CHAPTER 4

Pattern Classification using Generalized Recurrent Exponential Fuzzy Associative Memories

Marcos Eduardo Valle and Aline Cristina de Souza

Generalized recurrent exponential fuzzy associative memories (GRE-FAMs) are biologically inspired models designed for the storage and recall of fuzzy sets. They can be viewed as a recurrent multilayer neural network that employs a fuzzy similarity measure in its first hidden layer. In this chapter, we provide theoretical results concerning the storage capacity and noise tolerance of a single-step GRE-FAM. Furthermore, we describe how a GRE-FAM model can be applied for pattern classification. Computational experiments show that the accuracy of certain GRE-FAM classifiers is competitive with some well-known classifiers from the literature.

Keywords - Fuzzy associative memories, pattern classification, fuzzy systems.

Marcos Eduardo Valle and Aline Cristina de Souza
Department of Applied Mathematics
University of Campinas
Campinas, Brazil, SP 13083-859
e-mail: valle@ime.unicamp.br, s.alinedesouza@gmail.com

Editors: G.A. Papakostas, A.G. Hatzimichailidis, V.G. Kaburlasos, *Handbook of Fuzzy Sets Comparison - Theory, Algorithms and Applications*
DOI: 10.15579/gcsr.vol6.ch4, GCSR Vol. 6, pp. 79-102, 2016

4.1 Introduction

Associative memories (AMs) are mathematical models inspired by the human brain ability to store and recall information by means of associations [3, 15, 17, 23]. They are designed for the storage of a finite set of items called fundamental memories. Furthermore, an AM is expected to retrieve a stored information even upon the presentation of an incomplete or noisy item. Applications of AM models include pattern classification and recognition [13, 39, 49, 50], optimization [19], computer vision and image retrieval [5, 7, 37], prediction [38, 41, 42], and control [24, 25].

An AM model designed for the storage and recall of fuzzy sets is called *fuzzy associative memory* (FAM) [25, 42]. Precisely, a FAM is designed for the storage of associations $(A^1, B^1), (A^2, B^2), \ldots, (A^p, B^p)$, where A^ξ and B^ξ are fuzzy sets for all $\xi = 1, \ldots, p$. Afterward, the FAM model is expected to retrieve a certain B^ξ in response to the presentation of a partial or noisy version $\tilde{A}^\xi$ of A^ξ.

Research on FAM models originated in the early 1990s by Kosko [24, 25]. Generally speaking, Kosko's FAM stores a pair (A^ξ, B^ξ) in a matrix using either the correlation-minimum or the correlation-product encoding scheme. Despite successful applications of Kosko's FAMs to problems such as backing up a truck and trailer [24], target tracking [25], and voice cell control in ATM networks [32], they suffer from an extremely low storage capacity due to crosstalk between the stored items [9, 42].

To overcome the limitations of Kosko's FAMs, many researchers have developed FAM models with high storage capacity and improved noise tolerance. For instance, Chung and Lee generalized the matrix-based FAM models of Kosko by considering general maximum of triangular norm (max-T) compositions [9]. A recording recipe, called *implicative fuzzy learning*, has been devised for the storage of as many items as desired in a matrix-based FAM model [21, 29, 40]. A matrix-based FAM model synthesized using the implicative fuzzy learning is called *implicative fuzzy associative memories* (IFAM) [40]. Recently, Perfilieva and Vajgl characterized the noise tolerance of IFAMs using the notion of fuzzy preorder [33, 34]. Also, Bui et al. improved the noise tolerance of a certain IFAM using concepts from mathematical morphology [7].

Besides the active research on matrix-based FAMs, there is an increasing interest on non-distributive FAM models such as the Θ-FAMs introduced recently by Esmi et al. [12]. In general terms, a Θ-FAM yields the union $\cup_{\gamma \in \Gamma} B^\gamma$, where $\Gamma \subseteq \{1, \ldots, p\}$ is the set of the indexes that maximize a certain function of the input fuzzy set. For instance, an SM-FAM is obtained by considering the indexes that maximize the similarity measure between the fundamental memory A^ξ and the input X.

The *generalized recurrent exponential fuzzy associative memories* (GRE-FAMs), introduced recently by Souza et al., also belong to the class of non-distributive models [10]. Briefly, the GRE-FAM models have been derived from our previous recurrent exponential fuzzy associative memory by adding a hidden layer geared to avoid crosstalk between the stored items [44]. Like the famous Hopfield network [16, 18], GRE-FAMs only implement autoassociative memories, that is, they are designed for the storage and recall of fuzzy sets $A^1, \ldots, A^p$. Furthermore, they are recurrent models, that is, they produce a sequence of fuzzy sets $X_0, X_1, \ldots$ which presumably converges to the desired output. Indeed, the output of a single-step GRE-FAM can be made as close as desired to a certain combination of the fuzzy sets $A^1, \ldots, A^p$ which are the most similar to the input. Computational experiments revealed that the GRE-FAM models can be effectively used for reconstruction of noisy gray-scale images [10]. In this chapter, we successfully apply GRE-FAMs to pattern classification tasks.

The chapter is organized as follows. The next section reviews some basic concepts on fuzzy sets and similarity measures. Section 4.3 presents the GRE-FAM models. This section also contains some theoretical results concerning the storage capacity and retrieval capability of the GRE-FAMs. The application of the GRE-FAMs for pattern classification task is described in Section 4.4. Section 4.5 is concerned with the application of the GRE-FAM classifier to several benchmark classification problems. The chapter finishes with some concluding remarks in Section 4.6.

4.2 Fuzzy Sets and Similarity Measures

Let us begin by recalling some well-established basic concepts that will be used throughout the text. First of all, a fuzzy set A on a universe of discourse U is identified by its membership function $A : U \rightarrow [0,1]$. Hence, $A(u)$ denotes the degree to which the element $u \in U$ belongs to the fuzzy set A. The family of all fuzzy subsets of U is denoted by $\mathcal{F}(U)$. Similarly, $\mathcal{P}(U)$ represents the power set of U.

In this chapter, a fuzzy set A on a finite universe of discourse $U = \{u_1, u_2, \ldots, u_n\}$ is identified by a column vector $[a_1, a_2, \ldots, a_n]^T$, where $a_j = A(u_j)$ for all $j = 1, \ldots, n$. Hence, $\mathcal{F}(\{u_1, u_2, \ldots, u_n\})$ corresponds to the hypercube $[0,1]^n$.

As usual, we say that $A \in \mathcal{F}(U)$ is a subset of $B \in \mathcal{F}(U)$, and write $A \subseteq B$, if $A(u) \leq B(u)$ for all $u \in U$. Also, $\bar{A}$ denotes the standard complement of a fuzzy set A, that is, $\bar{A}(u) = 1 - A(u)$ for all $u \in U$.

A *similarity measure*, also known as *equality index* or *fuzzy equivalence*, is a function that associates to each pair of fuzzy sets a real number in the unity interval $[0,1]$, representing the degree to which those fuzzy sets are equal [11, 22]. Applications of similarity measures include fuzzy neural networks [4, 12], fuzzy clustering [47], linguistic approximation [51], rule base simplification [36], fuzzy reasoning [43], and image processing [6, 8].

There is a vast literature on similarity measure. Also, this concept is interpreted in different ways depending on the context. In the following, we consider the normalized version of the axiomatic definition provided by Xuecheng [46]:

Definition 1 (Similarity Measure)**.** A similarity measure is a function $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0,1]$ which satisfies the following properties, for any fuzzy sets $A, B, C, D \in \mathcal{F}(U)$:

1. $\mathcal{S}(A,B) = S(B,A)$.
2. $\mathcal{S}(A,A) = 1$.
3. If $A \subseteq B \subseteq C \subseteq D$, then $\mathcal{S}(A,D) \leq \mathcal{S}(B,C)$.
4. $\mathcal{S}(A,\bar{A}) = 0$, for all crisp set $A \in \mathcal{P}(U)$.

We say that $\mathcal{S}$ is a *strong similarity measure* if $\mathcal{S}(A,B) = 1$ implies $A = B$.

Definition 2. Let $A = [a_1, a_2, \ldots, a_n]^T$ and $B = [b_1, b_2, \ldots, b_n]^T$ be two fuzzy sets on $U = \{u_1, \cdots, u_n\}$. The following presents two strong similarity measures [6, 11, 30, 51].

1. *Gregson similarity measure*:

$$\mathcal{S}_G(A, B) = \frac{\sum_{j=1}^{n}(a_j \wedge b_j)}{\sum_{j=1}^{n}(a_j \vee b_j)}, \tag{4.1}$$

   where the symbols "$\wedge$" and "$\vee$" denote respectively the minimum and maximum operations.

2. *Complement of the Relative Hamming Distance*:

$$\mathcal{S}_H(A, B) = 1 - \frac{1}{n}\sum_{j=1}^{n}|a_j - b_j|. \tag{4.2}$$

Example 1. Let $A = [0.4, 0.5]^T$ and $B = [0.6, 0.4]^T$ be fuzzy sets on $\mathcal{F}(\{u_1, u_2\})$. In this case, we obtain the similarity degrees

$$\mathcal{S}_G(A, B) = 0.73 \quad \text{and} \quad \mathcal{S}_H(A, B) = 0.85.$$

Moreover, Fig.(4.1) shows the fuzzy sets A and B as well as the geometric regions corresponding to the families

$$F_G = \{X \in \mathcal{F}(\{u_1, u_2\}) : \mathcal{S}_G(A, X) \geq 0.8\}, \tag{4.3}$$

and

$$F_H = \{X \in \mathcal{F}(\{u_1, u_2\}) : \mathcal{S}_H(A, X) \geq 0.8\}. \tag{4.4}$$

Note that F_H is a square while F_G is a trapezoid. Also, $F_G \subseteq F_H$. In particular, we have $B \in F_H$ but $B \notin F_G$.

Remark 1. Gregson similarity measure is an example of a similarity measure based on intersection, union, and cardinality. Precisely, we have

$$\mathcal{S}_G(A, B) = \frac{Card(A \cap B)}{Card(A \cup B)}, \tag{4.5}$$

where "$\cap$" and "$\cup$" denote the classical intersection and union of fuzzy sets and the cardinality of a fuzzy set is the sum of its membership degrees.

The similarity measure $\mathcal{S}_H$ is a particular case of the complement of a relative distance $\mathcal{S}_p$ given by

$$\mathcal{S}_p(A, B) = 1 - \frac{d_p(A, B)}{d_p(\emptyset, U)} \tag{4.6}$$

where d_p denotes the L_p distance of order $p \geq 1$. The denominator $d_p(\emptyset, U)$, which corresponds to the largest distance between two fuzzy sets, ensures $S_p(A, B) \in [0, 1]$.

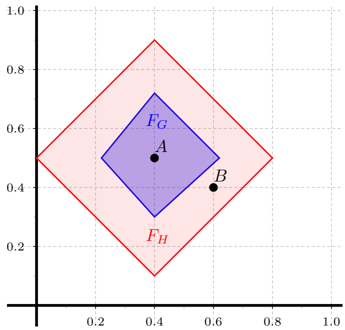

Figure 4.1: Geometry of the families F_G (blue) and F_H (red) given respectively by Eq.(4.3) and Eq.(4.4).

4.3 Generalized Recurrent Exponential Fuzzy Associative Memories

Generalized recurrent exponential fuzzy associative memories (GRE-FAMs), introduced recently by Souza et al. [10], are high-capacity associative memory models designed for the storage of a finite family $\mathcal{A} = \{A^1, A^2, \cdots, A^p\} \subseteq \mathcal{F}(U)$ of fuzzy sets. We shall refer to $A^\xi \in \mathcal{A}$ as a fundamental memory. A GRE-FAM model is used to retrieve a memorized fuzzy set under presentation of an input fuzzy set $X_0 \in \mathcal{F}(U)$. Formally, a GRE-FAM is defined as follows:

Definition 3 (GRE-FAM)**.** Consider a finite family of fundamental memories $\mathcal{A} = \{A^1, A^2, \cdots, A^p\} \subseteq \mathcal{F}(U)$, a real number $\alpha > 0$, and a similarity measure $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \to [0,1]$. Furthermore, let $G = (g_{\nu\mu})$ be a real-valued matrix of size $p \times p$. Given an initial fuzzy set $X_0 \in \mathcal{F}(U)$, a GRE-FAM defines a sequence $\{X_t\}_{t\geq 0}$ of fuzzy sets as follows for all $u \in U$ and $t = 0, 1, \ldots$:

$$X_{t+1}(u) = \varphi\left(\frac{\sum_{\xi=1}^{p}\sum_{\mu=1}^{p} A^\xi(u)\, g_{\xi\mu} e^{\alpha\mathcal{S}(A^\mu, X_t)}}{\sum_{\eta=1}^{p}\sum_{\mu=1}^{p} g_{\eta\mu} e^{\alpha\mathcal{S}(A^\mu, X_t)}}\right), \tag{4.7}$$

where $\varphi : \mathbb{R} \to [0,1]$ is the piece-wise linear function given by

$$\varphi(x) = 0 \vee (1 \wedge x). \tag{4.8}$$

Equivalently, the fuzzy set X_{t+1} produced by a GRE-FAM at iteration $t \geq 0$ is defined by the following equations:

$$u_{\mu t} = e^{\alpha\mathcal{S}(A^\mu, X_t)}, \quad \forall \mu = 1, \ldots, p, \tag{4.9}$$

$$v_{\xi t} = \sum_{\mu=1}^{p} g_{\xi\mu} u_{\mu t}, \quad \forall \xi = 1, \ldots, p, \tag{4.10}$$

$$w_{\xi t} = \frac{1}{\sum_{\eta=1}^{p} v_{\eta t}} v_{\xi t}, \quad \forall \xi = 1, \ldots, p, \tag{4.11}$$

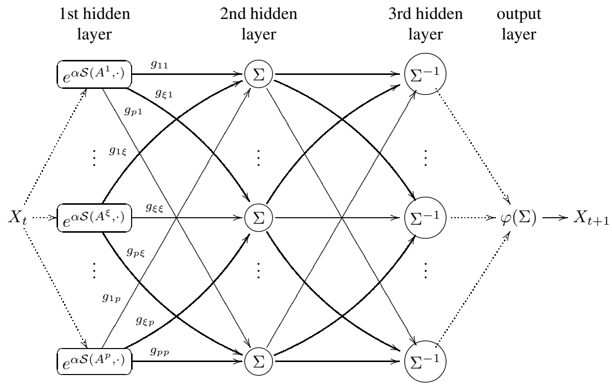

Figure 4.2: Block diagram of a GRE-FAM.

$$X_{t+1}(u) = \varphi\left(\sum_{\xi=1}^{p} w_{\xi t} A^{\xi}(u)\right), \quad \forall u \in U. \tag{4.12}$$

Therefore, a GRE-FAM can be interpreted as the fully connected four-layer[1] recurrent neural network shown in Fig.(4.2). In each node of the first hidden layer, an exponential function is applied to the similarity between the current state X_t and the fundamental memory A^ξ, for $\xi = 1, \ldots, p$. The second hidden layer consists of linear neurons. The nodes in the third layer simply normalize the output of the previous layer. Finally, the output layer computes a weighted sum of the fundamental memories $A^1, \ldots, A^p$ followed by a piece-wise linear function φ which ensures that $X_{t+1}(u) \in [0,1]$ for all $u \in U$.

The following example illustrates the recall phase of a GRE-FAM based on the complement of the relative Hamming distance $\mathcal{S}_H$.

Example 2. Consider the fundamental memory set $\mathcal{A} = \{A^1, A^2, \ldots, A^8\} \in [0,1]^2$ formed by the fuzzy sets

$$A^1 = \begin{bmatrix} 0.1 \\ 0.5 \end{bmatrix}, \quad A^2 = \begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix}, \quad A^3 = \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix}, \quad A^4 = \begin{bmatrix} 0.2 \\ 0.6 \end{bmatrix}, \tag{4.13}$$

$$A^5 = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix}, \quad A^6 = \begin{bmatrix} 0.4 \\ 0.8 \end{bmatrix}, \quad A^7 = \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}, \quad A^8 = \begin{bmatrix} 0.6 \\ 0.6 \end{bmatrix}, \tag{4.14}$$

the similarity measure $\mathcal{S}_H$, the parameter $\alpha = 1$, and let $G \in \mathbb{R}^{8\times 8}$ be the matrix

1 We would like to point out that the network topology presented in this work has an additional layer compared to the topology in [10]. As we shall see in Section 4.4, the additional layer plays an important role in pattern classification problems.

defined by

$$G = \begin{bmatrix} 3.87 & 0.00 & -3.68 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 8.30 & -3.68 & -3.82 & -0.68 & 0.19 & -4.73 & 4.47 \\ -3.68 & -3.68 & 10.86 & -3.68 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & -3.82 & -3.68 & 8.30 & 0.19 & -0.68 & 4.47 & -4.73 \\ 0.00 & -0.68 & 0.00 & 0.19 & 1.52 & -0.43 & -0.68 & 0.19 \\ 0.00 & 0.19 & 0.00 & -0.68 & -0.43 & 1.52 & 0.19 & -0.68 \\ 0.00 & -4.73 & 0.00 & 4.47 & -0.68 & 0.19 & 6.46 & -5.66 \\ 0.00 & 4.47 & 0.00 & -4.73 & 0.19 & -0.68 & -5.66 & 6.46 \end{bmatrix}. \quad (4.15)$$

Assume that the fuzzy set $X_0 = [0.4, 0.5]^T \in [0,1]^2$ is presented as input to the GRE-FAM model. Figure 4.3 shows the fundamental memories $A^1, \ldots, A^8$, marked by either a square or a triangle, as well as the input fuzzy set X_0, marked by a black filled circle. Note that, since $\mathcal{S}_H$ is close related to the L_1-distance (also known as the taxicab distance), A^3 is the fundamental memory most similar to X_0. Indeed, we have $\mathcal{S}_H(A^3, X_0) = 0.9$ while $\mathcal{S}_H(A^\mu, X_0) = 0.85$ for all $\mu \neq 3$. Now, the first hidden layer produces the vector $\mathbf{u}_0 = [u_{1,0}, \ldots, u_{8,0}]^T$ given by

$$\mathbf{u}_0 = \begin{bmatrix} 2.34 & 2.34 & 2.46 & 2.34 & 2.34 & 2.34 & 2.34 & 2.34 \end{bmatrix}^T. \quad (4.16)$$

From Eq.(4.10) and Eq.(4.11), the outputs of the second and third hidden layers can be arranged in the vectors

$$\mathbf{v}_0 = G\mathbf{u}_0 = \begin{bmatrix} 0 & -0.32 & 0.90 & -0.32 & 0.25 & 0.25 & 0.13 & 0.13 \end{bmatrix}^T, \quad (4.17)$$

$$\mathbf{w}_0 = \frac{\mathbf{v}_0}{\sum_{\eta=1}^{8} v_{\eta 0}} = \begin{bmatrix} 0 & -0.30 & 0.87 & -0.30 & 0.24 & 0.24 & 0.13 & 0.13 \end{bmatrix}^T, \quad (4.18)$$

respectively. Note that the sum of the weights $w_{\xi t}$ equals 1, that is, $\sum_{\xi=1}^{8} w_{\xi 0} = 1$. Moreover, although the components of $\mathbf{u}_0$ are all positive, the vector $\mathbf{v}_0 = G\mathbf{u}_0$ contains negative entries because of G. Consequently, some of the weights $w_{\xi t}$'s are also negative. From Eq.(4.12), we finish the first iteration with the fuzzy set $X_1 = [0.4, 0.5]^T$. Since $X_1 = X_0$, we have $X_t = [0.4, 0.5]$ for all $t \geq 0$. In other words, the fuzzy set $X_0 = [0.4, 0, 5]$ is a fixed point of the GRE-FAM based on $\mathcal{S}_H$.

Note that a GRE-FAM model is specified by the fundamental memory set $\mathcal{A}$, a similarity measure $\mathcal{S}$, the parameter $\alpha > 0$, and the real-valued matrix $G \in \mathbb{R}^{p \times p}$. We usually assume that $\mathcal{A}$ is given *a priori*. The similarity measure $\mathcal{S}$, as well as the parameter $\alpha > 0$, should be determined by the nature (or geometry) of the problem we deal with. Finally, the matrix G can be computed according to the following theorem, which have been stated without proof on [10]:

Theorem 1. *Given a family of fundamental memories $\mathcal{A} = \{A^1, A^2, \cdots, A^p\} \subseteq \mathcal{F}(U)$, a parameter $\alpha > 0$, and a similarity measure $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \to [0,1]$, define the matrix $C = (c_{\nu\mu}) \in \mathbb{R}^{p \times p}$ as follows:*

$$c_{v\mu} = e^{\alpha \mathcal{S}(A^v, A^\mu)}, \ \forall v, \mu = 1, ..., p. \quad (4.19)$$

If C is invertible, then any fundamental memory from the family $\mathcal{A}$ is a fixed point of the GRE-FAM with $G = C^{-1}$.

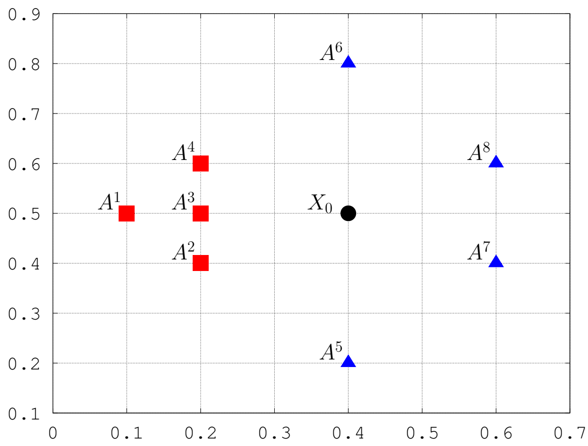

Figure 4.3: Fundamental memories $A^1, \ldots, A^8$ and the input fuzzy set X_0, marked by a black filled circle. In a classification task, the fuzzy sets $A^1, \ldots, A^4$, marked by red squares, belong to class 0 while the fuzzy sets $A^5, \ldots, A^8$, marked by blue triangle, belong to class 1.

Proof. Suppose that the matrix C given by Eq.(4.19) is invertible. If $G = C^{-1}$, then the following identity holds true for all indexes $\xi, \nu \in \{1, \ldots, p\}$:

$$\sum_{\mu=1}^{p} g_{\xi\mu} e^{\alpha \mathcal{S}(A^\mu, A^\nu)} = G(\xi, :)C(:, \nu) = \delta_{\xi\nu},$$

where $G(\xi, :)$ denotes the ξ-th row of the matrix G, $C(:, \nu)$ denotes the ν-th column of the matrix C, and $\delta_{\xi,\nu}$ is the Kronecker's delta defined by

$$\delta_{\xi\nu} = \begin{cases} 1, & \xi = \nu, \\ 0, & \xi \neq \nu. \end{cases}$$

Now, if a fundamental memory $X_0 = A^\nu$, with $\nu \in \{1, \ldots, p\}$, is presented to the GRE-FAM, then the following identities hold all $u \in U$:

$$\begin{aligned} X_1(u) &= \varphi\left(\frac{\sum_{\xi=1}^{p} \sum_{\mu=1}^{p} A^\xi(u) g_{\xi\mu} e^{\alpha \mathcal{S}(A^\mu, X_0)}}{\sum_{\eta=1}^{p} \sum_{\mu=1}^{p} g_{\eta\mu} e^{\alpha \mathcal{S}(A^\mu, X_0)}}\right) \\ &= \varphi\left(\frac{\sum_{\xi=1}^{p} A^\xi(u) \left[\sum_{\mu=1}^{p} g_{\xi\mu} e^{\alpha \mathcal{S}(A^\mu, A^\nu)}\right]}{\sum_{\eta=1}^{p} \left[\sum_{\mu=1}^{p} g_{\eta\mu} e^{\alpha \mathcal{S}(A^\mu, A^\nu)}\right]}\right) \\ &= \varphi\left(\frac{\sum_{\xi=1}^{p} A^\xi(u) \delta_{\xi\nu}}{\sum_{\eta=1}^{p} \delta_{\eta\nu}}\right) = \varphi\left(A^\nu(u)\right) = A^\nu(u). \end{aligned}$$

Therefore, A^ν is a fixed point of the GRE-FAM given by Eq.(4.7) with $G = C^{-1}$. $\square$

In words, Theorem 1 prescribes a matrix G such that any fundamental memory A^ξ is a fixed point of the GRE-FAM. Moreover, in this case, the linear neurons in the second hidden layer are used to mitigate the crosstalk between the fundamental memories. From now on, we shall assume that the matrix G is computed according to Theorem 1. In case C is not invertible, we define $G = C^\dagger$, where $C^\dagger$ denotes the pseudo-inverse of C.

The following example reveals that the matrix G given by Eq.(4.15) have been determined according to Theorem 1.

Example 3. Consider the fundamental memory set $\mathcal{A} = \{A^1, \ldots, A^8\}$ formed by the fuzzy sets given by Eq.(4.13) and Eq.(4.14). Also, let $\alpha = 1$ and consider the similarity measure $\mathcal{S}_H$ defined by Eq.(4.2). In this case, Eq.(4.19) yields the matrix

$$C = \begin{bmatrix} 2.72 & 2.46 & 2.59 & 2.46 & 2.01 & 2.01 & 2.01 & 2.01 \\ 2.46 & 2.72 & 2.59 & 2.46 & 2.23 & 2.01 & 2.23 & 2.01 \\ 2.59 & 2.59 & 2.72 & 2.59 & 2.12 & 2.12 & 2.12 & 2.12 \\ 2.46 & 2.46 & 2.59 & 2.72 & 2.01 & 2.23 & 2.01 & 2.23 \\ 2.01 & 2.23 & 2.12 & 2.01 & 2.72 & 2.01 & 2.23 & 2.01 \\ 2.01 & 2.01 & 2.12 & 2.23 & 2.01 & 2.72 & 2.01 & 2.23 \\ 2.01 & 2.23 & 2.12 & 2.01 & 2.23 & 2.01 & 2.72 & 2.46 \\ 2.01 & 2.01 & 2.12 & 2.23 & 2.01 & 2.23 & 2.46 & 2.72 \end{bmatrix} \tag{4.20}$$

For instance, the entry $c_{13} = e^{\alpha \mathcal{S}_H(A^1,A^3)} = e^{\alpha 0.95} = 2.59$ is the exponential of the similarity between A^1 and A^3. Note that the matrix C is symmetric with diagonal $e^\alpha = 2.72$. The inverse of C is the the matrix $G = C^{-1}$ given by Eq.(4.15), which is also symmetric.

Let us now estimate the number of operations required by a GRE-FAM based on Theorem 1. First, assume that an evaluation of the similarity measure $\mathcal{S}$ demands $\mathcal{O}(n)$ operations. Now, the matrix C given by Eq.(4.19) requires p^2 evaluations of the similarity measure $\mathcal{S}$. Also, $\mathcal{O}(p^3)$ operations are needed for the computation of $G = C^{-1}$. Therefore, the cost of designing a GRE-FAM is $\mathcal{O}(np^2 + p^3)$. Afterward, each step of a GRE-FAM model requires $\mathcal{O}(pn + p^2)$ operations.

Let us conclude this section by turning our attention to the output X_1 produced by a single-step GRE-FAM with G prescribed by Theorem 1. The following theorem shows that X_1 converges point-wise to an affine combination of the fundamental memories which have the highest similarity (in terms of $\mathcal{S}$) with the input X_0 as the parameter $\alpha > 0$ tends to infinity [10].

Theorem 2. *Consider a family of fundamental memories $\mathcal{A} = \{A^1, \cdots, A^p\} \subseteq \mathcal{F}(U)$ and let $\mathcal{S}$ denote a strong similarity measure. Suppose that the matrix C given by Eq.(4.19) is invertible for any $\alpha > 0$. Given an initial fuzzy set $X_0 \in \mathcal{F}(U)$, let $\Gamma \subseteq \{1, \ldots, p\}$ denote the set of the indexes of the fundamental memories which are the most similar to the input X_0 in terms of $\mathcal{S}$. Formally, we have*

$$\Gamma = \left\{\gamma : \mathcal{S}\left(A^\gamma, X_0\right) \geq \mathcal{S}\left(A^\xi, X_0\right), \forall \xi = 1, ..., p\right\}. \tag{4.21}$$

If $X_1 \in \mathcal{F}(U)$, given by Eq.(4.7) with $t = 0$, denotes the output of the single-step GRE-FAM, then

$$\lim_{\alpha\to\infty} X_1(u) = \frac{1}{Card(\Gamma)} \sum_{\gamma\in\Gamma} A^{\gamma}(u), \ \ \forall u \in U \tag{4.22}$$

Furthermore, the weight $w_{\xi 0}$ given by Eq.(4.11), which correspond to the output of the ξth neuron of the third hidden layer, satisfies the following equation for all $\xi = 1, \ldots, p$:

$$\lim_{\alpha\to\infty} w_{\xi 0} = \begin{cases} \frac{1}{Card(\Gamma)}, & \xi \in \Gamma, \\ 0, & \text{otherwise.} \end{cases} \tag{4.23}$$

Proof. Let $\sigma = \max_{\xi=1:p}\{\mathcal{S}(A^{\xi}, X_0)\}$ denote the maximum of the similarity between the input fuzzy set X_0 and the fundamental memories $A^1, \cdots, A^p$. The output of a single-step GRE-FAM is given by

$$X_1(u) = \varphi\left(\sum_{\xi=1}^{p} w_{\xi 0} A^{\xi}(u)\right), \tag{4.24}$$

where $w_{\xi 0}$, derived from Eq.(4.9)–Eq.(4.11), satisfies

$$w_{\xi 0} = \frac{\sum_{\mu=1}^{p} g_{\xi\mu} e^{\alpha \mathcal{S}(A^{\mu}, X_0)}}{\sum_{\eta=1}^{p}\sum_{\mu=1}^{p} g_{\eta\mu} e^{\alpha\mathcal{S}(A^{\mu}, X_0)}}, \ \ \forall \xi = 1, ..., p. \tag{4.25}$$

Multiplying both the numerator and the denominator of Eq.(4.25) by $e^{-\alpha\sigma}$ and breaking up the sums, we obtain:

$$w_{\xi 0} = \frac{\sum_{\gamma\in\Gamma} g_{\xi\gamma} + \sum_{\mu\notin\Gamma} g_{\xi\mu} e^{\alpha(\mathcal{S}(A^{\mu}, X_0)-\sigma)}}{\sum_{\eta=1}^{p}\sum_{\gamma\in\Gamma} g_{\eta\gamma} + \sum_{\eta=1}^{p}\sum_{\mu\notin\Gamma} g_{\eta\mu} e^{\alpha(\mathcal{S}(A^{\mu}, X_0)-\sigma)}} \tag{4.26}$$

Now, the matrix C given by Eq.(4.19) can be written as $C = e^{\alpha} D(\alpha)$, where the entries of $D(\alpha)$ are $d_{\nu\mu}(\alpha) = e^{\alpha(\mathcal{S}(A^{\nu}, A^{\mu})-1)}$. Moreover, $G = C^{-1} = e^{-\alpha} H(\alpha)$, where $H(\alpha) = D^{-1}(\alpha)$. Hence, by factoring $e^{-\alpha}$, we obtain from Eq.(4.26):

$$w_{\xi 0} = \frac{\sum_{\gamma\in\Gamma} h_{\xi\gamma}(\alpha) + \sum_{\mu\notin\Gamma} h_{\xi\mu}(\alpha) e^{\alpha(\mathcal{S}(A^{\mu}, X_0)-\sigma)}}{\sum_{\eta=1}^{p}\sum_{\gamma\in\Gamma} h_{\eta\gamma}(\alpha) + \sum_{\eta=1}^{p}\sum_{\mu\notin\Gamma} h_{\eta\mu}(\alpha) e^{\alpha(\mathcal{S}(A^{\mu}, X_0)-\sigma)}}. \tag{4.27}$$

Recalling that $\mathcal{S}(A^{\mu}, X_0) - \sigma < 0$ for all $\mu \notin \Gamma$, the second sum in both numerator and denominator tends to 0 as $\alpha \to \infty$. Moreover, since $\mathcal{S}$ is a strong similarity measure, $\lim_{\alpha\to\infty} h_{\nu\mu}(\alpha) = \delta_{\nu\mu}$, where $h_{\nu\mu}(\alpha)$ is the (ν, μ)-entry of $H(\alpha)$ and $\delta_{\nu\mu}$ denotes the Kronecker's delta. Hence,

$$\lim_{\alpha\to\infty} w_{\xi 0} = \frac{\sum_{\gamma\in\Gamma}\delta_{\xi\gamma}}{\sum_{\eta=1}^{p}\sum_{\gamma\in\Gamma}\delta_{\eta\gamma}} = \begin{cases} \frac{1}{Card(\Gamma)}, & \xi \in \Gamma, \\ 0, & \text{otherwise.} \end{cases} \tag{4.28}$$

Since φ is continuous and $\varphi(x) = x$ for all $x \in [0, 1]$, we conclude that

$$\lim_{\alpha\to\infty} X_1(u) = \lim_{\alpha\to\infty} \varphi\left(\sum_{\xi=1}^{p} w_{\xi 0} A^{\xi}(u)\right) = \varphi\left(\sum_{\xi=1}^{p} \left[\lim_{\alpha\to\infty} w_{\xi 0}\right] A^{\xi}(u)\right) \tag{4.29}$$

$$= \varphi\left(\sum_{\xi\in\Gamma} \frac{1}{Card(\Gamma)} A^{\xi}(u)\right) = \frac{1}{Card(\Gamma)} \sum_{\gamma\in\Gamma} A^{\gamma}(u). \tag{4.30}$$

for all $u \in U$. □

The following corollary shows that, under mild conditions, the output X_1 of a single-step GRE-FAM converges point-wise to the fundamental memory A^{γ} which is the most similar to the input X_0.

Corollary 1. *Consider a family of fundamental memories $\mathcal{A} = \{A^1, \cdots, A^p\} \subseteq \mathcal{F}(U)$ and let $\mathcal{S}$ denote a strong similarity measure. Suppose that the matrix C given by Eq.(4.19) is invertible for any $\alpha > 0$. Given an initial fuzzy set $X_0 \in \mathcal{F}(U)$, if the inequality $\mathcal{S}(A^{\gamma}, X_0) > \mathcal{S}(A^{\xi}, X_0)$ holds true for all $\xi \neq \gamma$, with $\xi, \gamma \in \{1, \ldots, p\}$, then the output of the single-step GRE-FAM given by Eq.(4.7) with $t = 0$ satisfies*

$$\lim_{\alpha\to\infty} X_1(u) = A^{\gamma}, \ \ \forall u \in U. \tag{4.31}$$

Moreover, the weight $w_{\xi 0}$ given by Eq.(4.11) converges to $\delta_{\xi\gamma}$ as α tends to infinity, that is,

$$\lim_{\alpha\to\infty} w_{\xi 0} = \begin{cases} 1, & \xi = \gamma, \\ 0, & \textit{otherwise}. \end{cases} \tag{4.32}$$

Example 4. Consider the fundamental memory set $\mathcal{A} = \{A^1, \ldots, A^8\}$ given by Eq.(4.13) and Eq.(4.14), the strong similarity measure $\mathcal{S}_H$, and let $X_0 = [0.4, 0.5]^T$ be the input fuzzy set. Note that the fundamental memory most similar to X_0 is A^3. Also, recall from Example 2 that the output of the single-step GRE-FAM based on $\mathcal{S}_H$ with $\alpha = 1$ is $X_1 = [0.4, 0.5]$. Therefore, for $\alpha = 1$, the Chebyshev distance between X_1 and A_3 is

$$\left\|A^3 - X_1\right\|_{\infty} = \max_{u\in U} \left|A^3(u) - X_0(u)\right| = 0.2. \tag{4.33}$$

Now, Fig.((4.4)) shows the distance $\|A^3 - X_1\|_{\infty}$ by the parameter α. Note that the Chebyshev distance decreases by increasing α. In fact, since $U = \{u_1, u_2\}$ is finite, Corollary 1 asserts that $\lim_{\alpha\to\infty} \|A^3 - X_1\|_{\infty} = 0$.

From Corollary 1, we conjecture that the basin of attraction of A^{γ} is the region

$$\mathcal{R}^{\gamma} = \left\{X \in \mathcal{F}(\mathcal{U}) : \mathcal{S}(A^{\gamma}, X) > \mathcal{S}(A^{\xi}, X), \forall \xi \neq \gamma\right\}, \tag{4.34}$$

when α is sufficiently large. In other words, $\mathcal{R}^{\gamma}$ corresponds to the family of fuzzy sets which are more similar to A^{γ} than any other fundamental memory A^{ξ}, $\xi \neq \gamma$.

At this point, recall that the pattern recalled by an autoassociative similarity measure FAM (SM-FAM) of Esmi et al. under presentation of $X \in \mathcal{F}(\mathcal{U})$ is the fuzzy set

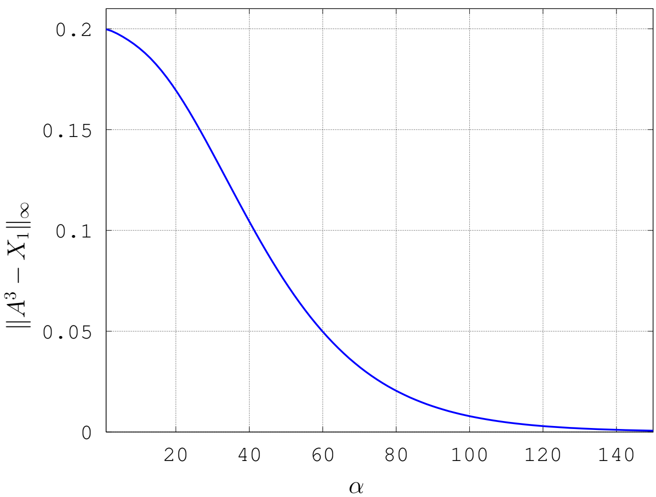

Figure 4.4: Chebyshev distance between the fundamental memory A^3 and the output X_1 of a single-step GRE-FAM model by the parameter α.

$$Y = \bigcup_{\gamma \in \Gamma} A^\gamma, \tag{4.35}$$

where Γ is the set of indexes given by Eq.(4.21) [12]. Hence, an autoassociative SM-FAM differs from a single-step GRE-FAM with a large parameter α in respect to how the fuzzy sets A^γ, $\gamma \in \Gamma$, are combined to produce the output. Furthermore, both FAM models probably coincide if Γ has an unique element and α is sufficiently large.

We would like to point out, however, that the parameter α cannot be made indefinitely large in many practical situations. For instance, due to overflow, we must roughly consider $\alpha \leq 700$ on a machine that supports IEEE floating point arithmetic. If the parameter α is not sufficiently large, the fuzzy set X_1 produced by a GRE-FAM may differ significantly from the output of an autoassociative SM-FAM. As we shall see in Section 4.5, this can be an advantage in some pattern classification problems.

4.4 Pattern Classification Using GRE-FAMs

In this section, we show how a single-step GRE-FAMs can be applied for pattern classification tasks. Formally, assume that we have a family of labeled fuzzy sets

$$\mathcal{A}_\mathcal{L} = \left\{ \left(A^\xi, \ell_\xi \right) : \xi = 1, ..., p \right\} \subseteq \mathcal{F}(U) \times \mathcal{L}, \tag{4.36}$$

where $\mathcal{L}$ is a finite set of labels and $\mathcal{A}^\xi$ are distinct non-empty fuzzy sets on U. We refer to $\mathcal{A}_\mathcal{L}$ as the training set. In a pattern classification task, the goal is to attribute a label $\ell \in \mathcal{L}$ to a given (unlabeled) input fuzzy set $X \in \mathcal{F}(U)$. Specifically, based on the training set $A_\mathcal{L}$, we synthesize a mapping $\mathcal{C} : \mathcal{F}(U) \to \mathcal{L}$, called classifier.

Example 5 (Fuzzy Nearest Neighbor Classifier)**.** A fuzzy version of the popular nearest neighbor classifier assigns to $X \in \mathcal{F}(U)$ the label of the most similar fuzzy set A^ξ, for $\xi = 1, \ldots, p$. In mathematical terms, a fuzzy nearest neighbor (fNN) classifier, denoted by $\mathcal{C}_{fNN}$, is defined as follows using a fuzzy similarity measure $\mathcal{S}$:

$$\mathcal{C}_{fNN}(X) = \ell_\eta, \text{ where } \eta \text{ satisfies } \mathcal{S}(A^\eta, X) > \mathcal{S}(A^\xi, X), \forall \xi = 1, ..., p. \quad (4.37)$$

Note that, if $\mathcal{S}$ is a strong similarity measure, the equation $\mathcal{C}_{fNN}(A^\xi) = \ell_\xi$ holds true for all $\xi = 1, \ldots, p$.

Example 6. Consider the family of labeled fuzzy sets $\mathcal{A}_\mathcal{L} = \{(A^\xi, \ell_\xi) : \xi = 1, \ldots, 8\}$ in which the fuzzy sets $A^1, A^2, \ldots, A^8$ are given by Eq.(4.13) and Eq.(4.14) and the class labels are

$$\ell_1 = \ell_2 = \ell_3 = \ell_4 = 0 \text{ and } \ell_5{=}\ell_6{=}\ell_7{=}\ell_8{=}1. \quad (4.38)$$

Also, let $X = X_0 = [0.4, 0.5]^T$ be the input fuzzy set. Recall that Fig.(4.3) shows $\mathcal{A}_\mathcal{L}$ as well as the input fuzzy set X_0. Here, the fuzzy sets labeled as 0 are marked by a red square while the fuzzy sets from class 1 are marked by blue triangles. From Example 2, we know that A^3 is the fundamental memory most similar to the input X_0 with respect to $\mathcal{S}_H$. Therefore, the fNN based on $\mathcal{S}_H$ yields $\mathcal{C}_{fNN}(X_0) = \ell_3 = 0$. In other words, the fNN classifier assign to X_0 the class 0. We would like to point out that the SM-FAM classifier of Esmi et al. [12], based on the similarity measure $\mathcal{S}_H$, also attributes to X the class label 0.

As we will see, the GRE-FAM classifier, denoted by $\mathcal{C}_g$, is closely related to the fNN classifier. Furthermore, $\mathcal{C}_g$ can also be viewed as a sparse representation (SR) classifier [45]. Generally speaking, a SR classifier relies on the following hypothesis [45]: A sample Y from class i can be approximately written as a linear combination of the training data from class i, that is,

$$Y(u) \approx \sum_{\xi:\ell_\xi = i} \alpha_\xi A^\xi(u), \quad \forall u \in U. \quad (4.39)$$

Alternatively, Y can be expressed using all training data $A^1, \ldots, A^p$ as follows where $\alpha_\xi = 0$ for all ξ such that $\ell_\xi \neq i$:

$$Y(u) \approx \sum_{\xi=1}^{p} \alpha_\xi A^\xi(u), \quad \forall u \in U. \quad (4.40)$$

In other words, the components of the vector $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_p]^T \in \mathbb{R}^p$ are zero for all indexes which are not associated to class i – hence the name sparse representation classifier.

Now, as an associative memory model, a GRE-FAM is expected to remove noise from a corrupted input. As a consequence, if X corresponds to a corrupted version of a sample from class $i \in \mathcal{L}$, then the output X_1 of a single-step GRE-FAM should also belong to class i. On one hand, according to Eq.(4.12), the output X_1 satisfies

$$X_1(u) = \varphi\left(\sum_{\xi=1}^{p} w_{\xi 0} A^\xi(u)\right), \quad \forall u \in U. \quad (4.41)$$

On the other hand, we expect from Eq.(4.40) that

$$X_1(u) \approx \sum_{\xi=1}^{p} \alpha_\xi A^\xi(u), \quad \forall u \in U, \tag{4.42}$$

where $\alpha_\xi = 0$ for all ξ such that $\ell_\xi \neq i$. Apart from the piece-wise linear activation function φ, which can be ignored if $\sum_{\xi=1}^{p} w_\xi A^\xi(u) \in [0,1]$, Eq.(4.41) and Eq.(4.42) suggest that the weights $w_{10}, w_{20}, \ldots, w_{p0}$ are sparse. Moreover, we propose to compute the coefficients α_ξ in Eq.(4.42) by means of the equation

$$\alpha_\xi = w_{\xi 0}\chi_i(\ell_\xi), \quad \forall \xi = 1, ..., p, \tag{4.43}$$

where $\chi_i : \mathcal{L} \to \{0,1\}$, for $i \in \mathcal{L}$, is the indicator function defined by

$$\chi_i(x) = \begin{cases} 1, & x = i, \\ 0, & \text{otherwise.} \end{cases} \tag{4.44}$$

Note that Eq.(4.43) implies $\alpha_\xi = w_{\xi 0}$ if $\ell_\xi = i$ and $\alpha_\xi = 0$ otherwise. Concluding, if the input X belongs to class i, we guess that

$$X_1(u) \approx \sum_{\xi=1}^{p} w_{\xi 0}\chi_i(\ell_\xi) A^\xi(u), \quad \forall u \in U, \tag{4.45}$$

where $X_1(u)$ and the weights $w_{\xi 0}$'s, given respectively by Eq.(4.12) and Eq.(4.11), correspond to the outputs of the fourth and third hidden layer of a single-step GRE-FAM. Finally, since we do not know *a priori* which class the input belongs, we assign to X the class ℓ that minimizes the distance between X_1 and the linear combination $\sum_{\xi=1}^{p} w_{\xi 0}\chi_\ell(\ell_\xi)A^\xi$. To be specific, we attribute to X a class label $\ell \in \mathcal{L}$ such that

$$d_2\left(X_1, \sum_{\xi=1}^{p} w_{\xi 0}\chi_\ell(\ell_\xi) A^\xi\right) \leq d_2\left(X_1, \sum_{\xi=1}^{p} w_{\xi 0}\chi_i(\ell_\xi) A^\xi\right), \ \forall i \in \mathcal{L}. \tag{4.46}$$

where d_2 denotes the L_2-distance. Algorithm 4.1 summarizes a GRE-FAM classifier.

Let us discuss briefly the computational complexity of the GRE-FAM classifier described by Algorithm 4.1. For simplicity, consider the usual situation in which the number of samples p is much greater than both the dimension n of the data and the number $c = \text{Card}(\mathcal{L})$ of classes, i.e., $n << p$ and $c << p$. In this case, $\mathcal{O}(p^3)$ operations are required for the computation of the matrix G, which should be either provided by the user or computed in steps 1 and 2 of Algorithm 4.1. The remaining steps, which corresponds to the recall phase of the GRE-FAM plus c evaluations of the distance d_2, require all together $\mathcal{O}(p^2)$ operations. Since the matrix G need to be computed only once, we conclude that $\mathcal{O}(p^3)$ operations are performed to synthesize (training phase) a GRE-FAM classifier $\mathcal{C}_g : \mathcal{F}(U) \to \mathcal{L}$ while an evaluation of $\mathcal{C}_g$ demands $\mathcal{O}(p^2)$ operations.

Algorithm 4.1 GRE-FAM Classifier

Data: Training set $\mathcal{A}_{\mathcal{L}} = \{(A^\xi, \ell_\xi) : \xi = 1, \ldots, p\} \subseteq \mathcal{F}(U) \times \mathcal{L}$, a fuzzy similarity measure $\mathcal{S}$, and the parameter $\alpha > 0$;
Optionally, provide the matrix $G \in \mathbb{R}^{p \times p}$ determined according to Theorem 1.
Input: Input $X \in \mathcal{F}(U)$.
Output: Class label $\ell \in \mathcal{L}$.

if *the matrix* G *was not provided* **then**
    compute it from $A^1, \ldots, A^p$ according to Theorem 1;
Determine the weight $w_{\xi 0}$, for $\xi = 1, \ldots, p$, using Eq.(4.11);
Define the recalled fuzzy set $X_1 \in \mathcal{F}(U)$ by means of Eq.(4.12);
Choose $\ell \in \mathcal{L}$ and compute $\eta_\ell = d_2\left(X_1, \sum_{\xi=1}^{p} w_{\xi 0} \chi_\ell(\ell_\xi) A^\xi\right)$;
for *all* $i \in \mathcal{L} \setminus \{\ell\}$ **do**
    Compute

$$\eta_i = d_2\left(X_1, \sum_{\xi=1}^{p} w_{\xi 0} \chi_i(\ell_\xi) A^\xi\right); \tag{4.47}$$

    if $\eta_i < \eta_\ell$ **then**
        Update $\ell = i$ and $\eta_\ell = \eta_i$;

Example 7. Again, consider the labeled family $\mathcal{A}_{\mathcal{L}} = \{(A^\xi, \ell_\xi) : \xi = 1, \ldots, 8\}$, where the fuzzy sets and labels are given by Eq.(4.13), Eq.(4.14), and Eq.(4.38). Also, as shown in Fig.(4.3), let $X_0 = [0.4, 0.5]^T$ be the input fuzzy set. From Example 2, we know that the single-step GRE-FAM based on $\mathcal{S}_H$ yields $X_1 = [0.4, 0.5]^T$ as output. Moreover, from Eq.(4.18) and Algorithm 4.1, we obtain

$$\eta_0 = d_2\left(X_1, \sum_{\xi=1}^{p} w_{\xi 0} \chi_0(\ell_\xi) A^\xi\right) \tag{4.48}$$

$$= d_2\left(X_1, w_{1,0} A^1 + w_{2,0} A^2 + w_{3,0} A^3 + w_{4,0} A^4\right) \tag{4.49}$$

$$= d_2\left(\begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix}, -0.3 \begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix} + 0.87 \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix} - 0.3 \begin{bmatrix} 0.2 \\ 0.6 \end{bmatrix}\right) \tag{4.50}$$

$$= d_2\left(\begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0.05 \\ 0.13 \end{bmatrix}\right) = 0.51. \tag{4.51}$$

Note that, despite the large weight $w_{3,0} = 0.87$, the linear combination

$$\sum_{\xi=1}^{p} w_{\xi 0} \chi_0(\ell_\xi) A^\xi$$

is neither similar to X_1 nor A^3 because of the negative weights $w_{2,0} = w_{3,0} = -0.3$.

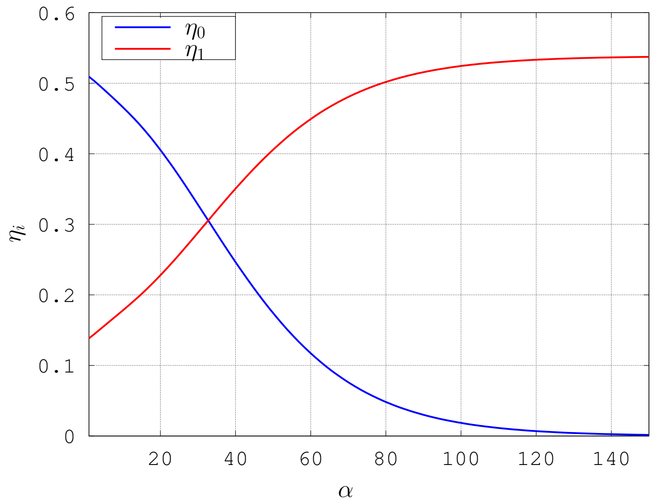

Figure 4.5: Values of η_i given by Eq.(4.47) by the parameter α of a GRE-FAM classifier.

Analogously, we have

$$\eta_1 = d_2\left(X_1, w_{5,0}A^5 + w_{6,0}A^6 + w_{7,0}A^7 + w_{8,0}A^8\right) = 0.14. \tag{4.52}$$

Since $\eta_1 < \eta_0$, the GRE-FAM classifier attributes to X_0 the class label 1, i.e., $\mathcal{C}_g(X_0) = 1$. Finally, Fig.(4.5) depicts the values of η_0 and η_1 by the parameter α. Note that $\eta_1 < \eta_0$ and, thus, $\mathcal{C}_g(X_0) = 1$ for $1 \leq \alpha \leq 32$. Similarly, we have $\mathcal{C}_g(X_0) = 0$ for $\alpha > 32$. Therefore, for α sufficiently large, the GRE-FAM classifier $\mathcal{C}_g$ coincide with both fNN and SM-FAM classifiers. We discuss the relationship between these classifiers after Remark 2.

Remark 2. Recall from Example 6 that the fNN and SM-FAM classifiers attributed to X_0 the class 0, which is the class label of the most similar fundamental memory. In contrast, the GRE-FAM classifier assigned to X_0 the class label 1. Although there is no correct solution for this classification problem, we suggest the following explanation for the solution provided by the GRE-FAM classifier. Notice in Fig.(4.3) that the fuzzy sets $A^1, \ldots, A^4$ from class 0 are more concentrated than the fuzzy sets $A^5, \ldots, A^8$ from class 1. Thus, since the fuzzy sets from class 0 are clustered about $A^3 = [0.2, 0.5]^T$, we guess that X_0 does not belong to class 0. The concentration of the fundamental memories are taken into account by the matrix G prescribed by Theorem 1, which is the inverse of a kind of similarity matrix.

Let us finish this section by pointing out the relationship between GRE-FAM, fNN, and SM-FAM classifiers. Suppose that there exists only one fundamental memory A^γ, with $A^\gamma \neq \emptyset$, such that $\mathcal{S}(A^\gamma, X) > \mathcal{S}(A^\xi, X)$ for all $\xi \neq \gamma$. In other words, A^γ is the unique fundamental memory most similar to the input X. From Eq.(4.37), the fNN evidently attributes to X the class label ℓ_γ of A^γ, that is, $\mathcal{C}_{fNN}(X) = \ell_\gamma$. Moreover, it is not hard to show that the SM-FAM of Esmi et al. also attributes to X the class label ℓ_γ [12]. Now, from Corollary 1, the output X_1 of the single-step GRE-FAM converges to A^γ as $\alpha \to \infty$. Also, the weight $w_{\xi 0}$ converges to the delta Kronecker $\delta_{\xi\gamma}$. Since d_2 is continuous, we have

$$\lim_{\alpha\to\infty} \eta_i = \lim_{\alpha\to\infty} d_2\left(X_1, \sum_{\xi=1}^{p} w_{\xi 0}\chi_i(\ell_\xi)A^\xi\right) = d_2\left(A^\gamma, \sum_{\xi=1}^{p} \delta_{\xi\gamma}\chi_i(\ell_\xi)A^\xi\right)$$
$$= d_2\left(A^\gamma, \chi_i(\ell_\gamma)A^\gamma\right), \quad \forall i \in \mathcal{L}. \tag{4.53}$$

On one hand, if A^γ belongs to class i, then

$$\lim_{\alpha\to\infty} \eta_i = d_2\left(A^\gamma, A^\gamma\right) = 0. \tag{4.54}$$

On the other hand, we have

$$\lim_{\alpha\to\infty} \eta_i = d_2\left(A^\gamma, \emptyset\right) > 0, \tag{4.55}$$

if A^γ does not belong to class i. Concluding, such as the fNN and the SM-FAM, the GRE-FAM classifier assigns to input X the class label ℓ_γ of the most similar fundamental memory A^γ when α is sufficiently large. In many practical situations, however, the parameter α cannot be made indefinitely large. As shown in Examples 6 and 7, the GRE-FAM classifiers may differ significantly from fNN as well as from SM-FAM if the parameter α is not sufficiently large. The next section compares further the GRE-FAM classifier with many other classifiers from the literature, including SM-FAM and fNN classifiers.

4.5 Computational Experiments

In this section, we evaluate the classification accuracy of GRE-FAM classifiers in some benchmark problems available on the internet. Furthermore, we compare them with some other classifiers from the literature. For simplicity, we will only consider the GRE-FAM models defined using the Gregson's similarity measure $\mathcal{S}_G$ and the complement of the relative Hamming distance $\mathcal{S}_H$. Moreover, we fix the parameter $\alpha = 30$.

Consider the following fifteen benchmark classification problems available at the *Knowledge Extraction Based on Evolutionary Learning* (KEEL) database repository: Appendicitis, cleveland, crx, ecoli, glass, heart, iris, monks, movementlibras, pima, sonar, spectfheart, vowel, wdbc, and wine [2]. Table 4.1 summarizes the information about the classification problems, also called data sets.

According to previous experiments described in the literature [1, 12], we evaluated the accuracy of the classifiers using 10-fold cross-validation. Specifically, we first partitioned each data set into 10 folds. Then, each folder served as a test set for one time while the union of the remaining 9 folders were used for training. The average of the accuracies obtained from these 10 experiments is used to evaluate the performance of a classifier. We would like to point out that we used the same partitioning as in [1, 2, 12] to ensure a fair comparison.

Table 4.1: Description of the classification problems.

	Instances	Categorical Features	Numerical Features	Classes
Appendicitis	106	0	7	2
Cleveland	297	0	13	5
Crx	653	9	6	2
Ecoli	336	0	7	8
Glass	214	0	9	7
Heart	270	0	13	2
Iris	150	0	4	3
Monks	432	0	6	2
Movementlibras	360	0	90	15
Pima	768	0	8	2
Sonar	208	0	60	2
Spectfheart	267	0	44	2
Vowel	990	0	13	11
Wdbc	569	0	30	2
Wine	178	0	13	3

Note that the data sets listed on Table 4.1 contain both categorical and numerical features. Therefore, some pre-processing are required to convert the original data into fuzzy sets. First, a categorical feature $f \in \{v_1, \ldots, v_c\}$, with $c > 1$, is transformed into a c-dimensional numerical feature $\mathbf{n} = (n_1, n_2, \ldots, n_c) \in \mathbb{R}^{\mathsf{c}}$ as follows for all $i = 1, \ldots, c$:

$$n_i = \begin{cases} 1, & f = v_i, \\ 0, & \text{otherwise}, \end{cases} \tag{4.56}$$

For example, the crx data set contain a categorical feature with 14 possibilities. Such categorical feature is transformed into 14 numerical features using Eq.(4.56). At the end, an instance of the transformed crx data set contain 46 numerical features instead of 9 categorical and 6 numerical features of the original classification problem. After converting categorical features into numerical values, an instance from a data set can be written as a pair $(\mathbf{x}, \ell)$, where $\mathbf{x} = [x_1, \ldots, x_n]^T \in \mathbb{R}^n$ is a vector of numerical features and $\ell \in \mathcal{L}$ denotes its class label. Furthermore, we can associated to each feature vector $\mathbf{x} \in \mathbb{R}^n$ a fuzzy set $A = [a_1, a_2, \ldots, a_n]^T$ by means of the equation

$$a_i = \frac{1}{1 + e^{-(x_i - \mu_i)/\sigma_i}} \in [0, 1], \quad \forall i = 1, \ldots, n, \tag{4.57}$$

where μ_i and σ_i represent respectively the mean and the standard deviation of ith component of all training instances. As a consequence, any training set can be written as a labeled family of fuzzy sets $\mathcal{A}_{\mathcal{L}} = \{(A^\xi, \ell_\xi) : \xi = 1, \ldots, p\}$.

Remark 3. Alternatively, the fuzzy set $A = [a_1, \ldots, a_n] \in [0, 1]^n$ can be derived from the feature vector $\mathbf{x} = [x_1, \ldots, x_n]^T \in \mathbb{R}^n$ by means of the equation

$$a_i = \varphi\left(\frac{x_i - x_i^{\text{min}}}{x_i^{\text{max}} - x_i^{\text{min}}}\right), \quad \forall i = 1, \ldots, n, \tag{4.58}$$

where x_i^{max} and x_i^{min} represent respectively the maximum and the minimum values of ith component of all training instances. Also, φ is the piece-wise linear function given by Eq.(4.8). In this case, however, some information may be lost if the ith component of the probed feature vector does not belong to the interval $[x_i^{\text{min}}, x_i^{\text{max}}]$, which has been determined from the training samples.

The last two columns of Table 4.2 show the classification accuracies produced by the GRE-FAM classifiers on each data set as well as the average accuracy on all classification problems. For comparison purpose, Table 4.2 also contains the accuracies produced by the following 12 classifiers: structural learning algorithm on vague environment (2SLAVE) [14], fuzzy hybrid genetic based machine learning algorithm (FH-GBML) [20], steady-state genetic algorithm to extract fuzzy classification rules from data (SGERD) [31], classification based on associations (CBA) [27], an improved version of the CBA method (CBA2) [28], classification based on multiple association rules (CMAR) [26], classification based on predictive association rules (CPAR) [48], a C4.5 decision tree (C4.5) [35], fuzzy association rule-based classification method for high-dimensional problems (FARC-HD) [1], a Θ-fuzzy associative memory ($\Theta-$FAM) [12], and the fNN presented in Example 5. We would like to point out that the accuracy values of the first 10 classifiers have been extracted from [1, 12]. Finally, the highest classification accuracy for each data set has been typed using boldface numbers.

Table 4.2 reveals that the GRE-FAM classifiers yielded very satisfactory classification performance. In fact, the GRE-FAM classifiers based on $\mathcal{S}_G$ and $\mathcal{S}_H$ produced the highest average accuracy rates of 84.84% and 85.98%, respectively. In addition, the GRE-FAM classifier based on $\mathcal{S}_H$ outperformed all the other classifiers in 5 of the 15 data sets under consideration, namely: Glass, movementlibras, sonar, wdbc, and wine. In particular, the classification performance of the GRE-FAM $\mathcal{S}_H$ classifier on the glass classification problem is approximately 12% higher than the CBA2 model, which yielded the highest accuracy reported in the literature before this work.

Table 4.2: Classification accuracy using 90-10 training-test split of the available data. The accuracy of the ten first classifiers has been extracted from [12].

	2SLAVE	FH-GBML	SGERD	CBA	CBA2	CMAR	CPAR	C4.5	FARC-HD	Θ-FAM	fNN S_G	fNN S_H	GRE-FAM S_G	GRE-FAM S_H
Appendicitis	82.91	86	84.48	89.6	89.6	**89.7**	87.8	83.3	84.2	81.18	82.09	82.09	85.09	85.09
Cleveland	48.82	53.51	51.59	**56.9**	54.9	53.9	54.9	54.5	55.2	51.17	44.24	43.55	53.88	55.6
Crx	74.06	86.6	85.03	83.6	85	85	**87.3**	85.3	86	82.02	63.6	63.73	85.58	85.13
Ecoli	**84.53**	69.38	74.05	78	77.1	77.7	76.2	79.5	82.2	76.78	79.2	78.90	83.97	83.65
Glass	58.05	57.99	58.49	70.8	71.3	70.3	68.9	67.4	70.2	70.49	73.81	73.81	78.19	**80.92**
Heart	71.36	75.93	73.21	83	81.5	82.2	80.7	78.5	**84.4**	78.15	67.78	66.67	80.74	82.59
Iris	94.44	94	94.89	93.3	93.3	94	**96**	**96**	**96**	**96**	95.33	94.67	94	94
Monks	97.26	98.18	80.65	**100**	**100**	**100**	**100**	**100**	99.8	98.63	88.34	90.44	97.27	98.86
Movementlibras	67.04	68.89	68.09	36.1	7.2	39.2	63.6	69.4	76.7	84.72	85.83	86.94	86.67	**88.06**
Pima	73.71	75.26	73.37	72.7	72.5	75.1	74.5	74	**75.7**	67.44	67.98	67.58	73.71	75.15
Sonar	71.42	68.24	71.9	75.4	77.9	78.8	75	70.5	80.2	80.69	84.55	84.64	83.1	**86.48**
Spectfheart	79.17	72.36	78.16	79.8	79.8	79.4	78.3	76.5	79.8	**81.3**	72.71	73.82	79.42	80.53
Vowel	71.11	67.07	65.83	63.6	74.9	60.4	63	81.5	71.8	97.07	93.74	**99.29**	99.09	99.09
Wdbc	92.33	92.26	90.68	94.7	95.1	94.9	95.1	95.2	95.3	96.14	93.67	93.67	96.31	**96.84**
Wine	89.47	92.61	91.88	93.8	93.8	96.7	95.6	93.3	94.3	97.24	84.87	83.20	95.52	**97.75**
Average	77.05	77.22	76.15	78.09	76.93	78.49	79.79	80.33	82.12	82.6	78.52	78.87	84.84	**85.98**

4.6 Concluding Remarks

In this chapter, we first reviewed the GRE-FAM models introduced by Souza et al. in [10]. In few words, a GRE-FAM is a multilayer recurrent neural network designed for the storage and recall of a family of fuzzy sets $\mathcal{A} = \{A^1, \ldots, A^p\}$. A GRE-FAM is characterized by the fundamental memories $A^1, \ldots, A^p$, a similarity measure $\mathcal{S}$, a parameter $\alpha > 0$, and a real-valued matrix $G \in \mathbb{R}^{p \times p}$. In this chapter, we showed how to compute a matrix G such that any fundamental memory is a fixed point of the GRE-FAM. As a consequence of this result, a GRE-FAM can implement high-capacity associative memory. Besides the high-storage capacity, we showed that the output X_1 of a single-step GRE-FAM converges point-wise to a linear combination of the fundamental memories most similar to the input X_0 as the parameter $\alpha \to \infty$.

The main contribution of this chapter is the application of GRE-FAMs to pattern classification. Inspired by sparse representation classifiers [45], we assumed that the output X_1 of a single-step GRE-FAM can be expressed as a linear combination of the fundamental memories that belong to the class of the input X_0. Using this hypothesis, we presented Algorithm 4.1, which summarizes the GRE-FAM classifier. Also, we pointed out that the relationship between GRE-FAM, fuzzy nearest neighbor (fNN), and similarity-measure fuzzy associative memory (SM-FAM) classifiers. An illustrative example revealed that the GRE-FAM classifier, with the matrix G prescribed by Theorem 1, may take into account the distribution of the training data.

The chapter finishes with applications of the GRE-FAM classifier to some well-know benchmark classification problems taken from the KEEL database [2]. The GRE-FAM classifiers produced competitive performances in terms of accuracy in comparison to many classifiers from the literature [1, 12]. Specifically, the GRE-FAM classifiers based on $\mathcal{S}_G$ and $\mathcal{S}_H$, both with the parameter $\alpha = 30$, yielded the highest average accuracy rates over all classification problems. Furthermore, the GRE-FAM classifier based on $\mathcal{S}_H$ outperformed all the other classifiers in 5 of the 15 data sets under consideration.

Acknowledgement

This work was supported in part by CNPq under grant no. 305486/2014-4 and FAPESP under grant nos. 2013/12310-4 and 2015/00745-1.

References

[1] J. Alcala-Fdez, R. Alcala, and F. Herrera. A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. *IEEE Transactions on Fuzzy Systems*, 19(5):857–872, 2011.

[2] J. Alcala-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. Garcia, L. Sanchez, and F. Herrera. KEEL data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, 17(2-3):255–287, 2011.

[3] J. Austin. *Handbook of Neural Computation*, chapter Associative memory, pages F1.4:1–F1.4:7. Oxford University Press, 1997.

[4] R. Ballini. Equality index and learning in recurrent fuzzy neural networks. In *12th IEEE International Conference on Fuzzy Systems (FUZZ '03)*, volume 1, pages 155–160, 2003.

[5] P.V. Binh and N.T. Hoa. *Context-Aware Systems and Applications*, volume 109 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, chapter Compare effective fuzzy associative memories for grey-scale image recognition, pages 258–267. Springer Berlin Heidelberg, 2013.

[6] I. Bloch. On fuzzy distances and their use in image processing under imprecision. *Pattern Recognition*, 32(11):1873–1895, 1999.

[7] T.D. Bui, T.H. Nong, and T.K. Dang. Improving learning rule for fuzzy associative memory with combination of content and association. *Neurocomputing*, 149(Part A):59–64, 2015.

[8] H. Bustince, M. Pagola, and E. Barrenechea. Construction of fuzzy indices from fuzzy di-subsethood measures: application to the global comparison of images. *Information Sciences*, 177(3):906–929, 2007.

[9] F.-L. Chung and T. Lee. On fuzzy associative memory with multiple-rule storage capacity. *IEEE Transactions on Fuzzy Systems*, 4(3):375–384, 1996.

[10] A.C. de Souza, M.E. Valle, and P. Sussner. Generalized recurrent exponential fuzzy associative memories based on similarity measures. In *16th World Congress of the International Fuzzy Systems Association (IFSA) and the 9th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT)*, pages 455–462, 2015.

[11] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York, USA, 1980.

[12] E. Esmi, P. Sussner, H. Bustince, and J. Fernandez. Theta-fuzzy associative memories (Theta-FAMs). *IEEE Transactions on Fuzzy Systems*, 23(2):313–326, 2015.

[13] E. Esmi, P. Sussner, M.E. Valle, F. Sakuray, and L. Barros. Fuzzy associative memories based on subsethood and similarity measures with applications to speaker identification. In *International Conference on Hybrid Artificial Intelligence Systems (HAIS 2012)*, volume 7209 of *Lecture Notes in Computer Science*, pages 479–490. Springer-Verlag Berlin Heidelberg, 2012.

[14] A. Gonzalez and R. Perez. Selection of relevant features in a fuzzy genetic learning algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 31(3):417–425, 2001.

[15] M.H. Hassoun, editor. *Associative Neural Memories: Theory and Implementation*. Oxford University Press, Oxford, U.K., 1993.

[16] M.H. Hassoun. *Fundamentals of Artificial Neural Networks*. MIT Press, Cambridge, MA, USA, 1995.

[17] M.H. Hassoun and P.B. Watta. *Handbook of Neural Computation*, chapter Associative memory networks, pages C1.3:1–C1.3:14. Oxford University Press, 1997.

[18] S. Haykin. *Neural Networks and Learning Machines*. Prentice-Hall, Upper Saddle River, NJ, USA, 3rd edition, 2009.

[19] J.J. Hopfield and D.W. Tank. Neural computation of decisions in optimization problems. *Biological Cybernetics*, 52(3):141–152, 1985.

[20] H. Ishibuchi, T. Yamamoto, and T. Nakashima. Hybridization of fuzzy GBML approaches for pattern classification problems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 35(2):359–365, 2005.

[21] F. Junbo, J. Fan, and S. Yan. A learning rule for fuzzy associative memories. In *IEEE International Joint Conference on Neural Networks (IJCNN)*, volume 7, pages 4273–4277, 1994.

[22] G.J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, Upper Saddle River, NY, USA, 1995.

[23] T. Kohonen. *Self-Organization and Associative Memory*. Springer-Verlag New York, Inc., New York, NY, USA, 2nd edition, 1987.

[24] S.G. Kong and B. Kosko. Adaptive fuzzy systems for backing up a truck-and-trailer. *IEEE Transactions on Neural Networks*, 3(2):211–223, 1992.

[25] B. Kosko. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Prentice Hall, Englewood Cliffs, NJ, USA, 1992.

[26] W. Li, J. Han, and J. Pei. CMAR: accurate and efficient classification based on multiple class-association rules. In *IEEE International Conference on Data Mining (ICDM)*, pages 369–376, 2001.

[27] B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. In *4th International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 80–86, 1998.

[28] B. Liu, Y. Ma, and C.-K. Wong. *Data Mining for Scientific and Engineering Applications*, volume 2 of *Massive Computing*, chapter Classification using association rules: weaknesses and enhancements, pages 591–605. Springer US, 2001.

[29] P. Liu. The fuzzy associative memory of max-min fuzzy neural networks with threshold. *Fuzzy Sets and Systems*, 107(2):147–157, 1999.

[30] R. Lowen and W. Peeters. Distances between fuzzy sets representing grey level images. *Fuzzy Sets and Systems*, 99(2):135–149, 1998.

[31] E.G. Mansoori, M.J. Zolghadri, and S.D. Katebi. SGERD: a steady-state genetic algorithm for extracting fuzzy classification rules from data. *IEEE Transactions on Fuzzy Systems*, 16(4):1061–1071, 2008.

[32] T.D. Ndousse. Fuzzy neural control of voice cells in ATM networks. *IEEE Journal on Selected Areas in Communications*, 12(9):1488–1494, 1994.

[33] I. Perfilieva. *Integrated Uncertainty in Knowledge Modelling and Decision Making*, volume 9376 of *Lecture Notes in Computer Science*, chapter A new model of a fuzzy associative memory, pages 32–42. Springer International Publishing, Switzerland, 2015.

[34] I. Perfilieva and M. Vajgl. Autoassociative fuzzy implicative memory on the platform of fuzzy preorder. In *16th World Congress of the International Fuzzy Systems Association (IFSA) and the 9th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT)*, pages 1598–1603, 2015.

[35] J.R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[36] M. Setnes, R. Babuska, U. Kaymak, and H.R. van Nauta Lemke. Similarity measures in fuzzy rule base simplification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 28(3):376–386, 1998.

[37] P. Sussner, E.L. Esmi, I. Villaverde, and M. Grana. The Kosko subsethood fuzzy

associative memory (KS-FAM): mathematical background and applications in computer vision. *Journal of Mathematical Imaging and Vision*, 42(2):134–149, 2012.

[38] P. Sussner, R. Miyasaki, and M.E. Valle. An introduction to parameterized IFAM models with applications in prediction. In *2009 IFSA World Congress and 2009 EUSFLAT Conference*, pages 247–252, 2009.

[39] P. Sussner and M.E. Valle. Gray-scale morphological associative memories. *IEEE Transactions on Neural Networks*, 17(3):559–570, 2006.

[40] P. Sussner and M.E. Valle. Implicative fuzzy associative memories. *IEEE Transactions on Fuzzy Systems*, 14(6):793–807, 2006.

[41] P. Sussner and M.E. Valle. Recall of patterns using morphological and certain fuzzy morphological associative memories with applications in classification and prediction. In *IEEE International Conference on Fuzzy Systems*, pages 209–216, 2006.

[42] P. Sussner and M.E. Valle. *Handbook of Granular Computing*, chapter Fuzzy associative memories and their relationship to mathematical morphology, pages 733–754. John Wiley and Sons, Inc., New York, USA, 2008.

[43] I.B. Turksen and Z. Zhong. An approximate analogical reasoning approach based on similarity measures. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(6):1049–1056, 1988.

[44] M.E. Valle and A.C. de Souza. On the recall capability of recurrent exponential fuzzy associative memories based on similarity measures. *Mathware & Soft Computing Magazine*, 22(1):33–39, 2015.

[45] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[46] L. Xuecheng. Entropy, distance measure and similarity measure of fuzzy sets and their relations. *Fuzzy Sets and Systems*, 52(3):305–318, 1992.

[47] R.R. Yager and D.P. Filev. Summarizing data using a similarity based mountain method. *Information Sciences*, 178(3):816–826, 2008.

[48] X. Yin and J. Han. CPAR: classification based on predictive association rules. In *SIAM International Conference on Data Mining*, pages 331–335, 2003.

[49] B.-L. Zhang, H. Zhang, and S.S. Ge. Face recognition by applying wavelet subband representation and kernel associative memory. *IEEE Transactions on Neural Networks*, 15(1):166–177, 2004.

[50] D. Zhang and W. Zuo. Computational intelligence-based biometric technologies. *IEEE Computational Intelligence Magazine*, 2(2):26–36, 2007.

[51] R. Zwick, E. Carlstein, and D.V. Budescu. Measures of similarity among fuzzy concepts: a comparative analysis. *International Journal of Approximate Reasoning*, 1(2):221–242, 1987.