces the best result with refined accuracy (see Table II). We simulated the re-initialization process of SCS by changing the $T_0$ from 3 to 9 (see Table III), when $T_0 = 5$ and $T_0 = 7$, SCS provided refined accurate segmentation images (see Table II and Figure 10). Compared to FS-CLA, the SCS algorithm requires more steps to learn. WTA will converge into a winner-take-all state using the same set of initials (see Table II and Figure 11). WTA provides a segmentation image that only enhances one side of a lesion with large $N_L = 48 \times 28224$ trails. According to our experimental results, FS-CLA provides refined accurate segmentation images that enhance the two lesions at the pons area of the brain. The iteration number to converge is also faster than the other two algorithms. Overall, the physicians recommended using FS-CLA as an aid to brain lesion medical diagnosis.

5. CONCLUSIONS

In this paper, a sequential type of the (batch-type) FS-LVQ is created; that is, the FS-CLA. The differences between the three types of competitive learning algorithms with WTA, SCS and FS-CLA were analyzed. Since the concept of fuzzy-soft competitive learning is used in FS-CLA where the fuzzy-soft learning can build a more fair and reasonable competitive neural environment, it provides a more stable competitive learning network. We also found that the FS-CLA is a parameter-free scheme and can solve the problem of using a re-initialization process in SCS. Several numerical comparisons for WTA, SCS and FS-CLA are made. The numerical examples showed that the FS-CLA has better performance in
MSE values than WTA and SCS. Finally, these three competitive learning algorithms were applied in segmenting an actual ophthalmological MRI data case. MRI has limitations due to the size of the lesion area. This may be caused by digital noise in the partial volume effects originality from the low sensor resolution. As a result, all three algorithms, WTA, SCS and FS-CLA, can detect lesion areas with the suggested initial vectors from the MR image data set histogram. However, in using the fuzzy-soft competition learning methodology, the large learning size with a large excited state will remain as a relatively large learning condition. Thus, the learning pixel quantities of the lesions may affect the NI and accuracy of the lesion outline. As it is more robust, FS-CLA has the ability to tolerate the above situations both in the number of iterations and the accuracy of the lesion outline. According to the comparisons, FS-CLA has better accuracy and computational efficiency than the other two algorithms. Therefore, the created FS-CLA is recommended for use in MRI segmentation as an aid for support diagnoses.

Acknowledgment

This work was supported in part by the National Science Council, Taiwan, under Miin-Shen Yang’s Grant: NSC 97-2118-M-033-002-MY2.

References