Fuzzy Kernel Associative Memories with Application in Classification

Aline Cristina de Souza and Marcos Eduardo Valle[✉]

Department of Applied Mathematics, University of Campinas,
Campinas, São Paulo 13081-970, Brazil
s.alinedesouza@gmail.com, valle@ime.unicamp.br

Abstract. In this paper we introduce the class of fuzzy kernel associative memories (fuzzy KAMs). Fuzzy KAMs are derived from single-step generalized exponential bidirectional fuzzy associative memories by interpreting the exponential of a fuzzy similarity measure as a kernel function. The output of a fuzzy KAM is obtained by summing the desired responses weighted by a normalized evaluation of the kernel function. Furthermore, in this paper we propose to estimate the parameter of a fuzzy KAM by maximizing the entropy of the model. We also present two approaches for pattern classification using fuzzy KAMs. Computational experiments reveal that fuzzy KAM-based classifiers are competitive with well-known classifiers from the literature.

Keywords: Fuzzy associative memory · Similarity measure
Entropy · Pattern classification

1 Introduction

Associative memories (AMs) are mathematical models inspired by the human brain ability to store and recall information by means of associations [1]. More specifically, they are designed for the storage of a finite set of association pairs $\{(\mathbf{x}^\xi, \mathbf{y}^\xi), \xi = 1, \cdots, p\}$, called *fundamental memory set*. We refer to $\mathbf{x}^1, \ldots, \mathbf{x}^p$ as the stimuli and $\mathbf{y}^1, \ldots, \mathbf{y}^p$ as the desired responses. Furthermore, an AM should exhibit some error correction capability, i.e, it should yield the desired response $\mathbf{y}^\xi$ even upon the presentation of a corrupted version $\tilde{\mathbf{x}}^\xi$ of the stimulus $\mathbf{x}^\xi$. An AM is said autoassociative if $\mathbf{x}^\xi = \mathbf{y}^\xi$ for all $\xi = 1, \cdots, p$, and heteroassociative if there is at least one $\xi \in \{1, \cdots, p\}$ such that $\mathbf{x}^\xi \neq \mathbf{y}^\xi$. Applications of AM models include diagnosis [2], emotional modeling [3], pattern classification [4–7], and image processing and analysis [8,9].

The Hopfield neural network is one of the most widely known neural network used to implement an AM [10]. Despite its many successful applications, the Hopfield network suffers from a very low storage capacity [1]. A simple but significant improvement in storage capacity of the Hopfield network is achieved by the *recurrent correlation associative memory* (RCAMs) [11]. RCAMs are closely related to the dense associative memory model introduced recently by

Krotov and Hopfield to establish the duality between AMs and deep learning [12, 13]. Furthermore, a particular RCAM, called *exponential correlation associative memory* (ECAM), is equivalent to a certain recurrent kernel associative memory proposed by Garcia and Moreno [14,15].

Like the traditional Hopfield neural network, the original ECAM is an autoassociative memory designed for the storage and recall of bipolar vectors. Many applications of AMs, however, require either an heteroassociative memory or the storage and recall of real-valued data. A heteroassociative version of the ECAM, called *exponential bidirectional associative memory* (EBAM), have been proposed by Jeng et al. [16]. As to the storage and recall of real-valued vectors, Chiueh and Tsai introduced the *multivalued exponential recurrent associative memory* (MERAM) [17].

Recently, we introduced the class of *generalized recurrent exponential fuzzy associative memories* (GRE-FAMs), which have been effectively applied for pattern classification [5,18]. Briefly, GRE-FAMs are autoassociative fuzzy memories obtained from a generalization of a fuzzy version of the MERAM. Recall that a fuzzy associative memory is a fuzzy system designed for the storage and recall of fuzzy sets [19,20]. The *generalized exponential bidirectional fuzzy associative memories* (GEB-FAMs), which generalize the GRE-FAMs for the heteroassociative case, have been applied for face recognition [21]. We would like to point out, however, that the dynamic of the GEB-FAMs are not fully understood yet. In view of this remark, we mostly considered single-step versions of these two AM models.

Summarizing, on the one hand, GEB-FAMs can be viewed as a fuzzy version of the bidirectional (heteroassociative) ECAM. On the other hand, ECAM is equivalent to a certain kernel associative memory. Put together, these two remarks suggest us to interpret the single-step GEB-FAMs using (fuzzy) kernel functions and, from now on, we shall refer to them as fuzzy kernel associative memories (fuzzy KAMs). Recall that a kernel is informally defined as a similarity measure that can be thought of as a dot product on a high-dimensional feature space [22]. Disregarding formal definitions, we interpret a fuzzy similarity measure as a fuzzy kernel. Such interpretation lead us to information theoretical learning whose goal is to capture the information in the parameters of a learning machine [23]. In this paper, we propose to fine tune the parameter of a fuzzy KAM using information theoretical learning.

The paper is structured as follows. Next section presents the fuzzy kernel associative memory and some theoretical results. In this section, we also describe how information theoretical learning can be used to determine the parameter of a fuzzy KAM. An application of autoassociative and heteroassociative fuzzy KAMs for pattern classification is given in Sect. 3. Computational experiments on pattern classification are provided in Sect. 4. The paper finishes with the concluding remarks in Sect. 5.

2 Fuzzy Kernel Associative Memory

A fuzzy similarity measure, or simply similarity measure, is a function that associates to each pair of fuzzy sets a real number that expresses the degree of equality of these sets. According to De Baets and Meyer [24], a fuzzy similarity measure is a symmetric binary fuzzy relation on the family of all fuzzy sets $\mathcal{F}(U)$. In mathematical terms, a similarity measure is a mapping $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \to [0, 1]$ such that $\mathcal{S}(A, B) = \mathcal{S}(B, A)$ for all fuzzy sets $A, B \in \mathcal{F}(U)$. We speak of a strong similarity measure if $\mathcal{S}(A, B) = 1$ if, and only if, $A = B$.

Given a fuzzy similarity measure, the mapping $\kappa : \mathcal{F}(U) \times \mathcal{F}(U) \to [0, 1]$ defined by the following equation for $\alpha > 0$ is also a fuzzy similarity measure

$$\kappa(A, B) = e^{\alpha(S(A,B)-1)}. \tag{2.1}$$

Disregarding formal definitions, we shall refer to κ as a fuzzy kernel.

We would like to call the reader's attention to the dependence of the fuzzy kernel κ on the parameter α when $\mathcal{S}$ is a strong similarity measure. On the one hand, κ approximates the strict equality of fuzzy sets as α increases. Precisely, we have $\kappa(A, B) = 1$ if $A = B$ and $\kappa(A, B) = 0$ otherwise as $\alpha \to \infty$. On the other hand, $\kappa(A, B) = 1$ for all $A, B \in \mathcal{F}(U)$ as $\alpha \to 0$. In other words, κ is unable to discriminate fuzzy sets for sufficiently small $\alpha > 0$. Hence, the parameter α controls the capability of the fuzzy kernel to distinguish fuzzy sets.

Let us now introduce the fuzzy kernel associative memory (fuzzy KAM):

Definition 1 (Fuzzy KAM). *Consider a fundamental memory set $\{(A^\xi, B^\xi) : \xi = 1, \ldots, p\} \subset \mathcal{F}(U) \times \mathcal{F}(V)$. Let $\alpha > 0$ be a real number, $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \to [0, 1]$ a similarity measure, and H a $p \times p$ real-valued matrix. A fuzzy KAM is a mapping $\mathcal{K} : \mathcal{F}(U) \to \mathcal{F}(V)$ defined by the following equation where $X \in \mathcal{F}(U)$ is the input and $Y = \mathcal{K}(X) \in \mathcal{F}(V)$ is the output:*

$$Y(v) = \varphi \left(\frac{\displaystyle\sum_{\xi=1}^{p} \sum_{\mu=1}^{p} h_{\xi\mu} \kappa(A^\mu, X) B^\xi(v)}{\displaystyle\sum_{\eta=1}^{p} \sum_{\mu=1}^{p} h_{\eta\mu} \kappa(A^\mu, X)} \right). \tag{2.2}$$

Here, the piece-wise linear function $\varphi(x) = \max(0, \min(1, x))$ ensures $Y(v) \in [0, 1]$, for all $v \in V$.

Alternatively, we can write the output of a fuzzy KAM as

$$Y(v) = \varphi \left(\sum_{\xi=1}^{p} w_\xi B^\xi(v) \right) \quad \text{where} \quad w_\xi = \frac{\displaystyle\sum_{\mu=1}^{p} h_{\xi\mu} \kappa(A^\mu, X)}{\displaystyle\sum_{\eta=1}^{p} \sum_{\mu=1}^{p} h_{\eta\mu} \kappa(A^\mu, X)}. \tag{2.3}$$

In words, the output $Y = \mathcal{K}(X)$ is given by a linear combinations of the desired responses B^ξ's. Moreover, the coefficients of the linear combinations are calculated by using the parametrized fuzzy kernel κ given by (2.1).

As pointed out in the introduction, an autoassociative fuzzy KAM is equivalent to the single-step generalized recurrent exponential fuzzy associative memories (GRE-FAM) designed for the storage of a finite family of fuzzy sets $\mathcal{A} = \{A^i, i = 1, \cdots, p\} \subset \mathcal{F}(U)$ [18]. Similarly, the heteroassociative fuzzy KAM corresponds to the single-step generalized exponential bidirectional fuzzy associative memories (GEB-FAM) in the heteroassociative case [25].

The matrix H plays a very important role in the storage capacity and noise tolerance of a fuzzy KAM. The next theorem shows how to define the matrix H so that the fundamental memories are all correctly encoded in the memory.

Theorem 1. *Let* $\mathcal{A} = \{(A^\xi, B^\xi) : \xi = 1, \cdots, p\} \subset \mathcal{F}(U) \times \mathcal{F}(V)$ *be the fundamental memory set,* $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \to [0,1]$ *a similarity measure,* $\alpha > 0$ *a real number, and* $\kappa : \mathcal{F}(U) \times \mathcal{F}(U) \to [0,1]$ *a fuzzy kernel defined by* (2.1)*. If the matrix* $K = (k_{ij}) \in \mathbb{R}^{p \times p}$*, whose entries are defined by*

$$k_{ij} = \kappa(A^i, A^j), \quad \forall i, j = 1, \ldots, p, \tag{2.4}$$

is invertible, then the fuzzy KAM obtained by considering the matrix $H = K^{-1}$ *satisfies the identity* $\mathcal{K}(A^\xi) = B^\xi$ *for all* $\xi = 1, \ldots, p$*.*

Let us briefly address the computational effort required to synthesize a fuzzy KAM based on Theorem 1. First, the fuzzy kernel κ is evaluated $(p^2 + p)/2$ times to compute the symmetric $p \times p$ matrix K defined by (2.4). Then, instead of determining the inverse $H = K^{-1}$, we compute the LU factorization (or the Cholesky factorization if H is symmetric and positive definite) of K using $\mathcal{O}(p^3)$ operations. Then, the multiplication of H by a vector is replaced by the solution of two triangular systems during the recall phase. Summarizing, $\mathcal{O}(p^3)$ operations are performed to synthesize a fuzzy KAM.

The parameter α plays an important role on the noise tolerance of a fuzzy KAM. Briefly, the higher the parameter α, the greater the weight of the fundamental memories most similar to the input X in the calculation of the output $\mathcal{K}(X)$. In other words, increasing α emphasizes the role of the fundamental memories most similar to the input. Thus, in some sense, the parameter α controls how each fundamental memory contributes to the output of a fuzzy KAM. Formally, the next theorem states that, as α tends to infinity, the output $\mathcal{K}(X)$ converges point-wise to the arithmetic mean of the desired responses B^ξ's whose associated stimulus A^ξ's are the most similar to the input X.

Theorem 2. *Let* $\mathcal{A} = \{(A^\xi, B^\xi) : \xi = 1, \cdots, p\} \subseteq \mathcal{F}(U) \times \mathcal{F}(V)$ *be a family of fundamental memories and* $\mathcal{S}$ *a strong similarity measure. Suppose that the matrix* K *given by* (2.4) *is invertible for any* $\alpha > 0$*. Given a fuzzy set* $X \in \mathcal{F}(U)$*, define* $\Gamma \subseteq \{1, \ldots, p\}$ *as the set of the indexes of the stimulus which are the most similar to the input* X *in terms of* $\mathcal{S}$*, that is:*

$$\Gamma = \{\gamma : \mathcal{S}(A^\gamma, X) \geq \mathcal{S}(A^\xi, X), \forall \xi = 1, \ldots, p\}. \tag{2.5}$$

Then,

$$\lim_{\alpha \to \infty} Y(v) = \frac{1}{\operatorname{Card}(\Gamma)} \sum_{\gamma \in \Gamma} B^{\gamma}(v), \quad \forall v \in V. \tag{2.6}$$

where $Y = \mathcal{K}(X)$ is the output of a fuzzy KAM. Furthermore, the weights w_ξ given by (2.3) satisfy the following equation for all $\xi = 1, \ldots, p$:

$$\lim_{\alpha \to \infty} w_\xi = \begin{cases} \dfrac{1}{\operatorname{Card} \Gamma}, & \xi \in \Gamma, \\ 0, & otherwise. \end{cases} \tag{2.7}$$

2.1   Estimating the Parameter of a Fuzzy KAM

In this section, we propose to estimate the parameter α of a fuzzy KAM using information theoretical learning [23]. The basic idea is to maximize the capability of the fuzzy kernel κ to discriminate between two different stimulus. To this end, we use the information-theoretic descriptor of entropy.

The concept of entropy, introduced by Shannon in 1948 [26], represents a quantitative measure of uncertainty and information of a probabilistic system [27,28]. The entropy of a n-state system is defined by the following equation where p_i denotes the probability of occurrence of the i-th state [27]:

$$E = \sum_{i=1}^{n} p_i \log(1/p_i). \tag{2.8}$$

The entropy given by (2.8) can be used as a measure of the amount of uncertainty of a system. Given a fundamental memory set $\mathcal{A} = \{(A^\xi, B^\xi) : \xi = 1, \ldots, p\} \subseteq \mathcal{F}(U) \times \mathcal{F}(V)$ and a strong similarity measure $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \to [0,1]$, we define the entropy of a fuzzy KAM $\mathcal{K}$ by means of the equation

$$E_\mathcal{K}(\alpha) = \sum_{i=1}^{p} \sum_{j=1}^{p} \kappa(A^i, A^j) \log\left(1/\kappa(A^i, A^j)\right) \tag{2.9}$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{p} -e^{\alpha(\mathcal{S}(A^i, A^j)-1)} \log\left(e^{\alpha(\mathcal{S}(A^i, A^j)-1)}\right) \tag{2.10}$$

$$= -\sum_{i=1}^{p} \sum_{j=1}^{p} \alpha(\mathcal{S}(A^i, A^j) - 1)) e^{\alpha(\mathcal{S}(A^i, A^j)-1)}. \tag{2.11}$$

Note that the entropy $E_\mathcal{K}$ of a fuzzy KAM is a function of the parameter α. Furthermore, $E_\mathcal{K}(\alpha)$ tend to zero if either $\alpha \to 0$ or $\alpha \to \infty$. Intuitively, $E_\mathcal{K}$ quantifies the capability of the fuzzy kernel κ to discriminate between A^i and A^j as a function of α. By maximizing $E_\mathcal{K}$, we expect to improve the noise tolerance of the fuzzy KAM. In view of this remark, we suggest to choose the parameter α^* that maximizes (2.11). Formally, we propose to define

$$\alpha^* = \operatorname*{argmax}_{\alpha > 0} E_\mathcal{K}(\alpha). \tag{2.12}$$

At this point, we would like to recall that Shannon derived (2.8) using classical probability theory. A fuzzy entropy that does not take into account probabilistic concepts in its definition have been provided by De Luca and Termini [29]. Although the fuzzy entropy would be more appropriate in our context, we have not observed significant improvements in our preliminary computational experiments using the fuzzy entropy compared to those obtained using the entropy of Shannon. Moreover, besides presenting lower computational cost, Shannon's entropy showed to be more robust. Therefore, we only consider the entropy of Shannon in this paper.

3 Classifiers Based on Fuzzy KAMs

A classifier is a mapping $\mathcal{C} : W \to \mathcal{L}$ that associates to each pattern $w \in W$ a label $l \in \mathcal{L}$ that represents the class which w belongs to. Classifiers are usually synthesized using a family of labeled samples, called *training set*. In this section, we present two approaches to define classifiers based on fuzzy KAMs. The first approach, which is inspired by sparse representation classifiers [30], is based on autoassociative fuzzy KAMs. The second approach contemplates the heteroassociative case.

3.1 The Autoassociative-Based Approach

Sparse representation classifiers [30] are based on the hypothesis that a sample Y from class i is approximately equal to a linear combination of the training data from class i. Formally, let $\mathcal{A}_\mathcal{L} = \{(A^\xi, \ell_\xi), \xi = 1, \cdots, p\} \subset \mathcal{F}(U) \times \mathcal{L}$ be the training set, where A^ξ are distinct non-empty fuzzy sets on U and $\mathcal{L}$ is a finite set of labels. If Y belongs to class i, then

$$Y(u) \approx \sum_{\xi:\ell_\xi=i} \alpha_\xi A^\xi(u), \quad \forall u \in U. \tag{3.13}$$

Equivalently, Y can be written as:

$$Y(u) \approx \sum_{\xi=1}^{p} \alpha_\xi A^\xi(u), \quad \forall u \in U, \tag{3.14}$$

where $\alpha_\xi = 0$ if $\ell_\xi \neq i$. In other words, Y can be written as a sparse linear combination of the training data.

Assume we have an autoassociative fuzzy KAM $\mathcal{K} : \mathcal{F}(U) \to \mathcal{F}(U)$ designed for the storage of the fundamental memory set $\{\mathcal{A}^1, \ldots, \mathcal{A}^p\}$. Given a pattern X from class i (or a noisy version $\tilde{X}$ of X), the autoassociative fuzzy KAM is expected to produce a pattern $\mathcal{K}(X) = Y$ that also belongs to class i. From (2.3), the output of the autoassociative fuzzy KAM satisfies

$$Y(u) = \varphi \left(\sum_{\xi=1}^{p} w_\xi A^\xi(u) \right), \tag{3.15}$$

where $\varphi(x) = \max(0, \min(1, x))$. By comparing (3.14) and (3.15), except for the piece-wise linear function φ which can be disregarded if $\sum_{\xi=1}^{p} w_\xi A^\xi(u) \in [0,1]$, we conclude that the linear combination in (3.15) should also be sparse. Therefore, the coefficients α_ξ in (3.14) can be approximated by

$$\alpha_\xi = w_\xi \chi_i(\ell_\xi), \quad \forall \xi = 1, \ldots, p, \qquad (3.16)$$

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

$$\chi_i(x) = \begin{cases} 1, & x = i, \\ 0, & \text{otherwise.} \end{cases} \qquad (3.17)$$

Observe that (3.16) implies $\alpha_\xi = w_\xi$ if $\ell_\xi = i$ and $\alpha_\xi = 0$ otherwise. Concluding, if the input X belongs to class i, we presuppose that

$$Y(u) \approx \sum_{\xi=1}^{p} w_\xi \chi_i(\ell_\xi) A^\xi(u), \quad \forall u \in U, \qquad (3.18)$$

In practice, however, we do not know *a priori* to which class the input X belongs. As a consequence, we assign to X the class $i \in \mathcal{L}$ that minimizes the distance between Y and the linear combination $\sum_{\xi=1}^{p} w_\xi \chi_\ell(\ell_\xi) A^\xi$. Formally, we attribute to X a class label $i \in \mathcal{L}$ such that

$$d_2 \left(Y, \sum_{\xi=1}^{p} w_\xi \chi_i(\ell_\xi) A^\xi \right) \leq d_2 \left(Y, \sum_{\xi=1}^{p} w_\xi \chi_j(\ell_\xi) A^\xi \right), \forall j \in \mathcal{L}, \qquad (3.19)$$

where d_2 denotes the L_2-distance.

3.2   Heteroassociative-Based Approach

In the second approach, we define a classifier using the heteroassociative case. Precisely, we synthesize a heteroassociative fuzzy KAM designed for the storage of a fundamental memory set $\{(A^\xi, B^\xi), \xi = 1, \ldots, p\} \subset \mathcal{F}(U) \times \{0,1\}^n$, where $n = \mathrm{Card}(\mathcal{L})$ denotes the number of classes, A^ξ represents a sample from a certain class, and $B^\xi \subset \{0,1\}^n$ indicates to which class A^ξ belongs. In mathematical terms, the (fuzzy) set B^ξ associated to the stimulus A^ξ of class i is defined by:

$$B^\xi(v) = \begin{cases} 1, & \text{if } v = i \\ 0, & \text{if } v \neq i. \end{cases} \qquad (3.20)$$

Now, given an input $X \in \mathcal{F}(U)$, the fuzzy KAM yields a fuzzy set $Y = \mathcal{K}(X)$. According to Eqs. (2.2) and (3.20), $Y(i)$ is the sum of the weights w_ξ's for ξ such that A^ξ belongs to the class i. Hence, we associate the input X to the i-th class, where i is the first index such that $Y(i) \geq Y(j)$, for all $j = 1, \ldots, n$.

4   Computational Experiments and Results

In this section, we carry out computational experiments to evaluate the performance of the fuzzy KAM-based classifiers. Let us begin by clarifying the benchmark classification problems that we have used.

4.1   Classification Problems

Let us consider the following twenty two classification problems available at the *Knowledge Extraction Based on Evolutionary Learning* (KEEL) database repository as well as at the UCI Machine Learning Repository: appendicitis, cleveland, crx, ecoli, glass, heart, iris, monks, movementlibras, pima, sonar, spectfheart, vowel, wdbc, wine, satimage, texture, german, yeast, spambase, phoneme, and page-blocks [31]. We would like to point out that, due to computational limitations, we refrained to consider the classification problems: magic, pen-based, ringnorm, and twonorm. Precisely, recall that $\mathcal{O}(p^3)$ operations are performed to synthesize a fuzzy KAM and, in these four databases, we have $p \approx 10^4$.

Similar to previous experiments described on the literature, the experiments were conducted by using ten-fold cross validation technique. This method consists of dividing the data-set in ten parts and performing 10 tests, each one using one of the parts as a test set and the others nine parts as developing/training set. Afterward, we compute the mean of the ten accuracy values obtained in each one of the ten tests. In order to ensure a fair comparison, we used the same partitioning as in [31,32].

Some of the data sets considered in this experiment contain both categorical and numerical features. Therefore, a pre-processing step to convert the original data into fuzzy sets was necessary. First, in order to have only numerical values, we transformed each categorical feature $f \in \{v_1, \ldots, v_c\}$, with $c > 1$, into a c-dimensional numerical feature $\mathbf{n} = (n_1, n_2, \ldots, n_c) \in \mathbb{R}^c$ as follows for all $i = 1, \ldots, c$:

$$n_i = \begin{cases} 1, & f = v_i, \\ 0, & \text{otherwise.} \end{cases} \tag{4.21}$$

For example, the crx data set contain nine categorical features, one of them with 14 possibilities. Such categorical feature was transformed into 14 numerical features using (4.21). 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 all categorical features were converted 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. Moreover, each feature vector $\mathbf{x} \in \mathbb{R}^n$ can be associated with a fuzzy set $A = [a_1, a_2, \ldots, a_n]^T$ by means of the equation

$$a_i = \frac{1}{1 + e^{-(x_i - \mu_i)/\sigma_i}} \in [0, 1], \quad \forall i = 1, \ldots, n, \tag{4.22}$$

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

Besides, we have removed some repeated elements from the fundamental memories sets of the spambase and page-blocks data sets.

In our computational experiments, we considered the fuzzy KAM defined by using the Gregson similarity measure and the parameter α^* that maximizes the entropy. The Gregson similarity measure $\mathcal{S}_G : \mathcal{F}(U) \times \mathcal{F}(U) \to [0,1]$ is given by:

$$\mathcal{S}_G(A, B) = \frac{\displaystyle\sum_{i=1}^{n} \min(A(u_i), B(u_i))}{\displaystyle\sum_{i=1}^{n} \max(A(u_i), B(u_i))}. \tag{4.23}$$

Note that $\mathcal{S}_G$ given by (4.23) is a strong similarity measure which can be interpreted as the quotient between the cardinality of the intersection by the cardinality of the union of A and B. Finally, we would like to point out that we studied extensively the role of a fuzzy similarity measure in GEB-FAM models applied for face recognition and the Gregson similarity measure achieved competitive results in comparison with others models from the literature [21].

Figure 1 shows the boxplot of the average accuracy produced by the autoassociative and heteroassociative fuzzy KAM-based classifiers as well as other nine models from the literature, namely: 2SLAVE [33], FH-GBML [34], SGERD [35], CBA [36], CBA2 [37], CMAR [38], CPAR [39], C4.5 [40], and FARC-HD [32]. The accuracy of the nine first classifiers have been extract from [32]. We can observe from Fig. 1 that the fuzzy KAM-based classifiers outperformed (or are at least competitive!) with the other classifiers from the literature. Let us
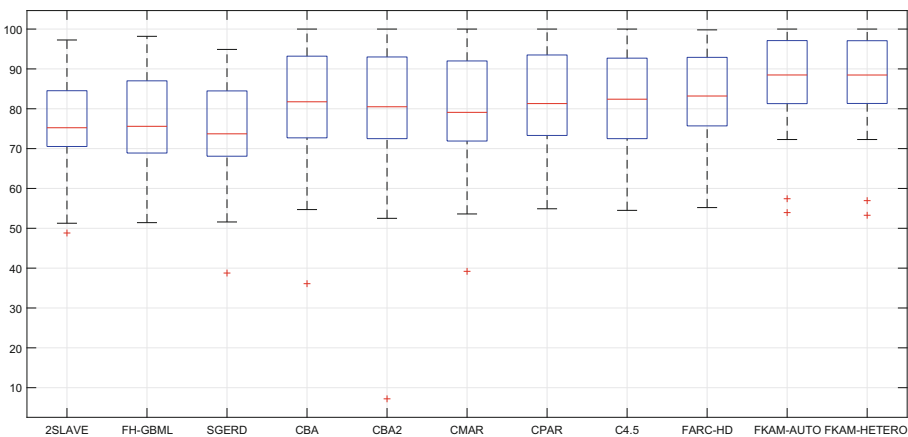

Fig. 1. Boxplot of classification accuracies of several models of the literature in twenty two problems. The accuracy of the nine first classifiers have been extract from [32].

conclude by pointing out that the outliers in the boxplot of the fuzzy KAM-based classifiers correspond to the Cleveland and Yeast classification problems.

5   Concluding Remarks

In this paper, we introduced the class of fuzzy kernel associative memories (fuzzy KAMs). Basically, a fuzzy KAM corresponds to single-step generalized exponential bidirectional fuzzy associative memories (GEB-FAMs), which have been introduced and investigated by us in the last years [5,18,21,25]. Like the single-step GEB-FAMs, fuzzy KAMs can be applied in classification problems. Indeed, in this paper we reviewed two approaches for pattern classification: one using the autoassociative case and the other based on the heteroassociative case.

The main contribution of this paper is the new interpretation of the exponential of a similarity measure as a kernel function. Although we did not elaborated rigorously on the notion of a fuzzy kernel, it allowed us to apply concepts from information theoretical learning to fine tune the parameter α of a fuzzy KAM. Precisely, in view of its simplicity, we proposed to determine the parameter α^* that maximizes the Shannon entropy of a fuzzy KAM.

Computational experiments with some well-know benchmark classification problems showed a superior performance, in terms of accuracy, of the fuzzy KAM-based classifiers over many other classifiers from the literature. In the future, we plan to formalize the notion of a fuzzy kernel and to investigate further the performance of the fuzzy KAM-based classifiers. We also intent to study other applications of the fuzzy KAM models.

References

1. Hassoun, M.H.: Fundamentals of Artificial Neural Networks. MIT Press, Cambridge (1995)
2. Njafa, J.P.T., Engo, S.N.: Quantum associative memory with linear and non-linear algorithms for the diagnosis of some tropical diseases. Neural Netw. **97**, 1–10 (2018)
3. Masuyama, N., Loo, C.K., Seera, M.: Personality affected robotic emotional model with associative memory for human-robot interaction. Neurocomputing **272**, 213–225 (2018)
4. Esmi, E., Sussner, P., Sandri, S.: Tunable equivalence fuzzy associative memories. Fuzzy Sets Syst. **292**, 242–260 (2016)
5. Valle, M.E., de Souza, A.C.: Pattern classification using generalized recurrent exponential fuzzy associative memories. In: George, A., Papakostas, A.G.H., Kaburlasos, V.G. (eds.) Handbook of Fuzzy Sets Comparison Theory, Algorithms and Applications, vol. 6, pp. 79–102. Science Gate Publishing (2016)
6. Li, L., Pedrycz, W., Li, Z.: Development of associative memories with transformed data. Appl. Soft Comput. **61**, 1141–1152 (2017)

7. Ramírez-Rubio, R., Aldape-Pérez, M., Yáñez-Márquez, C., López-Yáñez, I., Camacho-Nieto, O.: Pattern classification using smallest normalized difference associative memory. Pattern Recogn. Lett. **93**, 104–112 (2017)

8. Grana, M., Chyzhyk, D.: Image understanding applications of lattice autoassociative memories. IEEE Trans. Neural Netw. Learn. Syst. **27**(9), 1920–1932 (2016)

9. Valdiviezo-N, J.C., Urcid, G., Lechuga, E.: Digital restoration of damaged color documents based on hyperspectral imaging and lattice associative memories. SIViP **11**(5), 937–944 (2017)

10. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. Proc. Nat. Acad. Sci. **79**, 2554–2558 (1982)

11. Chiueh, T.D., Goodman, R.M.: Recurrent correlation associative memories. IEEE Trans. Neural Netw. **2**(2), 275–284 (1991). https://doi.org/10.1109/72.80338

12. Demircigil, M., Heusel, J., Löwe, M., Upgang, S., Vermet, F.: On a model of associative memory with huge storage capacity. J. Stat. Phys. **168**(2), 288–299 (2017)

13. Krotov, D., Hopfield, J.J.: Dense associative memory for pattern recognition (2016)

14. García, C., Moreno, J.A.: The hopfield associative memory network: improving performance with the kernel "Trick". In: Lemaître, C., Reyes, C.A., González, J.A. (eds.) IBERAMIA 2004. LNCS (LNAI), vol. 3315, pp. 871–880. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30498-2_87

15. Perfetti, R., Ricci, E.: Recurrent correlation associative memories: a feature space perspective. IEEE Trans. Neural Netw. **19**(2), 333–345 (2008)

16. Jeng, Y.J., Yeh, C.C., Chiueh, T.D.: Exponential bidirectional associative memories. Eletron. Lett. **26**(11), 717–718 (1990). https://doi.org/10.1049/el:19900468

17. Chiueh, T.D., Tsai, H.K.: Multivalued associative memories based on recurrent networks. IEEE Trans. Neural Netw. **4**(2), 364–366 (1993)

18. Souza, A.C., Valle, M.E., Sussner, P.: Generalized recurrent exponential fuzzy associative memories based on similarity measures. In: Proceedings of the 16th World Congress of the International Fuzzy Systems Association (IFSA) and the 9th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), vol. 1, pp. 455–462. Atlantis Press (2015). https://doi.org/10.2991/ifsa-eusflat-15.2015.66

19. Kosko, B.: Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence. Prentice Hall, Englewood Cliffs (1992)

20. Valle, M.E., Sussner, P.: A general framework for fuzzy morphological associative memories. Fuzzy Sets Syst. **159**(7), 747–768 (2008)

21. Souza, A.C., Valle, M.E.: Generalized exponential bidirectional fuzzy associative memory with fuzzy cardinality-based similarity measures applied to face recognition. In: Trends in Applied and Computational Mathematics (2018). Accepted for publication

22. Schölkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT press, Cambridge (2002)

23. Principe, J.C.: Information theory, machine learning, and reproducing kernel Hilbert spaces. Information Theoretic Learning. ISS, pp. 1–45. Springer, New York (2010). https://doi.org/10.1007/978-1-4419-1570-2_1

24. Baets, B.D., Meyer, H.D.: Transitivity-preserving fuzzification schemes for cardinality-based similarity measures. Eur. J. Oper. Res. **160**(3), 726–740 (2005). https://doi.org/10.1016/j.ejor.2003.06.036

25. Souza, A.C., Valle, M.E.: Memória associativa bidirecional exponencial fuzzy generalizada aplicada ao reconhecimento de faces. In: Valle, M.E., Dimuro, G., Santiago, R., Esmi, E. (eds.) Recentes Avanços em Sistemas Fuzzy, vol. 1, pp. 503–514. Sociedade Brasileira de Matemática Aplicada e Computacional (SBMAC), São Carlos - SP (2016). ISBN 978-85-8215-079-5

26. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423 (1948)

27. Pal, N., Pal, S.: Higher order fuzzy entropy and hybrid entropy of a set. Inf. Sci. **61**(3), 211–231 (1992)

28. Klir, G.J.: Uncertainty and Information: Foundations of Generalized Information Theory. Wiley-Interscience, Hoboken (2005)

29. De Luca, A., Termini, S.: A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. Inf. Control **20**(4), 301–312 (1972)

30. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 210–227 (2009)

31. Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. J. Multiple-Valued Log. Soft Comput. **17**(2–3), 255–287 (2011)

32. Alcalá-Fdez, J., Alcalá, R., Herrera, F.: A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. IEEE Trans. Fuzzy Syst. **19**(5), 857–872 (2011)

33. González, A., Pérez, R.: Selection of relevant features in a fuzzy genetic learning algorithm. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **31**(3), 417–425 (2001)

34. Ishibuchi, H., Yamamoto, T., Nakashima, T.: Hybridization of fuzzy GBML approaches for pattern classification problems. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **35**(2), 359–365 (2005)

35. Mansoori, E.G., Zolghadri, M.J., Katebi, S.D.: SGERD: a steady-state genetic algorithm for extracting fuzzy classification rules from data. IEEE Trans. Fuzzy Syst. **16**(4), 1061–1071 (2008)

36. Liu, B., Hsu, W., Ma, Y.: Integrating classification and association rule mining. In: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (1998)

37. Liu, B., Ma, Y., Wong, C.-K.: Classification using association rules: weaknesses and enhancements. In: Grossman, R.L., Kamath, C., Kegelmeyer, P., Kumar, V., Namburu, R.R. (eds.) Data Mining for Scientific and Engineering Applications. MC, vol. 2, pp. 591–605. Springer, Boston, MA (2001). https://doi.org/10.1007/978-1-4615-1733-7_30

38. Li, W., Han, J., Pei, J.: CMAR: accurate and efficient classification based on multiple class-association rules. In: 2001 Proceedings of IEEE International Conference on Data Mining. ICDM 2001, pp. 369–376. IEEE (2001)

39. Yin, X., Han, J.: CPAR: classification based on predictive association rules. In: Proceedings of the 2003 SIAM International Conference on Data Mining, pp. 331–335. SIAM (2003)

40. Quinlan, J.: C4. 5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco (1993). ISBN 1-55860-238-0