ARTICLE IN PRESS
FSS:6785

Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems ••• (••••) •••–•••

FUZZY sets and systems

www.elsevier.com/locate/fss

Tunable equivalence fuzzy associative memories

Estevão Esmi [a], Peter Sussner [a,*], Sandra Sandri [b]

[a] *University of Campinas, Department of Applied Mathematics, Campinas, SP 13083-859, Brazil*
[b] *Brazilian National Institute for Space Research, São José dos Campos, SP 12227-010, Brazil*

Received 31 January 2014; received in revised form 24 March 2015; accepted 11 April 2015

Abstract

This paper introduces a new class of fuzzy associative memories (FAMs) called *tunable equivalence fuzzy associative memories*, for short tunable E-FAMs or TE-FAMs, that are determined by the application of parametrized equivalence measures in the hidden nodes. Tunable E-FAMs belong to the class of Θ-FAMs that have recently appeared in the literature. In contrast to previous Θ-FAM models, tunable E-FAMs allow for the extraction of a fundamental memory set from the training data by means of an algorithm that depends on the evaluation of equivalence measures. Furthermore, we are able to optimize not only the weights corresponding to the contributions of the hidden nodes but also the contributions of the attributes of the data by tuning the parametrized equivalence measures used in a TE-FAM model. The computational effort involved in training tunable TE-FAMs is very low compared to the one of the previous Θ-FAM training algorithm.

Keywords: Parametrized equivalence measure; Fuzzy associative memory; Tunable equivalence fuzzy associative memory; Selection of fundamental memories; Supervised learning; Classification

1. Introduction

The class of Θ-fuzzy associative memories (Θ-FAMs) represents a class of fuzzy associative memories (FAMs) with a competitive hidden layer and with weights that can be tuned using a specific training algorithm [1].

A Θ-FAM is determined by functions Θ^ξ that are applied in the ξth hidden node. Previous publications on Θ-FAMs have focused on the cases where the functions Θ^ξ are given by fuzzy subsethood or similarity measures, leading to (weighted) subsethood, dual subsethood, and similarity measure FAMs [1,2]. These models, referred to using the acronyms S-FAMs, dual S-FAMs, and SM-FAMs (or weighted S-FAMs, dual S-FAMS, and SM-FAMs if one wishes to stress the inclusion of adjustable weights in their first layers), have found inspiration in fuzzy mathematical morphology (FMM) [3–6]. In this context, recall that in FMM the elementary operator of fuzzy erosion is defined in terms of the degree of inclusion or subsethood of objects (i.e., translated versions of the so called structuring element) in another object (given by the image under consideration) [7]. In addition, certain similarity measures employed

* Corresponding author.
E-mail addresses: eelaureano@gmail.com (E. Esmi), sussner@ime.unicamp.br (P. Sussner), sandri@lac.inpe.br (S. Sandri).

http://dx.doi.org/10.1016/j.fss.2015.04.004
0165-0114/

Please cite this article in press as: E. Esmi et al., Tunable equivalence fuzzy associative memories, Fuzzy Sets and Systems (2015), http://dx.doi.org/10.1016/j.fss.2015.04.004

in (weighted) SM-FAMs can be related to fuzzy hit-or-miss transforms. Therefore, (dual) subsethood and similarity measure FAMs can be viewed as fuzzy morphological neural networks, more specifically as *fuzzy morphological associative memories* (FMAMs) [8].

Previously, many well-known fuzzy associative memories (FAMs) from the literature had already been classified as FMAMs [8] since they perform elementary morphological operations in the complete lattice setting [9–11] of fuzzy mathematical morphology [6]. Examples of FMAMs include Kosko's max–min and max-product FAMs [12], the generalized FAM of Chung and Lee [13], Junbo's FAM [14], the max–min FAM with threshold of Liu [15], the fuzzy logical bidirectional associative memory of Belohlávek [16], as well as implicative fuzzy associative memories [17]. All of these FAM models represent fully-connected fuzzy neural networks without hidden layers [18,19]. As mentioned before, Θ-FAMs are equipped with a competitive hidden layer. In this respect, Θ-FAMs are similar to the well-known Hamming net [20] or Hamming associative memory and its extensions [21,22].

The aforementioned Θ-FAM models, i.e., weighted S-FAMs, dual S-FAMs, and SM-FAMs, were successfully applied to a number of classification problems and to a problem of vision-based self-localization in robotics [1]. The tuning of the weights was performed using a training algorithm that is guaranteed to converge in a finite number of steps and, under some weak conditions, to reach a local minimum of the proposed objective function. On the downside, the present Θ-FAM training algorithm is computationally very expensive, turning its use infeasible on very large datasets. Another option for tuning the weights of the aforementioned Θ-FAM models would be the use of a derivative-free non-linear optimization method such as a genetic algorithm. However, the application of a genetic algorithm to problems involving a large number of variables often becomes impracticable due to its high computational cost [23].

Here, we propose a computationally efficient alternative for tuning the weights of certain Θ-FAM models that are introduced in this paper. The first stage of the methodology proposed in this paper consists in extracting relevant fundamental memories from the training data by means of a selection algorithm that can only be applied to these types of Θ-FAMs. More precisely, this algorithm is based on the evaluation of parametrized equivalence measures, that can be chosen to serve as functions Θ^ξ in a Θ-FAM model. The resulting Θ-FAM models are named *tunable equivalence fuzzy associative memories*, for short, *tunable E-FAMs*. If the underlying equivalence measures are differentiable with respect to their parameters, then the parameters and the weights of a tunable E-FAM can be tuned in the second stage of our methodology using a conventional nonlinear optimization algorithm [24].

Since equivalence measures can be defined on any bounded lattice, we begin by providing some background information on lattice theory and equivalence measures. Section 3 introduces tunable E-FAM models in the context of Θ-FAMs. Section 4 describes our strategy for clustering the training data in terms of equivalence measures and for training the weights and parameters of tunable E-FAM models. Section 5 is concerned with the application of tunable E-FAMs to several classification problems from the KEEL and UCI databases [25,26]. Section 5 reveals that the classification rates produced by the tunable E-FAM algorithm compared favorably to the ones obtained by other classifiers for these problems in the recent literature [27,28]. Finally, we finish this paper with some concluding remarks in Section 6.

2. Mathematical background

In 1965, Lotfi A. Zadeh extended the classical notion of a set, also referred to as a crisp set, by introducing fuzzy sets [29]. A fuzzy set A consists of a set X, called universe, together with a membership function $\mu_A : X \to [0, 1]$ that yields the membership degree $\mu_A(x) \in [0, 1]$ for each $x \in X$. The symbol $\mathcal{F}(X)$ denotes the class of fuzzy sets on the universe X.

Mathematically speaking, a fuzzy set A on a universe X can be identified with its membership function μ_A and can simply be viewed as a function from the universe X to the unit interval $[0, 1]$. If X is finite, say $X = \{x_1, \dots, x_n\}$, then $\mathcal{F}(X)$ can be identified with $[0, 1]^n$ via the bijection that maps $A \in \mathcal{F}(X)$ to the vector $(A(x_1), \dots, A(x_n))^t \in [0, 1]^n$. Note that the notion of a crisp set arises if A (or, more precisely, μ_A) only adopts values in $\{0, 1\}$. In particular, the classical concept of a relation is given by a subset of $X \times Y$, where X and Y are arbitrary universes, and can thus be viewed as a function $X \times Y \to \{0, 1\}$. Extending this classical concept to the fuzzy domain, a fuzzy relation is given by a function $R : X \times Y \to [0, 1]$, where $R(x, y)$ can be interpreted as the degree of relationship between x and y [30]. In other words, a fuzzy relation on $X \times Y$ is nothing else than a fuzzy set on $X \times Y$.

Two fuzzy relations $R \in \mathcal{F}(X \times Y)$ and $S \in \mathcal{F}(Y \times Z)$, where X, Y, and Z are arbitrary universes, can be combined using a fuzzy relational composition that yields a fuzzy relation on the universe $X \times Z$. In this paper, we only employ fuzzy relational compositions of the type sup-t defined in Eq. (1) below. Here, t stands for a t-norm (an associative, commutative, and increasing mapping $t : [0, 1]^2 \to [0, 1]$ with identity element 1).

Goguen provided a generalization of the notions of crisp and fuzzy sets by defining an $\mathbb{L}$-fuzzy set as a function $X \to \mathbb{L}$, where $\mathbb{L}$ is usually required to be a partially ordered set, a lattice, or a complete lattice [31,32]. Let us briefly review these concepts in the following.

A pair $(P, \leq)$ consisting of a non-empty set P together with a reflexive, antisymmetric, and transitive binary relation "$\leq$" is called a *partially ordered set* or *poset* [33]. If the partial order relation $\leq$ clearly arises from the context, then we simply refer to the poset under consideration using the symbol P instead of $(P, \leq)$. If $X \subseteq P$, then an element $l \in P$ is said to be a *lower bound* of X if $l \leq x$ for all $x \in X$. Similarly, $u \in P$ is said to be an *upper bound* of X if $x \leq u$ for all $x \in X$. The *infimum* of $X \subseteq P$, denoted using the symbol $\bigwedge X$, is defined as the greatest lower bound of X. Similarly, the *supremum* of $X \subseteq P$, denoted using the symbol $\bigvee X$, is defined as the least upper bound of X. If $X = \{x, y\}$, then we may alternatively write $x \vee y$ and $x \wedge y$ instead of $\bigvee X$ and $\bigwedge X$, respectively.

A partially ordered set $\mathbb{L}$ is called a *lattice* if every finite, non-empty subset of $\mathbb{L}$ has an infimum and a supremum in $\mathbb{L}$ [31]. If we additionally have that $\bigwedge \mathbb{L} \in \mathbb{L}$ and $\bigvee \mathbb{L} \in \mathbb{L}$, then the lattice $\mathbb{L}$ is said to be *bounded*. In this case, the symbols $0_{\mathbb{L}}$ and $1_{\mathbb{L}}$ denote respectively $\bigwedge \mathbb{L}$ and $\bigvee \mathbb{L}$. A lattice $\mathbb{L}$ is *complete* if every subset of $\mathbb{L}$ has an infimum and a supremum in $\mathbb{L}$. In particular, every complete lattice is bounded [31].

A lattice $(\mathbb{L}, \leq)$ is called a *chain* if we have $x \leq y$ or $y \leq x$ for all $x, y \in \mathbb{L}$. In this case, one refers to $\leq$ as a *total order*. If $\mathbb{L}$ is additionally a complete lattice, then one speaks of a *complete chain*. The unit interval $[0, 1]$ and the set of the extended real numbers $\mathbb{R}_{\pm\infty} = \mathbb{R} \cup \{+\infty, -\infty\}$ together with the usual total orders represent examples of complete chains. Thus, the supremum of $r, s \in [0, 1]$ is s if and only if $r \leq s$. Similarly, $r \vee s = r$ if and only if $s \leq r$.

If t is a t-norm, then the sup-t composition of $R \in \mathcal{F}(X \times Y)$ and $S \in \mathcal{F}(Y \times Z)$ yields the following fuzzy relation $R \circ_t S \in \mathcal{F}(X \times Z)$:

$$(R \circ_t S)(x, z) = \bigvee_{y \in Y} R(x, y)\, t\, S(y, z) \quad \forall x \in X \text{ and } z \in Z. \tag{1}$$

Similarly, the sup-t composition of $R \in \mathcal{F}(X \times Y)$ and $S \in \mathcal{F}(Y)$ yields the following fuzzy set $R \circ_t S \in \mathcal{F}(X)$:

$$(R \circ_t S)(x) = \bigvee_{y \in Y} R(x, y)\, t\, S(y) \quad \forall x \in X. \tag{2}$$

If $(\mathbb{L}_1, \leq_1), \ldots, (\mathbb{L}_n, \leq_n)$ are posets, then a poset $\mathbb{L} = \mathbb{L}_1 \times \ldots \times \mathbb{L}_n$ arises by defining $\mathbf{x} \leq \mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{L}$ as follows:

$$\mathbf{x} \leq \mathbf{y} \Leftrightarrow x_i \leq_i y_i, \forall i = 1, \ldots, n. \tag{3}$$

If $\mathbb{L}_i$ are lattices, bounded lattices, or complete lattices for all $i = 1, \ldots, n$, then $\mathbb{L} = \mathbb{L}_1 \times \ldots \times \mathbb{L}_n$ also represents respectively a lattice, a bounded lattice, or a complete lattice and is called the product lattice (of $\mathbb{L}_1, \ldots, \mathbb{L}_n$). If $\mathbb{L}_i = \mathbb{M}$ for all $i = 1, \ldots, n$, then the symbol $\mathbb{M}^n$ stands for $\mathbb{L}_1 \times \ldots \times \mathbb{L}_n$. To simplify our notation, we denote $(\mathbb{R}_{\pm\infty})^n$ using the symbol $\mathbb{R}^n_{\pm\infty}$.

Let $\mathbb{L}^X$ denote the class of $\mathbb{L}$-fuzzy sets on an arbitrary universe X, i.e., the class of functions $X \to \mathbb{L}$. If $\mathbb{L}$ represents a poset then the partial ordering $\leq$ on $\mathbb{L}$ induces a partial ordering on $\mathbb{L}^X$ that is denoted using the same symbol $\leq$ and that is defined as follows for all $f, g \in \mathbb{L}^X$:

$$f \leq g \Leftrightarrow f(x) \leq g(x) \quad \forall x \in X. \tag{4}$$

The lattice structure on $\mathbb{L}$ induces the same lattice structure on $\mathbb{L}^X$. For example, if $\mathbb{L}$ constitutes a bounded or a complete lattice then $\mathbb{L}^X$ constitutes respectively a bounded or a complete lattice as well. Consider for instance the complete lattice $[0, 1]$. The class of fuzzy sets over the universe X, denoted using the symbol $\mathcal{F}(X)$, is given by the complete lattice $[0, 1]^X$.

Extending Fodor's and Roubens' definition of equivalence [34,35] as a binary operation on $[0, 1]$, Bustince et al. defined the notion of equivalence measure as a function $\mathcal{F}(X) \times \mathcal{F}(X) \to [0, 1]$ [36]. Here, we propose a further extension of this notion by substituting the complete lattice $\mathcal{F}(X)$ with any bounded lattice.

Definition 1. Let $\mathbb{L}$ be a bounded lattice. An equivalence measure on $\mathbb{L}$ is a function $E : \mathbb{L}^2 \to [0, 1]$ that satisfies the following conditions:

E1) $E(x, y) = E(y, x)$ for all $x, y \in \mathbb{L}$;
E2) $E(0_{\mathbb{L}}, 1_{\mathbb{L}}) = 0$;
E3) $E(x, x) = 1$ for all $x \in \mathbb{L}$;
E4) if $x \leq y \leq z$, then $E(x, z) \leq E(x, y)$ and $E(x, z) \leq E(y, z)$.

In the literature of fuzzy set theory, one can find two other concepts that are closely related to the one of (fuzzy) equivalence, namely fuzzy similarity measure and equality index [36–42]. In fact, these technical terms are either used interchangeably or their definitions differ with respect to slight modifications of Definition 1 (several of these differences were addressed in [36]). The main difference between the concepts of similarity and equivalence seems to be the requirement that the similarity between a crisp set and its complement should be zero [40,41]. This property is crucial for the definition of entropy or fuzziness measures by means of similarity measure since it is reasonable to assume that the entropy of a crisp set equals 0 [36,40–42]. Fuzzy equivalence and similarity measures have been extensively studied in the last few years and many authors have proposed formulas to produce fuzzy equivalence measures on the class of fuzzy sets over a universe U [36,39–41]. In [43], a mapping $E : \mathbb{R}^2 \to [0, 1]$ satisfying the conditions E1, E2 and E3 is called an *order compatible fuzzy relation* (OCFR) and E4 is called "compatibility with total order". The latter property (E4) distinguishes OCFRs from fuzzy relations that generalize the notion of equivalence relation of classical set theory by requiring t-transitivity.

An infinite number of equivalence measures on $\mathcal{F}(X)$ can be generated from fuzzy subsethood or inclusion measures [1,36,41,42,44]. The concept of equivalence measure on $\mathcal{F}(X)$ can also be linked to the ones of fuzzy entropy and distance measures [39–41].

Depending on the kind of application, additional properties may be desirable. In the special case where $\mathbb{L} = \mathcal{F}(X)$, the next definition yields Zeng's & Li's definition of a similarity measure [42]. Since there are several different definitions of similarity measure in the literature [36–38,40,42], we prefer to speak of a *strong equivalence* measure. Note that Definition 2 arises from Definition 1 by imposing an additional condition.

Definition 2. An equivalence measure E is said to be *strong* if $E(x, y) = 1 \Leftrightarrow x = y$.

Note that a restricted equivalence function [45] represents a special case of a strong equivalence measure $[0, 1]^2 \to [0, 1]$.

Example 1. Let $[a, b] \subset \mathbb{R}$, where $a < b$. For each $\lambda \in (0, 1]$, the function $E_\lambda : [a, b]^2 \to [0, 1]$ given by

$$E_\lambda(x, y) = \max\left(0,\ 1 - \frac{|x-y|}{\lambda|b-a|}\right) \quad \forall x, y \in [a, b], \tag{5}$$

yields a strong equivalence measure on $[a, b]$ [46].

Example 2. Let $\mathbb{R}_{\pm\infty}$ be the set of the extended real numbers, i.e., $\mathbb{R}_{\pm\infty} = \mathbb{R} \cup \{-\infty, +\infty\}$, and let σ be a real nonzero number, i.e., $\sigma \neq 0$. The function $G_\sigma : \mathbb{R}^2_{\pm\infty} \to [0, 1]$ defined as

$$G_\sigma(x, y) = \begin{cases} 1 & \text{if } x = y, \\ \exp(-\frac{(x-y)^2}{\sigma^2}) & \text{if } x, y \in \mathbb{R} \text{ and } x \neq y, \\ 0 & \text{otherwise}, \end{cases} \tag{6}$$

represents a strong equivalence measure on $\mathbb{R}_{\pm\infty}$.

Example 3. Let $\bar{\mathbb{R}}_{\geq 0}$ be the set of non-negative extended real numbers, i.e., $\bar{\mathbb{R}}_{\geq 0} = \{x \in \mathbb{R} \cup \{+\infty\} \mid x \geq 0\}$, and let σ be a positive real number. The function $H_\sigma : \bar{\mathbb{R}}^2_{\geq 0} \to [0, 1]$ defined as

$$H_\sigma(x, y) = \begin{cases} 1 & \text{if } x = y, \\ \exp(-\sigma|x-y|) & \text{if } x, y \in \mathbb{R} \text{ and } x \neq y, \\ 0 & \text{otherwise}, \end{cases} \tag{7}$$

represents a strong equivalence measure on $\bar{\mathbb{R}}_{\geq 0}$.

Recall that *an aggregation function* is an increasing function M from $[0, 1]^n$ to $[0, 1]$ that satisfies $M(0, \dots, 0) = 0$ and $M(1, \dots, 1) = 1$ [47]. A simple example of an aggregation function is given by the *weighted mean* $M_{\mathbf{w}} : [0, 1]^n \to [0, 1]$ that is defined as follows:

$$M_{\mathbf{w}}(a_1, \dots, a_n) = \sum_{i=1}^{n} w_i a_i, \tag{8}$$

where $\mathbf{w} \in [0, 1]^n$ with $\sum_{j=1}^{n} w_j = 1$.

Given equivalence measures on bounded lattices $\mathbb{L}_1, \dots, \mathbb{L}_n$, aggregation functions can be used to generate an equivalence measure on the product lattice $\mathbb{L}_1 \times .... \times \mathbb{L}_n$ as stated in the next proposition.

Proposition 1. *Let E_i be equivalence measures that are defined on bounded lattices $(\mathbb{L}_i, \leq_i)$ for $i = 1, \dots, n$ and let M be an aggregation function. If $\mathbb{L}$ denotes the product lattice $\mathbb{L}_1 \times \dots \times \mathbb{L}_n$, then an equivalence measure on $\mathbb{L}$ is given by the following function $E : \mathbb{L}^2 \to [0, 1]$:*

$$E(\mathbf{x}, \mathbf{y}) = M(E_1(x_1, y_1), \dots, E_n(x_n, y_n)) \; \forall \mathbf{x}, \mathbf{y} \in \mathbb{L}. \tag{9}$$

Moreover, if E_i are strong *equivalence measures for $i = 1, \dots, n$ and if M satisfies $M(x_1, \dots, x_n) = 1 \Leftrightarrow x_i = 1, \forall i = 1, \dots, n$, then E is also a* strong *equivalence measure.*

Proof. Evident. □

Suppose that the aggregation function M additionally satisfies the following conditions: $M(x_1, \dots, x_n) = 0$ implies that $x_1 = x_2 = \dots = x_n = 0$ and $M(x_1, \dots, x_n) = 1$ implies that $x_1 = x_2 = \dots = x_n = 1$. According to Bustince et al. [45], Eq. (9) yields a similarity measure in the sense of Liu [40] if each E_i is a restricted equivalence function from $[0, 1]^2$ to $[0, 1]$.

Example 4. Let $\mathbf{w} \in [0, 1]^n$ such that $\sum_{i=1}^{n} w_i = 1$ and let $M_{\mathbf{w}}$ be the weighted mean defined in Eq. (8). Proposition 1 implies that n equivalence measures E_i on bounded lattices $\mathbb{L}_i$ give rise to the following equivalence measure $E_{\mathbf{w}}$ on the product lattice $\mathbb{L} = \mathbb{L}_1 \times \dots \times \mathbb{L}_n$:

$$E_{\mathbf{w}}(\mathbf{x}, \mathbf{y}) = M_{\mathbf{w}}(E_1(x_1, y_1), \dots, E_n(x_n, y_n)) = \sum_{i=1}^{n} w_i E_i(x_i, y_i) \; \forall \mathbf{x}, \mathbf{y} \in \mathbb{L}. \tag{10}$$

In particular, we obtain the following equivalence measures:

a)

$$E_{\lambda,\mathbf{w}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} w_i E_{\lambda_i}(x_i, y_i) \quad \forall \mathbf{x}, \mathbf{y} \in [\mathbf{a}, \mathbf{b}], \tag{11}$$

where $[\mathbf{a}, \mathbf{b}] \subset \mathbb{R}^n$ with $\mathbf{a} \leq \mathbf{b}$ and $\lambda \in (0, 1]^n$;

b)

$$G_{\sigma,\mathbf{w}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} w_i G_{\sigma_i}(x_i, y_i) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n_{\pm\infty}, \tag{12}$$

where $\sigma \in \mathbb{R}^n$ such that $\sigma_i \neq 0$ for $i = 1, \dots, n$;

c)

$$H_{\sigma,\mathbf{w}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} w_i H_{\sigma_i}(x_i, y_i) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n_{\geq 0}, \tag{13}$$

where $\sigma \in \mathbb{R}^n$ such that $\sigma_i \neq 0$ for $i = 1, \dots, n$.

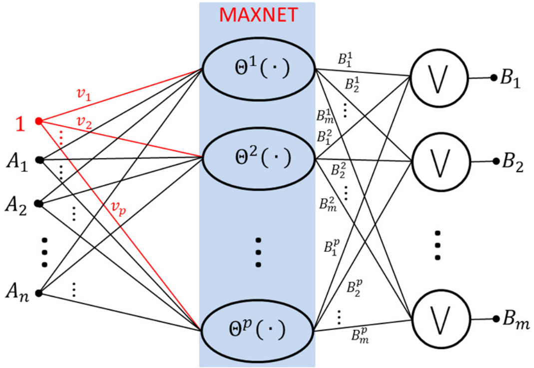

Fig. 1. Topology of a Θ-FAM.

Parametrized equivalence measures such as the ones given in Eqs. (10)–(13) can serve as functions Θ^ξ in tunable E-FAMs that are introduced in the next section. In this paper, we focus on equivalence measures where the weighted mean $M_{\mathbf{w}}$ plays the role of the aggregation functions M of Eq. (9). This type of equivalence measure has been employed by Martins-Bedé et al. in fuzzy case-based reasoning [46] and by Drummond et al. to restore consistency in fuzzy rule-based systems [48]. A particular case of an OCFR, proposed in [43], yields another example of a parametrized equivalence measure.

3. Introduction to tunable E-FAMs in the context of Θ-FAMs

The purpose of any associative memory (AM) is to store a set of associations $\mathcal{M} = \{(\mathbf{x}^\xi, \mathbf{y}^\xi) \mid \xi = 1, \ldots, p\}$, also known as the *fundamental memory set* or the set of *fundamental memories*, such that a desired output pattern $\mathbf{y}$ can be retrieved upon presentation of a possibly noisy or incomplete version of an input pattern $\mathbf{x}$. Formally speaking, an AM describes a mapping Φ that is geared to associating $\mathbf{x}^\xi$ with $\mathbf{y}^\xi$ for all $\xi = 1, \ldots, p$. Ideally, for each $\xi = 1, \ldots, p$, we have that Φ satisfies $\Phi(\mathbf{x}^\xi) = \mathbf{y}^\xi$ and additionally $\Phi(\tilde{\mathbf{x}}^\xi) = \mathbf{y}^\xi$ for corrupted or noisy versions of $\mathbf{x}^\xi$. In practice, many AM models are only able to store a subset of $\mathcal{M}$ and corrupted or noisy versions of $\mathbf{x}^\xi$ can only be retrieved approximately, i.e., $\Phi(\tilde{\mathbf{x}}^\xi) \simeq \mathbf{y}^\xi$. If an AM represents a fuzzy neural network [18], i.e., an artificial neural network whose inputs or weights are fuzzy, then we speak of a *fuzzy associative memory* (FAM).

A particular type of FAM models is given by the *tunable equivalence fuzzy associative memories* (tunable E-FAMs or simply TE-FAMs) that we introduce in this section. Let us first present the definition of a related model called Θ-fuzzy associative memory that motivated the present work.

3.1. A brief review and a generalization of Θ-fuzzy associative memories

The concept of a Θ-fuzzy associative memory (Θ-FAMs) was recently introduced in the literature as a fuzzy associative memory (FAM) having a competitive hidden layer. Let us briefly recall its original definition [1].

Suppose that we are provided with a fundamental memory set $\{(A^\xi, B^\xi) \in \mathcal{F}(X) \times \mathcal{F}(Y) : \xi = 1, \ldots, p\}$, where X and Y are arbitrary universes. Furthermore, let $\Theta^\xi : \mathcal{F}(X) \to [0, 1]$ be operators such that $\Theta^\xi(A^\xi) = 1$ for $\xi = 1, \ldots, p$ and let $\mathbf{v}$ be an arbitrary vector in $\mathbb{R}^p$. The *Θ-FAM based on* Θ^ξ *and* $\mathbf{v}$ produces the following output $\mathcal{O}(A) \in \mathcal{F}(Y)$ upon presentation of an input fuzzy set $A \in \mathcal{F}(X)$:

$$\mathcal{O}(A) = \bigcup_{j \in I_{\mathbf{v}}(A)} B^j, \tag{14}$$

where

$$I_{\mathbf{v}}(A) = \left\{ j \in \{1, \ldots, p\} : v_j \Theta^j(A) = \max_{\xi=1,\ldots,p} v_\xi \Theta^\xi(A) \right\}. \tag{15}$$

Fig. 1 displays the topology of a Θ-FAM for finite universes $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$. Note that the ξth node computes $(v_\xi \cdot 1) \cdot \Theta^\xi(A)$.

Esmi et al. provided theoretical results concerning the storage capacities as well as the error correction capabilities of Θ-FAM models [1]. In particular, sufficient conditions for the perfect recall of all fundamental memories and a characterization of the basins of attraction around each A^ξ can be found in [1]. The weight vector $\mathbf{v}$ can be optimized using a supervised training algorithm that was specifically designed for Θ-FAMs. Under some weak conditions, this supervised algorithm is guaranteed to converge to a local minimum of the objective function in a finite number of steps, but the computational cost can be extremely high for large datasets. This drawback motivated us to develop a new supervised algorithm for a particular class of Θ-FAMs (see Section 4), that includes a procedure for extracting a fundamental memory set from the training data.

Operators Θ^ξ that satisfy $\Theta^\xi(A^\xi) = 1$ for $\xi = 1, \dots, p$ can be defined in terms of fuzzy subsethood measures as well as fuzzy similarity (or equivalence) measures [1]. In particular, if we have that each operator Θ^ξ is given by $\Theta^\xi(\cdot) = SM(A^\xi, \cdot)$ for some similarity measure SM on $\mathcal{F}(X)$ in the original definition above [49], then we speak of a *weighted similarity measure fuzzy associative memory*, for short a weighted SM-FAM [1,2,44].

Before introducing tunable E-FAMs, we will extend the original definition of a Θ-FAM to deal with inputs in any bounded lattice. Despite this generalization, we may still speak of fuzzy associative memories, i.e., a particular type of fuzzy neural networks, because Θ-FAMs with inputs in an arbitrary bounded lattice have fuzzy weights [18].

First, note that the subset $I_\mathbf{v}(A)$ of $\{1, \dots, p\}$ can be uniquely identified with a fuzzy set $Z^A \in [0,1]^p$ whose membership function corresponds to its indicator function (i.e., $Z^A_i = 1$ if $i \in I_\mathbf{v}(A)$ and $Z^A_i = 0$, otherwise) for $A \in \mathcal{F}(X)$. Moreover, if $F : \mathbb{R}^p \to [0,1]^p$ is such that

$$F(a_1, \dots, a_p)_i = \begin{cases} 1 & \text{if } a_i = \bigvee_{j=1}^p a_j, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for } i = 1, \dots, p, \tag{16}$$

for all $\mathbf{a} \in \mathbb{R}^p$, then we have that $Z^A = F(v_1\Theta^1(A), \dots, v_p\Theta^p(A))$ for all $A \in \mathcal{F}(X)$. Based on the last observation and the properties of a t-norm, we can rewrite Eq. (14) by means of a sup-t composition, where t is an arbitrary t-norm, as follows:

$$\mathcal{O}(A) = \bigcup_{j \in I_\mathbf{v}(A)} B^j = R \circ_t F(v_1\Theta^1(A), \dots, v_p\Theta^p(A)). \tag{17}$$

Here, R is the fuzzy relation on $Y \times \{1, \dots, p\}$ that is given by $R(y, \xi) = B^\xi(y)$ for all $y \in Y$ and $\xi \in \{1, \dots, p\}$.

In view of these remarks, we propose to slightly generalize the original definition of a Θ-FAM as follows:

Definition 3. Let $\mathcal{M} = \{(\mathbf{x}^\xi, B^\xi) \in \mathbb{L} \times \mathcal{F}(Y) \mid \xi \in P\}$ for a bounded lattice $\mathbb{L}$, an arbitrary universe Y, and a finite set of indices P, say $P = \{1, \dots, p\}$. Given operators $\Theta^\xi : \mathbb{L} \to [0,1]$ satisfying $\Theta^\xi(\mathbf{x}^\xi) = 1\ \forall \xi \in P$, a weight vector $\mathbf{v} \in \mathbb{R}^p$, a function $F : \mathbb{R}^p \to [0,1]^p$, and an arbitrary t-norm, the following mapping $\mathcal{O} : \mathbb{L} \to \mathcal{F}(Y)$ yields a Θ-fuzzy associative memory, for short, a Θ-FAM:

$$\mathcal{O}(\mathbf{x}) = R \circ_t F(v_1\Theta^1(\mathbf{x}), \dots, v_p\Theta^p(\mathbf{x})), \tag{18}$$

where R denotes the fuzzy relation in $Y \times P$ given by $R(y, \xi) = B^\xi(y)$ for all $y \in Y$ and $\xi \in P$.

3.2. Tunable E-FAMs

Note that if E^ξ denote equivalence measures on $\mathbb{L}$, then the operators $E^\xi(\cdot, \mathbf{x}^\xi)$ can play the role of Θ^ξ in Definition 3 since $\Theta^\xi(\mathbf{x}^\xi) = E^\xi(\mathbf{x}^\xi, \mathbf{x}^\xi) = 1$ by property E3) of Definition 1. Thus, the notion of weighted SM-FAM can be generalized as follows.

Definition 4. Consider a Θ-FAM given by Eq. (18) of Definition 3. If each function $\Theta^\xi : \mathbb{L} \to [0,1]$ of Definition 3 is given by $\Theta^\xi(\cdot) = E^\xi(\cdot, \mathbf{x}^\xi)$ for some equivalence measure E^ξ on $\mathbb{L}$, then the corresponding Θ-FAM is called an *equivalence fuzzy associative memory (E-FAM)*. Note that an E-FAM generates the following output $\mathcal{E}(\mathbf{x})$ for an input $\mathbf{x} \in \mathbb{L}$:

$$\mathcal{E}(\mathbf{x}) = R \circ_t F(v_1 E^1(\mathbf{x}, \mathbf{x}^1), \dots, v_p E^p(\mathbf{x}, \mathbf{x}^p)), \tag{19}$$

where R denotes the fuzzy relation in $Y \times P$ given by $R(y, \xi) = B^\xi(y)$ for all $y \in Y$ and $\xi \in P$.

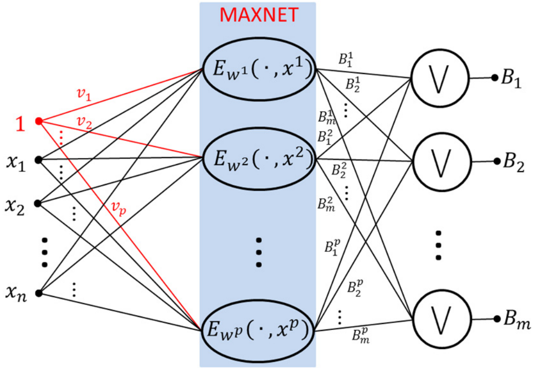

Fig. 2. Topology of a tunable E-FAM of type $\mathcal{E}_{W,\mathbf{v}}$ with F as in Eq. (16).

We speak of a *tunable equivalence fuzzy associative memory*, for short a *TE-FAM*, if each E^ξ equals a parametrized equivalence measure.

Consider the special case where $\mathbb{L}$ denotes a product of bounded lattices $\mathbb{L}_i$, each one of which being equipped with an equivalence measure E_i, where $i = 1, \ldots, n$. Throughout the rest of this manuscript we will concentrate on TE-FAMs based on $E_{\mathbf{w}^\xi}$ for some $\mathbf{w}^\xi \in [0,1]^n$ that satisfy $\sum_{i=1}^n w_i^\xi = 1$.

Let us gather $\mathbf{w}^1, \ldots, \mathbf{w}^p \in [0,1]^n$ as column vectors of a matrix $W = [\mathbf{w}^1, \ldots, \mathbf{w}^p] \in [0,1]^{n\times p}$ and denote the corresponding TE-FAM using the symbol $\mathcal{E}_{W,\mathbf{v}}$. Given an input pattern $\mathbf{x} \in \mathbb{L}$, the TE-FAM $\mathcal{E}_{W,\mathbf{v}}$ generates the following output $\mathcal{E}_{W,\mathbf{v}}(\mathbf{x}) \in \mathcal{F}(Y)$:

$$\begin{aligned} &\mathcal{E}_{W,\mathbf{v}}(\mathbf{x}) = B \circ_t F(v_1 E_{\mathbf{w}^1}(\mathbf{x}, \mathbf{x}^1), \ldots, v_p E_{\mathbf{w}^p}(\mathbf{x}, \mathbf{x}^p)) \\ &\quad \forall \mathbf{x} \in \mathbb{L} = \mathbb{L}_1 \times \ldots\ldots \mathbb{L}_n. \end{aligned} \tag{20}$$

Recall that the equivalence measures $E_{\mathbf{w}^\xi}$ are constructed on the product $\mathbb{L}$ of bounded lattices $\mathbb{L}_1, \ldots, \mathbb{L}_n$ according to the formula of Eq. (10). Thus, we incorporated another set of weights w_i^ξ, that permit adjusting the relevance of the ith attribute for the ξth hidden node in terms of $w_i^\xi E_i(\cdot, x_i^\xi)$. The weights v_ξ, $\xi = 1, \ldots, p$ are responsible for adjusting the contributions of the ξth hidden node.

Fig. 2 visualizes the topology of a TE-FAM of type $\mathcal{E}_{W,\mathbf{v}}$ for F as in Eq. (16) and a finite universe Y.

4. A new training algorithm for tunable E-FAMs

The Θ-FAM training algorithm presented in [1] can also be applied to tunable E-FAMs with fixed equivalence measures in the hidden nodes. At some stage of this previous Θ-FAM training algorithm, every training pattern is used as a fundamental memory. As mentioned before, this supervised algorithm is guaranteed to converge to a local minimum of the objective function in a finite number of steps under some weak conditions, but can be computationally expensive for large datasets.

Let us now introduce a training algorithm that is tailored towards tunable E-FAMs of type $\mathcal{E}_{W,\mathbf{v}}$ with competitive hidden nodes, in which case F is as in Eq. (16). Specifically, we will present a new supervised training algorithm called *Algorithm TE* that can be divided into two main stages:

1. Extraction of the fundamental memory set $\mathcal{M} = \{(\mathbf{x}^1, B^1), \ldots, (\mathbf{x}^p, B^p)\}$ from the training set $\mathcal{T}$;
2. Optimization of the weight vectors $\mathbf{w}^\xi$ and normalization of the weight vector $\mathbf{v}$.

Note that Stage 1 can be executed for any E-FAM and determines the exact topology of this two-layer feedforward fuzzy neural network since initially p, that is, the number of hidden nodes, and $\mathbf{x}^1, \ldots, \mathbf{x}^p$ are unknown. Stage 1 aims at generating a fundamental memory set $\mathcal{M}$ with $|\mathcal{M}| \leq |\mathcal{T}|$ in order to facilitate the optimization of the vectors $\mathbf{w}^\xi$ (if existent) and/or $\mathbf{v}$ based on the resulting fundamental memory set $\mathcal{M} \subseteq \mathcal{T}$. Let us now provide some details on the two individual stages.

Please cite this article in press as: E. Esmi et al., Tunable equivalence fuzzy associative memories, Fuzzy Sets and Systems (2015), http://dx.doi.org/10.1016/j.fss.2015.04.004

4.1. Selection of fundamental memories

Suppose that we are provided with a training set $\mathcal{T} = \{(\mathbf{x}^\xi, B^\xi) \in \mathbb{L} \times \mathcal{F}(Y), \xi = 1, \dots, k\}$. Let E^ξ be equivalence measures on a bounded lattice $\mathbb{L}$ for $\xi = 1, \dots, k$. Instead of presenting all elements of $\mathcal{T}$ as fundamental memories to the E-FAM based on E^ξ, where $\xi = 1, \dots, k$, we will select a subset $\mathcal{M}$ of $\mathcal{T}$ as a fundamental memory set of an E-FAM. The process of extracting fundamental memories from the entire training set $\mathcal{T}$ hinges on evaluating $E^\gamma(\mathbf{x}^\xi, \mathbf{x}^\gamma)$ for some $\xi \neq \gamma$ in order to ensure that $\mathcal{M}$ yields in some sense a good representation of $\mathcal{T}$. In particular, selecting a fundamental set $\mathcal{M}$ as a subset of $\mathcal{T}$ aims to improve the generalization capability of the E-FAM and to reduce the risk of overfitting. Let us now give some detailed explanations.

Let $e_{\xi,\gamma}$ symbolize $E^\gamma(\mathbf{x}^\xi, \mathbf{x}^\gamma)$ for $\xi, \gamma = 1, \dots, k$. If $(\mathbf{x}^\xi, B^\xi)$, $(\mathbf{x}^\gamma, B^\gamma) \in \mathcal{T}$ and $B^\xi = B^\gamma$, then $e_{\xi,\gamma}$ can be viewed as the potential, measured as a number in $[0, 1]$, of $\mathbf{x}^\gamma$ to replace $\mathbf{x}^\xi$. For each $(\mathbf{x}^\xi, B^\xi)$, we define $\bar{e}^\xi$ as the maximal equivalence or similarity of $\mathbf{x}^\xi$ and any $\mathbf{x}^\gamma$ such that $B^\gamma \neq B^\xi$. Formally, we have:

$$\bar{e}_\xi = \bigvee_{\gamma \,:\, B^\gamma \neq B^\xi} e_{\xi,\gamma}. \tag{21}$$

For every $\rho \in [0, 1)$ and every $\xi \in \{1, \dots, k\}$, we define $\mathcal{R}^\xi_\rho$ as the *set of indices γ such that $\mathbf{x}^\gamma$ can replace $\mathbf{x}^\xi$ with a potential larger than ρ* and such that

$$\bar{e}_\gamma \le \bar{e}_\xi < e_{\xi,\gamma}. \tag{22}$$

Formally, we have that

$$\mathcal{R}^\xi_\rho = \{\gamma \in \{1, \dots, k\} \mid B^\gamma = B^\xi,\ \bar{e}_\gamma \le \bar{e}_\xi < e_{\xi,\gamma},\ \text{and } \rho < e_{\xi,\gamma}\}. \tag{23}$$

If $\mathcal{C}^\xi$ denotes the set of $\mathbf{x}^\gamma$ such that $B^\gamma = B^\xi$ for all $\xi = 1, \dots, k$, then $\bar{e}_\gamma \le \bar{e}_\xi$ expresses the following situation: The maximal equivalence of $\mathbf{x}^\gamma$ and any $\mathbf{x}^\zeta \notin \mathcal{C}^\xi$ is at most as large as the maximal equivalence of $\mathbf{x}^\xi$ and any $\mathbf{x}^\zeta \notin \mathcal{C}^\xi$. The condition $\bar{e}_\xi < e_{\xi,\gamma}$ means that the maximal equivalence of $\mathbf{x}^\xi$ and any $\mathbf{x}^\zeta \notin \mathcal{C}^\xi$ is less than the degree of equivalence of $\mathbf{x}^\xi$ and $\mathbf{x}^\gamma$.

Note that the cardinality of $\mathcal{R}^\xi_\rho$ is decreasing with respect to ρ and that $\mathcal{R}^\xi_\rho$ is guaranteed to contain ξ if $\bar{e}_\xi < 1$. In particular, we have $\xi \in \mathcal{R}^\xi_\rho$ if E^γ is a strong equivalence measure and $\mathcal{C}^\xi \cap \mathcal{C}^\gamma = \emptyset$ for all $B^\xi \neq B^\gamma$.

Algorithm 1 below shows how to generate a fundamental memory set $\mathcal{M}$ of a tunable E-FAM as a subset of $\mathcal{T}$. More precisely, we have $\mathcal{M} = \bigcup_{\xi=1}^k \mathcal{M}^\xi$, where each $\mathcal{M}^\xi$ corresponds to a subset of $\mathcal{R}^\xi_\rho$. A brief glance at Algorithm 1 reveals that its complexity is bounded from above by $O(k^2)$.

Example 5. Let us illustrate how Algorithm 1 works by means of a binary classification problem that is determined by the small artificial training set $\mathcal{T}^\star = \{(\mathbf{x}^\xi, B^\xi) \mid \xi = 1, \dots, 6\} \subset \mathbb{R}^2_{\pm\infty} \times [0, 1]$ shown in Fig. 3. Specifically, we have

Algorithm 1 Extracting fundamental memories from the training set.

Input: Given $\mathcal{T} = \{(x^\xi, B^\xi) \in \mathbb{L} \times \mathcal{F}(Y), \xi = 1, \dots, k\}$;

Compute $e_{\xi,\gamma}$ for $\xi, \gamma = 1, \dots, k$;
Compute $\bar{e}_\xi$ for $\xi = 1, \dots, k$;
Define $\rho \in [0, 1)$;
Compute $\mathcal{R}^\xi_\rho$ for $\xi = 1, \dots, k$;
Define $\mathcal{S}^\gamma = \left\{\xi \in \{1, \dots, k\} \mid \gamma \in \mathcal{R}^\xi_\rho\right\}$ for $\gamma = 1, \dots, k$;
Compute $s_\gamma = \sum_{\xi \in \mathcal{S}^\gamma} e_{\xi,\gamma}$ for $\gamma = 1, \dots, k$;
Set $\mathcal{M} \leftarrow \emptyset$;
for $\xi = 1, \dots, k$ **do**
    for all $\gamma \in \mathcal{R}^\xi_\rho$ **do**
        Compute $v_{\xi,\gamma} \leftarrow e_{\xi,\gamma} s_\gamma$;
    end for
    Set $m_\xi \leftarrow \bigvee_{\zeta \in \mathcal{R}^\xi_\rho} v_{\xi,\zeta}$;
    For each $\gamma \in \mathcal{R}^\xi_\rho$ s.t. $v_{\xi,\gamma} = m_\xi$ update $\mathcal{M} \leftarrow \mathcal{M} \cup \left\{(\mathbf{x}^\gamma, B^\gamma)\right\}$;
end for

Output: $\mathcal{M}$.

Please cite this article in press as: E. Esmi et al., Tunable equivalence fuzzy associative memories, Fuzzy Sets and Systems (2015), http://dx.doi.org/10.1016/j.fss.2015.04.004

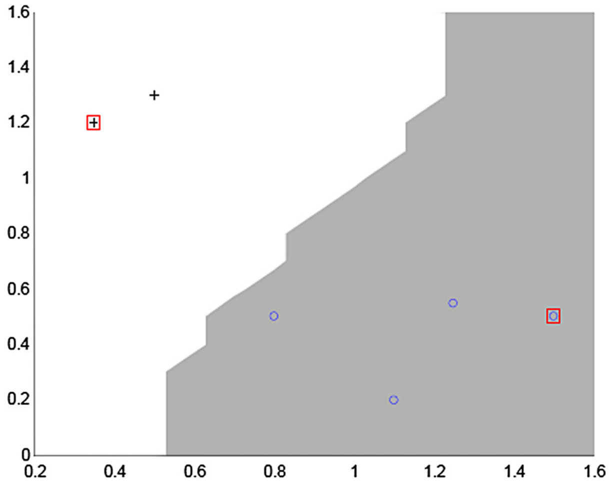

Fig. 3. The symbols $\circ$ and $+$ mark the locations of the patterns belonging to the classes "0" and "1", respectively. Algorithm 1 yields the set $\mathcal{M}$ consisting of two patterns whose locations are marked by red squares. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Results of Steps 4, 5, and 6 of Algorithm 1.

ξ	1	2	3	4	5	6
$\mathcal{R}_\rho^\xi$	$\{1,2\}$	$\{2\}$	$\{1,2,3\}$	$\{1,2,3,4\}$	$\{5,6\}$	$\{6\}$
$\mathcal{S}^\xi$	$\{1,3,4\}$	$\{1,2,3,4\}$	$\{3,4\}$	$\{4\}$	$\{5\}$	$\{5,6\}$
s_ξ	2.84	3.66	1.91	1.00	1.00	1.98

$$(\mathbf{x}^1, B^1) = \big((1.1, 0.2), 0\big),\ (\mathbf{x}^2, B^2) = \big((1.5, 0.5), 0\big),\ (\mathbf{x}^3, B^3) = \big((1.25, 0.55), 0\big),$$
$$(\mathbf{x}^4, B^4) = \big((0.8, 0.5), 0\big),\ (\mathbf{x}^5, B^5) = \big((0.5, 1.3), 1\big),\ (\mathbf{x}^6, B^6) = \big((0.35, 1.2), 1\big).$$

Let E^ξ for $\xi = 1, \ldots, 6$ be equal to the equivalence measure E on $\mathbb{R}^2_{\pm\infty}$ that is given by the mean of G_1, i.e., G_σ as defined in Eq. (6) with $\sigma = 1$. Thus we have $E^\xi(\mathbf{x}, \mathbf{y}) = G_{(1,1),(0.5,0.5)}(\mathbf{x}, \mathbf{y}) = 0.5(G_1(x_1, y_1) + G_1(x_2, y_2))$ for all $\xi = 1, \ldots, 6$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2_{\pm\infty}$.

Let $\mathbf{E} \in [0, 1]^{6\times 6}$ and $\bar{\mathbf{e}} \in [0, 1]^6$ be respectively the matrix and the vector whose entries $e_{i,j}$ and $\bar{e}_i$ are computed in Steps 1 and 2 of Algorithm 1:

$$\mathbf{E} = \begin{pmatrix} 1.00 & 0.88 & 0.93 & 0.91 & 0.50 & 0.47 \\ 0.88 & 1.00 & 0.97 & 0.81 & 0.45 & 0.44 \\ 0.93 & 0.97 & 1.00 & 0.91 & 0.57 & 0.55 \\ 0.91 & 0.81 & 0.91 & 1.00 & 0.72 & 0.71 \\ 0.50 & 0.45 & 0.57 & 0.72 & 1.00 & 0.98 \\ 0.47 & 0.44 & 0.55 & 0.71 & 0.98 & 1.00 \end{pmatrix} \quad \text{and} \quad \bar{\mathbf{e}} = \begin{pmatrix} 0.50 \\ 0.45 \\ 0.57 \\ 0.72 \\ 0.72 \\ 0.71 \end{pmatrix}.$$

Tables 1 and 2 show respectively the values computed in Steps 4–6 and 9–13 of Algorithm 1, with $\rho = 0$. Thus, Algorithm 1 produces $\mathcal{M} = \{\big((1.5, 0.5), 0\big), \big((0.35, 1.2), 1\big)\}$.

Given a set of associations $\mathcal{T} = \{(\mathbf{x}^\xi, B^\xi) : \xi = 1, \ldots, k\}$ and a tunable E-FAM $\mathcal{E}$, the next theorem shows that we can use the result $\mathcal{M} \subseteq \mathcal{T}$ produced by Algorithm 1 in order to achieve perfect recall of B^ξ upon presentation of $\mathbf{x}^\xi$ for all $\xi = 1, \ldots k$. More precisely, we have:

Theorem 1. *Let $\mathcal{T} = \{(\mathbf{x}^\xi, B^\xi) \in \mathbb{L} \times \mathcal{F}(Y), \xi = 1, \ldots, k\}$, where $\mathbb{L}$ is a bounded lattice. Suppose that E^γ are equivalence measures on $\mathbb{L}$ such that $E^\gamma(\mathbf{x}^\xi, \mathbf{x}^\gamma) < 1$ for all $B^\xi \neq B^\gamma$, where $\xi, \gamma = 1, \ldots, k$. If Algorithm 1 generates $\mathcal{M} = \{(\mathbf{x}^\xi, B^\xi), \xi = 1, \ldots, p\} \subseteq \mathcal{T}$, then the corresponding E-FAM $\mathcal{E}$ (cf. Eq. (19)) with competitive hidden nodes and*

Table 2
Results of Steps 9–13 of Algorithm 1.

ξ	$\{v_{\xi,\gamma} \mid \gamma \in \mathcal{R}_\rho^\xi\}$	m_ξ	Update $\mathcal{M}$
1	$\{v_{1,1} = 2.84, \mathbf{v_{1,2} = 3.22}\}$	3.22	$\mathcal{M} \leftarrow \mathcal{M} \cup \{(\mathbf{x}^2, 0)\}$
2	$\{\mathbf{v_{2,2} = 3.66}\}$	3.66	$\mathcal{M} \leftarrow \mathcal{M} \cup \{(\mathbf{x}^2, 0)\}$
3	$\{v_{3,1} = 2.64, \mathbf{v_{3,2} = 3.55}, v_{3,3} = 1.91\}$	3.55	$\mathcal{M} \leftarrow \mathcal{M} \cup \{(\mathbf{x}^2, 0)\}$
4	$\{v_{4,1} = 2.58, \mathbf{v_{4,2} = 2.96}, v_{4,3} = 1.74, v_{4,4} = 1.00\}$	2.96	$\mathcal{M} \leftarrow \mathcal{M} \cup \{(\mathbf{x}^2, 0)\}$
5	$\{v_{5,5} = 1.00, \mathbf{v_{5,6} = 1.94}\}$	1.94	$\mathcal{M} \leftarrow \mathcal{M} \cup \{(\mathbf{x}^6, 1)\}$
5	$\{\mathbf{v_{6,6} = 1.98}\}$	1.98	$\mathcal{M} \leftarrow \mathcal{M} \cup \{(\mathbf{x}^6, 1)\}$

weight vector $\mathbf{v} = (1, \dots, 1) \in \mathbb{R}^p$ perfectly retrieves B^ξ upon presentation of the pattern cues $\mathbf{x}^\xi$ for all $\xi = 1, \dots, k$, i.e.,

$$\mathcal{E}(\mathbf{x}^\xi) = B^\xi \quad \forall \xi = 1, \dots, k. \tag{24}$$

Proof. Since $e_{\xi,\gamma} = E^\gamma(\mathbf{x}^\xi, \mathbf{x}^\gamma) < 1$ for all $\xi, \gamma = 1, \dots, k$ such that $B^\xi \neq B^\gamma$, we have that $\bar{e}_\xi < 1$ for all $\xi = 1, \dots, k$ (cf. Eq. (21)). Thus, we have that $\xi \in \mathcal{R}_\rho^\xi$ for $\xi = 1, \dots, k$ because $\bar{e}_\xi \vee \rho < 1 = E^\xi(x^\xi, x^\xi) = e_{\xi,\xi}$ and $\bar{e}_\xi \leq \bar{e}_\xi$. Hence $\gamma \in \mathcal{S}^\gamma$ for all $\gamma = 1, \dots, k$. Therefore, $\mathcal{R}_\rho^\xi$ and $\mathcal{S}^\gamma$ are not empty for all $\xi, \gamma = 1, \dots, k$ which implies that, for all $\gamma, \xi = 1, \dots k$, the quantities s_γ as well as $v_{\xi,\zeta}$ for $\zeta \in \mathcal{R}_\rho^\xi$ are well defined.

Moreover, since $1 \leq |\mathcal{R}_\rho^\xi| \leq k$ we have that there exists an index ζ in $\mathcal{R}_\rho^\xi$ such that $m_\xi = v_{\xi,\zeta}$. A brief glance at Step 13 of Algorithm 1 reveals that $(\mathbf{x}^\zeta, B^\zeta) \in \mathcal{M}$. By the definition of $\mathcal{R}_\rho^\xi$, we have that $B^\zeta = B^\xi$ and $e_{\xi,\zeta} > \bar{e}_\xi$. Thus, for any association $(\mathbf{x}^\gamma, B^\gamma) \in \mathcal{M}$ such that

$$E^\gamma(\mathbf{x}^\xi, \mathbf{x}^\gamma) = \bigvee_{\mathbf{x}^\kappa : (\mathbf{x}^\kappa, B^\kappa) \in \mathcal{M}} E^\kappa(\mathbf{x}^\xi, \mathbf{x}^\kappa), \tag{25}$$

we have that $B^\gamma = B^\xi$ since

$$\bigvee_{\mathbf{x}^\kappa : (\mathbf{x}^\kappa, B^\kappa) \in \mathcal{M}} E^\kappa(\mathbf{x}^\xi, \mathbf{x}^\kappa) \geq e_{\xi,\zeta} > \bar{e}_\xi. \tag{26}$$

By the definition of $\mathcal{E}$ in terms of F given in Eq. (16), $\mathbf{v}$, and the sup-t composition (cf. Eq. (19)), we have that $\mathcal{E}(\mathbf{x}^\xi) = B^\xi$. □

Corollary 1. *Let $\mathcal{T} = \{\{(\mathbf{x}^\xi, B^\xi) \in \mathbb{L} \times \mathcal{F}(Y), \xi = 1, \dots, k\}$ such that $\mathbf{x}^\xi \neq \mathbf{x}^\gamma$ for all $\xi \neq \gamma$, where $\mathbb{L}$ is a bounded lattice. Suppose that E^ξ is a* strong *equivalence measure on $\mathbb{L}$ for every $\xi = 1, \dots, k$. If Algorithm 1 generates $\mathcal{M} \subseteq \mathcal{T}$, then the respective tunable E-FAM $\mathcal{E}$ with weight vector $\mathbf{v} = (1, \dots, 1) \in \mathbb{R}^{|\mathcal{M}|}$ perfectly retrieves B^ξ upon presentation of the pattern cues $\mathbf{x}^\xi$ for all $\xi = 1, \dots, k$, i.e.,*

$$\mathcal{E}(\mathbf{x}^\xi) = B^\xi \quad \forall \xi = 1, \dots, k. \tag{27}$$

Proof. Since E^ξ is a strong equivalence measure and $\mathbf{x}^\xi \neq \mathbf{x}^\gamma$ for $\xi \neq \gamma$, we have that $E^\gamma(\mathbf{x}^\xi, \mathbf{x}^\gamma) < 1$ for all $\xi \neq \gamma$. In particular, $E^\gamma(\mathbf{x}^\xi, \mathbf{x}^\gamma) < 1$ for all $\xi, \gamma = 1, \dots, k$ such that $B^\xi \neq B^\gamma$. Thus, Theorem 1 can be applied to prove the claim of corollary. □

4.2. A supervised training algorithm for tunable E-FAMs

Let us address the problem of adjusting the parameters of a tunable E-FAM. Consider a tunable E-FAM with competitive hidden nodes of the form $\mathcal{E}_{W,\mathbf{v}}$ depicted in Fig. 2 and suppose that we have already chosen a fundamental memory set $\mathcal{M} = \{(\mathbf{x}^\xi, B^\xi) \in \mathbb{L} \times \mathcal{F}(Y) \mid \xi = 1, \dots, p\}$. This model can deal with inputs from the product lattice $\mathbb{L} = \mathbb{L}_1 \times \dots \times \mathbb{L}_n$, where $\mathbb{L}_i$ are bounded lattices for $i = 1, \dots, n$. Recall that this model depends on the choice of vectors $\mathbf{w}^1, \dots, \mathbf{w}^p \in \mathbb{R}^n_{\geq 0}$, compiled in a matrix $W = [\mathbf{w}^1, \dots, \mathbf{w}^p] \in \mathbb{R}^{n \times p}_{\geq 0}$, and $\mathbf{v} \in \mathbb{R}^p$. The competition that takes place in the hidden layer can be expressed in terms of the application of the function F in Eq. (16). Unfortunately, the function F as well as the sup-t composition are not differentiable which complicates the optimization of the parameters W and $\mathbf{v}$ in Eq. (20). Let us now investigate how to overcome these problems.

First note that, by the definition of F, we have that the set of fixed points of F is $\{0,1\}^p \setminus \{(0,\dots,0)\}$. Moreover, if $\tilde{\mathbf{z}}^\xi = (v_1 E_{\mathbf{w}^1}(\mathbf{x}^\xi,\mathbf{x}^1),\dots,v_p E_{\mathbf{w}^p}(\mathbf{x}^\xi,\mathbf{x}^p))$ is a fixed point of F such that $\tilde{z}_i^\xi = 0$ for all $B^\xi \neq B^i$, then we obviously have that $\mathcal{E}_{W,\mathbf{v}}(\mathbf{x}^\xi) = B^\xi$. Since the last condition is sufficient for the perfect recall of B^ξ upon presentation of $\mathbf{x}^\xi$ to the memory $\mathcal{E}_{W,\mathbf{v}}$, we can formulate an objective function that aims to find W and $\mathbf{v}$ such that $\tilde{z}^\xi$ is a fixed point where $\tilde{z}_i^\xi = 0$ if and only if $B^\xi \neq B^i$. Note that the condition $\tilde{z}_i^\xi = 1$ for $B^\xi = B^i$ can be seen as an attempt to distribute the information of fundamental memories having the same consequent part in $\mathcal{F}(Y)$ over more than one hidden neuron. Based on these observations, we propose to minimize the following objective function:

$$\sum_{\xi=1}^{p}\left(\frac{1}{d_\xi}\sum_{\gamma:B^\gamma=B^\xi}(1-v_\xi E_{\mathbf{w}^\xi}(\mathbf{x}^\gamma,\mathbf{x}^\xi))^2 + \left(1-\frac{1}{d_\xi}\right)\sum_{\gamma:B^\gamma\neq B^\xi}(v_\xi E_{\mathbf{w}^\xi}(\mathbf{x}^\gamma,\mathbf{x}^\xi))^2\right), \tag{28}$$

where $d_\xi = |\{\kappa : B^\kappa = B^\xi\}|$. Defining $u_i^\xi = v_\xi w_i^\xi$ for $\xi = 1,\dots,p$ and $i = 1,\dots,n$, we can express $v_\xi E_{\mathbf{w}^\xi}(\mathbf{x},\mathbf{x}^\xi) = \sum_{i=1}^n v_\xi w_i^\xi E_i(x_i,x_i^\xi)$ in terms of the inner product of the vectors $\mathbf{a}^\xi(\mathbf{x}) = (E_1(x_1,x_1^\xi),\dots,E_n(x_n,x_n^\xi))^t$ and $\mathbf{u}^\xi = (u_1^\xi,\dots,u_n^\xi)^t$, i.e., $v_\xi E_{\mathbf{w}^\xi}(\mathbf{x},\mathbf{x}^\xi) = (\mathbf{a}^\xi(\mathbf{x}))^t\mathbf{u}^\xi$. Furthermore, let $\mathbf{z}^\xi \in \{0,1\}^p$ be such that $z_i^\xi = 0$ if only if $B^\xi \neq B^i$ for all $\xi = 1,\dots,p$ and $D^\xi = \mathrm{diag}(\tilde{d}_1^\xi,\dots,\tilde{d}_p^\xi) \in [0,1]^{p\times p}$ where:

$$\tilde{d}_i^\xi = \begin{cases} \frac{1}{d_\xi} & \text{if } B^i = B^\xi, \\ 1-\frac{1}{d_\xi} & \text{otherwise,} \end{cases} \qquad \forall \xi = 1,\dots,p,\ i = 1,\dots,n. \tag{29}$$

If $A^\xi = [\mathbf{a}^\xi(\mathbf{x}^1),\dots,\mathbf{a}^\xi(\mathbf{x}^p)]^t \in [0,1]^{p\times p}$, then the objective function above can be rewritten as follows:

$$\sum_{\xi=1}^{p}(\mathbf{z}^\xi - A^\xi\mathbf{u}^\xi)^t D^\xi(\mathbf{z}^\xi - A^\xi\mathbf{u}^\xi) \tag{30}$$

Here, we use the following procedure to tune the parameters W and $\mathbf{v}$:

1. Find $\mathbf{u}^\star \in \mathbb{R}^{pn}_{\geq 0}$ which solves the following optimization problem

$$\begin{aligned} &\min(\mathbf{z}-\mathbf{A}\mathbf{u})^t\mathbf{D}(\mathbf{z}-\mathbf{A}\mathbf{u}) \\ &\quad \text{subject to } \mathbf{u} \geq 0, \end{aligned} \tag{31}$$

   where $\mathbf{z} \in \{0,1\}^{p^2}$, $\mathbf{A} \in [0,1]^{p^2\times pn}$, and $\mathbf{D} \in [0,1]^{p^2\times p^2}$ are given as follows

$$\mathbf{z} = \begin{pmatrix} \mathbf{z}^1 \\ \vdots \\ \mathbf{z}^p \end{pmatrix},\ \mathbf{A} = \begin{pmatrix} A^1 & 0 & \dots & 0 \\ 0 & A^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A^p \end{pmatrix},\ \mathbf{D} = \begin{pmatrix} D^1 & 0 & \dots & 0 \\ 0 & D^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & D^p \end{pmatrix}.$$

2. For $\xi = 1,\dots,p$, define the vector $\mathbf{u}^\xi \in \mathbb{R}^n_{\geq 0}$ point-wise as $u_i^\xi = u^\star_{p(1-\xi)+i}$ for all $i = 1,\dots,n$.
3. Set $v_\xi = ||\mathbf{u}^\xi||_1$ and $w_i^\xi = \frac{u_i^\xi}{v_\xi}$ for all $\xi = 1,\dots,p$ and $i = 1,\dots,n$.

Let us finish this subsection with some remarks concerning Step 1, i.e., the minimization of the objective function $f(\mathbf{u}) = (\mathbf{z}-\mathbf{A}\mathbf{u})^t\mathbf{D}(\mathbf{z}-\mathbf{A}\mathbf{u})$ subject to $u_i \geq 0$ for $i = 1,\dots,pn$. First, note that the Hessian matrix $\nabla^2 f(\mathbf{u}) = 2\mathbf{A}^t\mathbf{D}\mathbf{A}$ is positive semi-definite since $\mathbf{D}$ is also positive semi-definite. Therefore, we are dealing with a convex optimization problem or—to be more precise—a quadratic program or quadratic programming problem [50]. As is the case for any

Please cite this article in press as: E. Esmi et al., Tunable equivalence fuzzy associative memories, Fuzzy Sets and Systems (2015), http://dx.doi.org/10.1016/j.fss.2015.04.004

Table 3
Description of the datasets.

	Instances	Categorical features	Numerical features	Classes	Repository
Appendicitis	106	0	7	2	KEEL
Breast	699	0	9	2	UCI
Cleveland	297	0	13	5	KEEL
Crx	653	9	6	2	KEEL
Ecoli	336	0	7	8	KEEL
Glass	214	0	9	7	KEEL
Heart	270	0	13	2	KEEL/UCI
Iris	150	0	4	3	KEEL/UCI
Liver	345	0	6	2	UCI
Monks	432	0	6	2	KEEL
Movementlibras	360	0	90	15	KEEL
Pima	768	0	8	2	KEEL/UCI
Satimage	6435	0	36	6	UCI
Segment	2310	0	19	7	UCI
Sonar	208	0	60	2	KEEL
Spectfheart	267	0	44	2	KEEL
Vowel	990	0	13	11	KEEL
Wdbc	569	0	30	2	KEEL
Wine	178	0	13	3	KEEL/UCI
Yeast	1484	0	7	10	UCI

convex optimization problem, all locally optimal points are globally optimal. In other words, all local minimum points are global minimum points. In addition, the inequality constraints $u_i \geq 0$, for $i = 1, \ldots, np$, are linear and therefore the Karush–Kuhn–Tucker conditions can be used as sufficient optimality criteria [51].

5. Experimental results

We evaluated the performance of tunable E-FAM models using fifteen datasets from the Knowledge Extraction Based on Evolutionary Learning (KEEL) Dataset Repository [25] and nine datasets from the UCI Machine Learning Repository of the University of California, Irvine [26]. Table 3 provides the numbers of instances, categorical and numerical features, and classes for each dataset under consideration. Note that the Iris, Heart, Pima, and Wine datasets can be found in both the KEEL and the UCI repository. Given the choice, we opted for extracting the datasets from the KEEL repository that disposes of a partitioning of each dataset into ten folds which is convenient for the subsequent experiments.

Each dataset in Table 3 comprises a finite number of instances of vectors that are divided in distinct classes and each component represents either a numerical or categorical attribute. Formally, each dataset is given by a finite subset of $T_1 \times \ldots \times T_{\tilde{n}} \times Y$ for some $\tilde{n} > 0$, say $\{(t_1^\xi, \ldots, t_{\tilde{n}}^\xi, y^\xi) \in T_1 \times \ldots \times T_{\tilde{n}} \times Y \mid \xi = 1, \ldots, k\}$, where each T_i is either a set of numerical values of the form $[t_{min}^i, t_{max}^i] \subset \mathbb{R}$ or a set of categorical values of the form $\{l_1^i, \ldots, l_c^i\}$, $c > 0$, and Y is a set of $m > 0$ categorical values $y_1, \ldots, y_m$ representing class labels.

Since a tunable E-FAM represents a map from a product of bounded lattices to a class of fuzzy sets $\mathcal{F}(Y)$, we associated each attribute domain T_i with a product $\mathbb{L}_i$ of bounded lattices $\mathbb{L}_{i_1}, \ldots, \mathbb{L}_{i_c}$ for some $i_c > 0$. Each $\mathbb{L}_{i_j}$ can be furnished with an equivalence measure E_{i_j} as follows. On the one hand, if T_i is an interval $[t_{min}^i, t_{max}^i]$ of numerical values, then T_i constitutes a complete chain and we set $\mathbb{L}_i = T_i$. An equivalence measure $E_i = E_{\lambda_i}$ on T_i arises from Eq. (5). Here we calculated λ_i by dividing the standard deviation of $\{t_i^\xi \mid \xi = 1, \ldots, k\}$ by $|t_{max}^i - t_{min}^i|$. On the other hand, if T_i is of the form $\{l_1^i, \ldots, l_c^i\}$ for some $c > 0$, then we associated T_i with the class of fuzzy sets $[0, 1]^c$, i.e., the product of the complete chains $\mathbb{L}_{i_j} = [0, 1]$, $j = 1, \ldots, c$. In this case, we mapped each categorical value l_q^i to the binary vector $\mathbf{a}^q \in \{0, 1\}^c \subset [0, 1]^c$ satisfying $a_j^q = 1$ if and only if $q = j$. Note that $\mathbf{a}^q$ can also be interpreted as a fuzzy set. Then, we defined an equivalence measure E_{i_j} on each $\mathbb{L}_{i_j} = [0, 1]$, where $j = 1, \ldots, c$, by setting $E_{i_j} = E_{\lambda_j}$ with λ_j given by the standard deviation of $\{a_j^q \mid q = 1, \ldots, c\}$. Finally, since the finite universe Y can be

Table 4
Classification rates achieved by the TE-FAM approach in comparison with the ones presented in [27].

	2SLAVE	FH-GBML	SGERD	CBA	CBA2	CMAR	CPAR	C4.5	FARC-HD	**TE-FAM**
Appendicitis	82.91	86	84.48	89.6	89.6	**89.7**	87.8	83.3	84.2	84.09
Cleveland	48.82	53.51	51.59	56.9	54.9	53.9	54.9	54.5	55.2	**61.01**
Crx	74.06	86.6	85.03	83.6	85	85	**87.3**	85.3	86	86.66
Ecoli	**84.53**	69.38	74.05	78	77.1	77.7	76.2	79.5	82.2	80.97
Glass	58.05	57.99	58.49	70.8	**71.3**	70.3	68.9	67.4	70.2	63.85
Heart	71.36	75.93	73.21	83	81.5	82.2	80.7	78.5	**84.4**	82.59
Iris	94.44	94	94.89	93.3	93.3	94	**96**	**96**	**96**	**96**
Monks	97.26	98.18	80.65	**100**	**100**	**100**	**100**	**100**	99.8	97.73
Movementlibras	67.04	68.89	68.09	36.1	7.2	39.2	63.6	69.4	**76.7**	69.17
Pima	73.71	75.26	73.37	72.7	72.5	75.1	74.5	74	75.7	**76.18**
Sonar	71.42	68.24	71.9	75.4	77.9	78.8	75	70.5	80.2	**83.62**
Spectfheart	79.17	72.36	78.16	79.8	79.8	79.4	78.3	76.5	79.8	**81.67**
Vowel	71.11	67.07	65.83	63.6	74.9	60.4	63.00	**81.5**	71.8	80
Wdbc	92.33	92.26	90.68	94.7	95.1	94.9	95.1	95.2	**95.3**	93.67
Wine	89.47	92.61	91.88	93.8	93.8	**96.7**	95.6	93.3	94.3	96.63
Mean	77.05	77.22	76.15	78.09	76.93	78.49	79.79	80.33	82.12	**82.25**

viewed as a categorical attribute, each class label $y_j \in Y$ can be associated with the crisp set $B^j \in \mathcal{F}(Y)$ that satisfies $B^j_q = 1$ if and only if $q = i$.

In this manner, we mapped every element of $T_1 \times \ldots \times T_{\tilde{n}} \times Y$ to a unique element of $\mathbb{L} \times \mathcal{F}(Y)$, where $\mathbb{L}$ is a product of bounded lattices with constituents $\mathbb{L}_{i_j}$ for $i = 1, \ldots, \tilde{n}$ and $j = 1, \ldots, i_c$. Overall we have that $\mathbb{L}$ is a product of $n = \sum_{i=1}^{\tilde{n}} i_c \geq \tilde{n}$ complete chains.

In our simulations, we trained the tunable E-FAM models using Algorithm TE that is detailed in Section 4. Recall that Algorithm TE has two stages: 1) extraction of fundamental memory set $\mathcal{M}$ and 2) selection of optimal weights. In Stage 1, corresponding to the extraction of $\mathcal{M}$ via Algorithm 1, we defined the parameter ρ as the mean of the values $\bar{e}^1, \ldots, \bar{e}^k$ that are obtained in the previous step (i.e., Step 2). Note that the parameter $\rho = \frac{1}{k}\sum_{\xi=1}^{k} \bar{e}^\xi$ acts as a lower bound for selecting potential candidates to replace a certain element $\mathbf{x}^\gamma$ by an element of the same class. Each value $\bar{e}^\xi$ denotes the maximum of $E(\mathbf{x}^\xi, \mathbf{x}^\gamma)$ for all $\mathbf{x}^\gamma$ such that $\mathbf{x}^\gamma$ and $\mathbf{x}^\xi$ belong to different classes. In Stage 2, we solve the optimization problem given in Eq. (31) using the well-known *Trust-Region-Reflective Optimization* method [52,53] with the initial vector $(\frac{1}{n}, \ldots, \frac{1}{n})^t \in [0, 1]^n$ and a tolerance parameter of 10^{-6} for the constraint violation and stopping criteria.

We compared the classification results obtained by tunable E-FAMs of type $\mathcal{E}_{W,\mathbf{v}}$ with the ones produced by other competitive classifiers that were applied to the same problems in two recent articles [27,28]. Specifically, the first article [27] reports the mean classification rates produced by the following approaches in applications of 10-fold cross-validation to each dataset from the KEEL Dataset Repository of Table 3: a C4.5 decision tree [54], structural learning algorithm on vague environment (2SLAVE) [55], classification based on associations (CBA) [56], an improved version of the CBA method (CBA2) [57], classification based on multiple association rules (CMAR) [58], classification based on predictive association rules (CPAR) [59], fuzzy hybrid genetic based machine learning algorithm (FH-GBML) [60], steady-state genetic algorithm for extracting fuzzy classification rules from data (SGERD) [61], and fuzzy association rule-based classification method for high-dimensional problems (FARC-HD) [27]. To ensure a fair comparison, we employed the same partitioning of the data as in [25,27]. The second article [28] reports the mean classification rates produced by the following approaches in applications of 10-fold cross-validation to each dataset from the UCI repository of Table 3: a radial basis function (RBF) network [62], two memetic multiobjective particle swarm optimization RBF networks (MPSONs) [28], two memetic non-dominated sorting genetic algorithm RBF networks (MGANs) [28,63], a support vector machine (SVM) [64], a C4.5 decision tree [54], and a random forest [65].

All experiments concerning the TE-FAM approach were implemented using MATLAB®. Table 5 exhibits the mean CPU time in each stage of the training phase of the tunable E-FAMs for each dataset of Table 4 (using an Intel® Core™ i5-3330 CPU with a processing speed of 3.00 GHz).

Table 5
$|\mathcal{M}|/|\mathcal{T}|$ and CPU times (in seconds) for training the TE-FAM as well as CPU times (in seconds) required by the weighted KS-FAM (w. KS-FAM).

	Instances	Mean % $|\mathcal{M}|/|\mathcal{T}|$	Stage 1	Stage 2	Total	w. KS-FAM
Appendicitis	106	62.16	0.01	0.12	0.12	1.21
Cleveland	297	78.71	0.03	0.29	0.32	126.41
Crx	653	63.42	0.48	4.04	4.51	465.37
Ecoli	336	65.15	0.02	0.22	0.24	167.34
Glass	214	68.07	0.01	0.17	0.19	46.93
Heart	270	68.23	0.02	0.22	0.24	27.31
Iris	150	38.89	0.01	0.09	0.10	1.62
Monks	432	87.22	0.03	0.24	0.27	37.56
Movementlibras	360	64.88	0.31	8.27	8.58	248.35
Pima	768	73.77	0.14	0.50	0.64	2605.55
Sonar	208	72.44	0.09	0.98	1.06	15.19
Spectfheart	267	88.43	0.10	1.45	1.55	22.77
Vowel	990	35.54	0.35	0.44	0.80	6032.68
Wdbc	569	46.83	0.25	1.18	1.43	156.81
Wine	178	45.38	0.01	0.15	0.16	4.73

Table 4 displays the mean classification rates obtained by TE-FAMs and the first set of the aforementioned models [27] in the testing phase. Note that our approach performed relatively well exhibiting the highest classification rates for the Iris, Pima, Sonar, and Spectfheart datasets as well as the best mean classification accuracy overall, achieving a slight improvement of the result produced by the FARC-HD classifier.

In a recent article, Alcalá-Fdez et al. [27] performed several statistical tests on a number of datasets in order to compare the performance of the FARC-HD model with all of the classifiers listed in Table 4 except for the TE-FAM. According to this article, the best accuracy model among these nine classifiers is given by the FARC-HD. Therefore, we decided to conduct a statistical comparison of our TE-FAM model with the FARC-HD by means of the Wilcoxon signed-ranks test [66,67].

An application of this statistical test to the TE-FAM and the FARC-HD using the fifteen datasets of Table 4 yields respectively the ranks $R^+ = 63$ and $R^- = 56$. Although the TE-FAM achieved a higher rank than the FARC-HD, the null-hypothesis that the FARC-HD and TE-FAM models perform equally well cannot be rejected since R^- is greater than the critical value 25 or, equivalently, the p-value 0.82 is greater than $\alpha = 0.05$. Thus, the Wilcoxon signed-ranks test yields no significant difference between these two models.

Recently, Esmi et al. also conducted experiments on the datasets of Table 4 using Θ-FAMs whose weight vectors were trained using the previous Θ-FAM training algorithm [1]. For each of the fifteen datasets, appropriate functions Θ^ξ were selected using 10-fold cross-validation [68]. The Θ-FAM approach presented in [1] yielded an even higher mean classification rate of 82.60% on these fifteen datasets but involved a huge computational cost in comparison to the one of the TE-FAM approach (cf. Table 5). Even for a fixed choice of functions Θ^ξ, the training phase required a large amount of processing time. For example, the training phase of a weighted KS-FAM [1], i.e., a Θ-FAM based on the well-known Kosko subsethood measure [12], required a CPU time of about 1 hour and 30 minutes for the Vowel data alone as shown in the last column of Table 5.

Alcalá-Fdez et al. also reported the CPU times required by the FARC-HD model in applications to the datasets listed in Tables 4 and 5 [27]. These CPU times are not directly comparable with the ones listed in Table 5 since Alcalá-Fdez et al. employed a different software and equipment, namely an open source (GPLv3) Java software tool called *KEEL* [25] and a Pentium Corel 2 Quad, 2.5 GHz CPU with 4 GB of memory running in Linux. However, Table XV of [27] indicates that the computational cost for training the FARC-HD model is much higher than the one for training the TE-FAM.

Note that Θ-FAM models such as TE-FAMs share some conceptual similarities with artificial neural network models such as RBF, fuzzy, and Bayesian neural networks. Whereas radial basis functions play the role of hidden layer activation functions in RBF networks, this role is played by the functions Θ^ξ in Θ-FAMs, in particular by parametrized equivalence measures in TE-FAMs.

Therefore, we followed a reviewer's suggestion and conducted an additional set of experiments in order to compare the classification performances of TE-FAMs and certain artificial neural network models. Specifically, we consid-

Please cite this article in press as: E. Esmi et al., Tunable equivalence fuzzy associative memories, Fuzzy Sets and Systems (2015), http://dx.doi.org/10.1016/j.fss.2015.04.004

Table 6
Classification rates achieved by the TE-FAM approach in comparison with the ones presented in [28].

Models	Breast	Heart	Iris	Liver	Pima	Satimage	Segment	Wine	Yeast	Mean
MGANf1f2	96.78	79.07	83.78	62.63	72.78	83.33	86.9	72.18	90	80.83
MGANf1–f3	97.8	80.79	84.44	63.5	68.35	83.33	86.22	72.04	90.01	80.72
MPSONf1f2	97.66	85.2	89.11	68.48	77.34	83.33	88.46	75.67	90.42	83.96
MPSONf1–f3	**97.81**	**85.54**	96.22	74.26	**78.25**	93	91.39	81.65	**90.5**	87.62
RBF network	95.71	84.62	95.56	62.14	74.35	85.29	84.99	**98.11**	58.88	82.18
C4.5	93.81	76.92	95.56	71.85	76.52	88.17	96.54	94.34	55.96	83.3
Random forest	96.19	82.42	95.56	64.08	72.61	90.73	**97.69**	94.34	60	83.73
SVM	96.49	54.88	**96.67**	59.42	65.1	21.3	65.37	44.38	43.26	60.76
TE-FAM	95.71	82.59	96	**82.61**	76.18	**96.02**	90.58	96.63	77.85	**88.24**

ered eight models, namely, a conventional RBF network, MPSON with two objectives optimization (MPSONf1f2), MGAN with two objectives optimization (MGANf1f2), MPSON with three objectives optimization (MPSONf1–f3), MGAN with three objectives optimization (MGANf1–f3), SVM, random forest, and C4.5 [28]. We restricted our attention to the nine datasets[1] discussed in [28] that were drawn from the UCI repository and that have no missing values [26].

Table 6 lists the classification rates produced by the TE-FAM approach in comparison to the ones produced by the eight other models. The TE-FAM approach exhibited a very satisfactory performance, achieving the best mean classification accuracy overall and with respect to the Liver and Satimage datasets. The second best mean classification accuracy was achieved by the MPSONf1–f3 approach [28] that yielded the best performance in four (Breast, Yeast, Pima, and Heart) of the nine datasets. Except for the Liver and Yeast datasets, the TE-FAM and MPSONf1–f3 approaches yielded comparable classification rates.

6. Conclusion

The Θ-FAM model, that recently appeared in the literature, represents a two-layer FAM. In this work, we extended the original definition of a Θ-FAM by allowing inputs in an arbitrary bounded lattice $\mathbb{L}$, arbitrary activation functions, and sup-t aggregation functions of output nodes. Hence, Θ-FAMs can be viewed as approaches towards lattice computing, "an evolving collection of tools and mathematical modeling methodologies with the capacity to process lattice ordered data per se including logic values, numbers, sets, symbols, graphs, etc." [69–73].

We defined equivalence fuzzy associative memories (E-FAMs) as special types of Θ-FAMs whose hidden layer aggregation functions are given by equivalence measures. We speak of tunable equivalence fuzzy associative memories (TE-FAMs) if the hidden layer aggregation functions are given by parametrized equivalence measures. Then we presented a new supervised training algorithm for TE-FAMs, called Algorithm TE, for optimizing the contributions of the attributes of the data by adjusting the parameters of the equivalence measures as well as the weights corresponding to the contributions of the hidden nodes. Algorithm TE first extracts a fundamental memory set $\mathcal{M}$ from a given training set $\mathcal{T}$ before tuning the parameters of the network. The construction of $\mathcal{M} \subseteq \mathcal{T}$ can be performed for any E-FAM. In this context, recall that the number of hidden nodes equals $|\mathcal{M}|$ in a Θ-FAM. We proved that an E-FAM with fundamental memory set $\mathcal{M}$, constructed using Algorithm TE, and constant, non-zero weights in the first layer produces no training error under some weak conditions on the equivalence measures used in the hidden nodes. Then we showed how to adjust the parameters of a tunable E-FAM by solving a quadratic optimization problem in the second stage of Algorithm TE.

Subsequently, we applied TE-FAMs based on parametrized equivalence measures of the form $E_{\lambda,\mathbf{w}}$ (cf. Eq. (11)) in conjunction with Algorithm TE to several classification problems from the KEEL-Dataset Repository that is available on the Internet [25]. The TE-FAM approach exhibited a slightly better mean classification accuracy than several competitive fuzzy rule-based classifiers that have recently been applied to the same fifteen problems of Table 4 by Alcala-Fdez et al. [27]. We observed that TE-FAMs, that were trained using Algorithm TE, required a relatively short CPU times in comparison to the previous Θ-FAM approach [1]. Finally, we employed nine datasets from the UCI

[1] Note that the Iris, Heart, Pima, and Wine datasets appear in Tables 4, 5, and 6.

repository [26] to compare the classification accuracies obtained by the TE-FAM approach with the ones obtained by other well-known classifiers, including RBF networks whose topology resembles the one of Θ-FAMs as pointed out at the end of Section 5. In these simulations, the TE-FAM approach also performed well in comparison to the other models that were considered in a recent article [28].

In the future, we intend to investigate possible improvements regarding Algorithm TE, in particular with respect to the optimization of the parameters, and to develop Θ-FAMs for a wider range of pattern recognition problems, in particular for regression [74].

References

[1] E. Esmi, P. Sussner, H. Bustince, J. Fernández, θ-Fuzzy associative memories, IEEE Trans. Fuzzy Syst. 23 (2) (2015) 313–326.
[2] P. Sussner, E.L. Esmi, I. Villaverde, M. Graña, The Kosko subsethood fuzzy associative memory (KS-FAM): mathematical background and applications in computer vision, J. Math. Imaging Vis. 42 (2012) 134–149.
[3] D. Sinha, E.R. Dougherty, Fuzzy mathematical morphology, J. Vis. Commun. Image Represent. 3 (3) (1992) 286–302.
[4] I. Bloch, H. Maitre, Fuzzy mathematical morphologies: a comparative study, Pattern Recognit. 28 (9) (1995) 1341–1387.
[5] M. Nachtegael, E.E. Kerre, Connections between binary, gray-scale and fuzzy mathematical morphologies, Fuzzy Sets Syst. 124 (1) (2001) 73–85.
[6] P. Sussner, M.E. Valle, Classification of fuzzy mathematical morphologies based on concepts of inclusion measure and duality, J. Math. Imaging Vis. 32 (2) (2008) 139–159.
[7] D. Sinha, P. Sinha, E.R. Dougherty, S. Batman, Design and analysis of fuzzy morphological algorithms for image processing, IEEE Trans. Fuzzy Syst. 5 (4) (1997) 570–583.
[8] M.E. Valle, P. Sussner, A general framework for fuzzy morphological associative memories, Fuzzy Sets Syst. 159 (7) (2008) 747–768.
[9] G.J.F. Banon, J. Barrera, Decomposition of mappings between complete lattices by mathematical morphology, part 1, general lattices, Signal Process. 30 (3) (1993) 299–327.
[10] H. Heijmans, C. Ronse, The algebraic basis of mathematical morphology I. Dilations and erosions, Comput. Vis. Graph. Image Process. 50 (3) (1990) 245–295.
[11] J. Serra, Image Analysis and Mathematical Morphology, Volume 2: Theoretical Advances, Academic Press, New York, 1988.
[12] B. Kosko, Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence, Prentice Hall, Englewood Cliffs, NJ, 1992.
[13] F. Chung, T. Lee, On fuzzy associative memory with multiple-rule storage capacity, IEEE Trans. Fuzzy Syst. 4 (3) (1996) 375–384.
[14] F. Junbo, J. Fan, S. Yan, A learning rule for fuzzy associative memories, in: Proceedings of the IEEE International Joint Conference on Neural Networks, vol. 7, 1994, pp. 4273–4277.
[15] P. Liu, The fuzzy associative memory of max–min fuzzy neural networks with threshold, Fuzzy Sets Syst. 107 (1999) 147–157.
[16] R. Belohlávek, Fuzzy logical bidirectional associative memory, Inf. Sci. 128 (1–2) (2000) 91–103.
[17] P. Sussner, M.E. Valle, Implicative fuzzy associative memories, IEEE Trans. Fuzzy Syst. 14 (6) (2006) 793–807.
[18] J.J. Buckley, Y. Hayashi, Fuzzy neural networks: a survey, Fuzzy Sets Syst. 66 (1994) 1–13.
[19] P. Sussner, M.E. Valle, Fuzzy associative memories and their relationship to mathematical morphology, in: W. Pedrycz, A. Skowron, V. Kreinovich (Eds.), Handbook of Granular Computing, John Wiley and Sons, Inc., New York, 2008, Ch. 33.
[20] R.P. Lippmann, An introduction to computing with neural nets, IEEE Trans. Acoust. Speech Signal Process. ASSP-4 (1987) 4–22.
[21] N. Ikeda, P. Watta, M. Artiklar, M.H. Hassoun, A two-level Hamming network for high performance associative memory, Neural Netw. 14 (9) (2001) 1189–1200.
[22] X. Mu, P. Watta, M. Hassoun, A weighted voting model of associative memory, IEEE Trans. Neural Netw. 18 (3) (2007) 756–777.
[23] B. Schölkopf, A. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, Adaptive Computation and Machine Learning, MIT Press, 2002.
[24] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, Belmont, MA, 1995.
[25] J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera, KEEL data-mining software tool: data set repository, integration of algorithms and experimental analysis framework, J. Mult.-Valued Log. Soft Comput. 17 (2–3) (2011) 255–287.
[26] A. Asuncion, D.J. Newman, UCI machine learning repository, University of California, Irvine, School of Information and Computer Sciences, http://www.ics.uci.edu/~mlearn/MLRepository.html, 2007.
[27] J. Alcala-Fdez, R. Alcala, F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning, IEEE Trans. Fuzzy Syst. 19 (5) (2011) 857–872.
[28] S.N. Qasem, S.M. Shamsuddin, S.Z.M. Hashim, M. Darus, E. Al-Shammari, Memetic multiobjective particle swarm optimization-based radial basis function network for classification problems, Inf. Sci. 239 (2013) 165–190.
[29] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353.
[30] A. Di Nola, S. Sessa, W. Pedrycz, E. Sanchez, Fuzzy Relation Equations and Their Applications to Knowledge Engineering, Kluwer Academic Publishers, Norwell, MA, 1989.
[31] G. Birkhoff, Lattice Theory, 3rd edition, American Mathematical Society, Providence, 1993.
[32] J. Goguen, L-fuzzy sets, J. Math. Anal. Appl. 18 (1) (1967) 145–174.
[33] G.A. Grätzer, Lattice Theory: First Concepts and Distributive Lattices, Freeman, San Francisco, CA, 1971.

Please cite this article in press as: E. Esmi et al., Tunable equivalence fuzzy associative memories, Fuzzy Sets and Systems (2015), http://dx.doi.org/10.1016/j.fss.2015.04.004

[34] J. Fodor, M. Roubens, Fuzzy preference modeling and multicriteria decision support, in: Theory and Decision Library, Kluwer Academic Publishers, 1994.
[35] J. Fodor, R.R. Yager, Fuzzy set-theoretic operators and quantifiers, in: D. Dubois, H. Prade (Eds.), Fundamentals of Fuzzy Sets, in: Handb. Fuzzy Sets Ser., vol. 7, Springer US, 2000, pp. 125–193.
[36] H. Bustince, M. Pagola, E. Barrenechea, Construction of fuzzy indices from fuzzy DI-subsethood measures: application to the global comparison of images, Inf. Sci. 177 (3) (2007) 906–929.
[37] D. Dubois, H. Prade, A unifying view of comparison indices in a fuzzy set-theoretic framework, in: R.R. Yager (Ed.), Recent Developments in Fuzzy Set and Possibility Theory, Pergamon Press, 1982, pp. 3–13.
[38] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall, Upper Saddle River, NY, 1995.
[39] J. Fan, W. Xie, Some notes on similarity measure and proximity measure, Fuzzy Sets Syst. 101 (3) (1999) 403–412.
[40] X. Liu, Entropy, distance measure and similarity measure of fuzzy sets and their relations, Fuzzy Sets Syst. 52 (1992) 305–318.
[41] H.-Y. Zhang, W.-X. Zhang, Hybrid monotonic inclusion measure and its use in measuring similarity and distance between fuzzy sets, Fuzzy Sets Syst. 160 (1) (2009) 107–118.
[42] W. Zeng, H. Li, Inclusion measures, similarity measures, and the fuzziness of fuzzy sets and their relations, Int. J. Intell. Syst. 21 (6) (2006) 639–653.
[43] S. Sandri, F. Martins-Bedê, A method for deriving order compatible fuzzy relations from convex fuzzy partitions, Fuzzy Sets Syst. 239 (2014) 91–103.
[44] E. Esmi, P. Sussner, M. Valle, F. Sakuray, L. Barros, Fuzzy associative memories based on subsethood and similarity measures with applications to speaker identification, in: E. Corchado, et al. (Eds.), Hybrid Artificial Intelligent Systems, in: Lect. Notes Comput. Sci., vol. 7209, Springer Verlag, Berlin/Heidelberg, 2012, pp. 479–490.
[45] H. Bustince, E. Barrenechea, M. Pagola, Restricted equivalence functions, Fuzzy Sets Syst. 157 (17) (2006) 2333–2346.
[46] F.T. Martins-Bedé, L. Godo, S. Sandri, L.V. Dutra, C.C. Freitas, O.S. Carvalho, R.J. Guimarães, R.S. Amaral, Classification of schistosomiasis prevalence using fuzzy case-based reasoning, in: Proceedings of the 10th International Work-Conference on Artificial Neural Networks: Part I: Bio-Inspired Systems: Computational and Ambient Intelligence, IWANN '09, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 1053–1060.
[47] G.J. Klir, T.A. Folger, Fuzzy Sets, Uncertainty, and Information, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1987.
[48] I. Drummond, L. Godo, S. Sandri, Restoring consistency in systems of fuzzy gradual rules using similarity relations, in: Lecture Notes in Computer Science, vol. 2507, 2002, pp. 386–396.
[49] J. Fan, W. Xie, J. Pei, Subsethood measure: new definitions, Fuzzy Sets Syst. 106 (2) (1999) 201–209.
[50] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, New York, NY, USA, 2004.
[51] M. Bazaraa, H. Sherali, C. Shetty, Nonlinear Programming: Theory and Algorithms, Wiley, 2006.
[52] T. Coleman, Y. Li, On the convergence of interior-reflective newton methods for nonlinear minimization subject to bounds, Math. Program. 67 (1–3) (1994) 189–224.
[53] T.F. Coleman, Y. Li, An interior trust region approach for nonlinear minimization subject to bounds, Tech. rep., Ithaca, NY, USA 1993.
[54] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
[55] A. Gonzalez, R. Perez, Selection of relevant features in a fuzzy genetic learning algorithm, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 31 (3) (2001) 417–425.
[56] B. Liu, W. Hsu, Y. Ma, Integrating classification and association rule mining, in: Proceedings of the International Conference on Knowledge Discovery and Data Mining, SIGKDD, New York, USA, 1998, pp. 80–86.
[57] B. Liu, Y. Ma, C.-K. Wong, Classification using association rules: weaknesses and enhancements, in: C.K.R. Grossman, V. Kumar (Eds.), Data Mining for Scientific and Engineering Applications, Kluwer Academic Publishers, 2001, pp. 591–601.
[58] W. Li, J. Han, J. Pei, CMAR: accurate and efficient classification based on multiple class-association rules, in: Proceedings of IEEE International Conference on Data Mining, ICDM 2001, 2001, pp. 369–376.
[59] X. Yin, J. Han, CPAR: classification based on predictive association rules, in: Proceedings of 3rd SIAM International Conference on Data Mining, SDM, San Francisco, CA, USA, 2003, pp. 331–335.
[60] H. Ishibuchi, T. Yamamoto, T. Nakashima, Hybridization of fuzzy GBML approaches for pattern classification problems, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 35 (2) (2005) 359–365.
[61] E. Mansoori, M. Zolghadri, S. Katebi, SGERD: a steady-state genetic algorithm for extracting fuzzy classification rules from data, IEEE Trans. Fuzzy Syst. 16 (4) (2008) 1061–1071.
[62] J. Leonard, M.A. Kramer, Radial basis function networks for classifying process faults, IEEE Control Syst. Mag. 11 (3) (1991) 31–38.
[63] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197.
[64] V.N. Vapnik, Statistical Learning Theory, 1st edition, Wiley, 1998.
[65] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[66] D.J. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, 4th edition, Chapman and Hall/CRC, 2007.
[67] F. Wilcoxon, Individual comparisons by ranking methods, Biom. Bull. 1 (6) (1945) 80–83.
[68] E. Esmi, P. Sussner, Some experimental results on the use of θ-FAMs, University of Campinas, Dept. of Applied Math., Mathematical Imaging and Computational Intelligence Group, http://www.milab.ime.unicamp.br/ieeeThetaFam.pdf, 2011.
[69] V. Kaburlasos, A. Kehagias, Fuzzy inference system (FIS) extensions based on lattice theory, IEEE Trans. Fuzzy Syst. 22 (3) (2014) 531–546.
[70] V. Kaburlasos, S. Papadakis, G. Papakostas, Lattice computing extension of the FAM neural classifier for human facial expression recognition, IEEE Trans. Neural Netw. Learn. Syst. 24 (10) (2013) 1526–1538.
[71] M. Graña, D. Chyzhyk, M. García-Sebastián, C. Hernández, Lattice independent component analysis for functional magnetic resonance imaging, Inf. Sci. 181 (10) (2011) 1910–1928.

[72] P. Sussner, M. Nachtegael, T. Mélange, G. Deschrijver, E. Esmi, E. Kerre, Interval-valued and intuitionistic fuzzy mathematical morphologies as special cases of **L**-fuzzy mathematical morphology, J. Math. Imaging Vis. 43 (2012) 50–71.

[73] M.E. Valle, P. Sussner, Quantale-based autoassociative memories with an application to the storage of color images, Pattern Recognit. Lett. 34 (14) (2013) 1589–1601.

[74] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, Oxford, UK, 1995.

Please cite this article in press as: E. Esmi et al., Tunable equivalence fuzzy associative memories, Fuzzy Sets and Systems (2015), http://dx.doi.org/10.1016/j.fss.2015.04.004