Neurocomputing 139 (2014) 298–309

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Generalization rules for the suppressed fuzzy *c*-means clustering algorithm

László Szilágyi [a,b,1], Sándor M. Szilágyi [c,*]

[a] Sapientia University of Transylvania, Faculty of Technical and Human Sciences, Şoseaua Sighişoarei 1/C, 540485 Tîrgu Mureş, Romania
[b] Budapest University of Technology and Economics, Department of Control Engineering and Information Technology, Magyar tudósok krt. 2, H-1117 Budapest, Hungary
[c] Petru Maior University, Department of Informatics, Str. Nicolae Iorga Nr. 1, 540088 Tîrgu Mureş, Romania

ARTICLE INFO

Article history:
Received 14 February 2013
Received in revised form
13 September 2013
Accepted 4 February 2014
Communicated by M. Sato-Ilic
Available online 13 April 2014

Keywords:
Clustering
Fuzzy *c*-means clustering
Suppressed fuzzy *c*-means clustering
Fuzzy *c*-means with improved partition

ABSTRACT

Intending to achieve an algorithm characterized by the quick convergence of hard *c*-means (HCM) and finer partitions of fuzzy *c*-means (FCM), suppressed fuzzy *c*-means (s-FCM) clustering was designed to augment the gap between high and low values of the fuzzy membership functions. Suppression is produced via modifying the FCM iteration by creating a competition among clusters: for each input vector, lower degrees of membership are proportionally reduced, being multiplied by a previously set constant suppression rate, while the largest fuzzy membership grows to maintain the probabilistic constraint. Even though so far it was not treated as an optimal algorithm, it was employed in a series of applications, and reported to be accurate and efficient in various clustering problems. In this paper we introduce some generalized formulations of the suppression rule, leading to an infinite number of new clustering algorithms. Further on, we identify the close relation between s-FCM clustering models and the so-called FCM algorithm with generalized improved partition (GIFP-FCM). Finally we reveal the constraints under which the generalized s-FCM clustering models minimize the objective function of GIFP-FCM, allowing us to call our suppressed clustering models optimal. Based on a large amount of numerical tests performed in multidimensional environment, several generalized forms of suppression proved to give more accurate partitions than earlier solutions, needing significantly less iterations than the conventional FCM.

1. Introduction

C-means clustering algorithms belong to unsupervised classification methods which group a set of input vectors into a previously defined number (c) of classes. Initially there was the hard *c*-means (HCM) algorithm [31], which employed the bivalent (crisp) logic to describe partitions. HCM usually converges quickly, but is considerably sensitive to initialization [1], and frequently gets stuck in local minima leading to mediocre partitions.

The introduction of fuzzy logic [41] into classification theory led to the definition of the fuzzy partition [34], in which every input vector can belong to several classes with various degrees of membership, and the formulation of fuzzy clustering problems [10]. The first *c*-means clustering algorithm that employs fuzzy partitions is the so-called fuzzy *c*-means (FCM) proposed by Bezdek [3], which uses a probabilistic constraint to define the fuzzy membership functions. FCM reportedly creates finer partitions than HCM, it has a reduced but still observable sensitivity to initial cluster prototypes, but it converges much slower. In spite of this drawback, FCM is one of the most popular clustering algorithms [23] not only in engineering studies, but also in a wide variety of sciences from climatology [22] to economical forecasting [24].

There exist several reported attempts to reduce the execution time of FCM, without causing serious damage to the partition quality. Early solutions—developed for computers with reduced computing power—turned to data approximation. The first accelerated FCM algorithms [5,25] used integer computation only. Cheng et al. [8] proposed data reduction based on random sampling, leading to a fast approximative FCM clustering. Higher speed has been also reached via data reduction. Eschrich et al. [11] aggregated similar input vectors into a weighted example. Their solution was able to speed up FCM by an order of magnitude. Data aggregation was employed in image segmentation, to cluster gray intensity levels instead of individual pixels, speeding up the execution by two orders of magnitude [36].

Alternately, Lázaro et al. [30] proposed a parallel hardware implementation to the FCM algorithm and successfully applied it in signal processing. Kolen and Hutcheson [27] reorganized FCM to

* Corresponding author. Tel./fax: +40 265 262275.
E-mail addresses: lalo@ms.sapientia.ro (L. Szilágyi), szsandor72@yahoo.com (S.M. Szilágyi).
[1] Tel.: +40 265 208170; fax: +40 265 206211.

http://dx.doi.org/10.1016/j.neucom.2014.02.027
0925-2312/

reduce the necessary memory storage by eliminating the partition matrix. Their solution was found significantly quicker than the classical formulation of FCM, especially in case of a large amount of input data. Further remarkable solutions for clustering unloadable amount of input data were given by Hathaway and Bezdek [14], and Havens et al. [15].

The suppressed fuzzy c-means (s-FCM) algorithm, introduced by Fan et al. [12], proposed to make a step from FCM towards HCM, by manipulating with the fuzzy membership functions computed in each iteration of the FCM's alternating optimization (AO) scheme. The authors defined a previously set constant suppression rate $\alpha \in [0,1]$, which determined the behavior of the algorithm. In each iteration, after having determined the new fuzzy membership functions for input vector with index k, denoted by $u_{1k}, u_{2k}, \ldots, u_{ck}$, the algorithm looks for the largest (winner) membership value u_{wk}, with $w = \arg\max_i \{u_{ik}, i = 1,2,\ldots,c\}$, suppresses all non-winner memberships by multiplication with α and raises the winner membership value u_{wk} in such a way that the probabilistic constraint is not harmed. The authors remarked that setting the suppression rate $\alpha = 0$ makes s-FCM behave like HCM, while $\alpha = 1$ reduces the algorithm to the conventional FCM. The authors failed to explain in detail what is happening in case of other values of α, they only said that the competition created among classes leads to beneficial effects in both execution time and partition quality. Later, Hung et al. [19,20] proposed some suppression rate selection schemes, and employed s-FCM to segment magnetic resonance images. Hung et al. [19] reported improved segmentation results for the suppressed FCM compared to the conventional one. Further modifications and applications of s-FCM were proposed recently in [35,21,32,39].

In a previous paper [37], we have studied in detail the competition of clusters caused by suppression. We have introduced a quasi-learning rate (QLR), similar to the learning rate known from conventional competitive clustering algorithms [26]. This QLR gives a mathematically precise characterization of the competition. In this paper we will show that the QLR established in [37] allows us to define a series of generalization rules for the suppression, opening the gate towards novel efficient and accurate clustering algorithms. We will further show that all sorts of suppressed FCM algorithms are optimal, as they minimize a special version of the objective function introduced by Zhu et al. in their generalized theory of FCM algorithms with improved partition [42].

The rest of this paper is structured as follows. Section 2 presents the necessary details of s-FCM and the competition it creates, which stand at the basis of our investigations. Section 3 introduces several types of generalized suppression rules for s-FCM, giving an analytical and graphical description of the proposed methods. Section 4 discusses the question of optimality of all suppressed FCM clustering models. Section 5 reports and discusses the benchmark tests performed with the generalized s-FCM algorithm variants. Conclusions are given in the last section.

2. Related works

2.1. The fuzzy and hard c-means algorithms

The conventional FCM algorithm partitions a set of object data into a number of c clusters based on the minimization of a quadratic objective function. The objective function to be minimized is defined as

$$J_{\mathrm{FCM}} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m \|\mathbf{x}_k - \mathbf{v}_i\|_A^2 = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m d_{ik}^2, \quad (1)$$

where $\mathbf{x}_k$ represents the input data ($k = 1\ldots n$), $\mathbf{v}_i$ represents the prototype or centroid value or representative element of cluster i ($i = 1\ldots c$), $u_{ik} \in [0,1]$ is the fuzzy membership function showing the degree to which vector $\mathbf{x}_k$ belongs to cluster i, $m > 1$ is the fuzzyfication parameter, and d_{ik} represents the distance (any inner product norm defined by a symmetrical positive definite matrix A) between vector $\mathbf{x}_k$ and cluster prototype $\mathbf{v}_i$. FCM uses a probabilistic partition, meaning that the fuzzy memberships of any input vector $\mathbf{x}_k$ with respect to classes satisfy the probability constraint

$$\sum_{i=1}^{c} u_{ik} = 1. \quad (2)$$

The minimization of the objective function J_{FCM} is achieved by alternately applying the optimization of J_{FCM} over $\{u_{ik}\}$ with $\mathbf{v}_i$ fixed, $i = 1\ldots c$, and the optimization of J_{FCM} over $\{\mathbf{v}_i\}$ with u_{ik} fixed, $i = 1\ldots c$, $k = 1\ldots n$ [3]. During each iteration, the optimal values are deduced from the zero gradient conditions and Lagrange multipliers, and obtained as follows:

$$u_{ik}^\star = \frac{d_{ik}^{-2/(m-1)}}{\sum_{j=1}^{c} d_{jk}^{-2/(m-1)}} \quad \forall\ i = 1\ldots c,\ \forall\ k = 1\ldots n, \quad (3)$$

$$\mathbf{v}_i^\star = \frac{\sum_{k=1}^{n} u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^{n} u_{ik}^m} \quad \forall\ i = 1\ldots c. \quad (4)$$

According to the AO scheme of the FCM algorithm, Eqs. (3) and (4) are alternately applied, until cluster prototypes stabilize. This stopping criterion compares the sum of norms of the variations of the prototype vectors $\mathbf{v}_i$ within the latest iteration with a predefined small threshold value ε.

Hard c-means is a special case of FCM, which uses $m=1$, and thus the memberships are obtained by the winner-takes-all rule. Each cluster prototype will be the average of the input vectors assigned to the given cluster.

2.2. FCM with improved partition

Partitions provided by FCM have an undesired property, which is visualized in Fig. 1 for the case of a one-dimensional problem and $c=3$ clusters. In the proximity of the boundary between two neighbor clusters, a zero fuzzy membership value for the third cluster would be appreciated. Instead of that, these fuzzy membership functions with respect to any third cluster have elevated values at the boundary of the other two clusters. Due to this behavior, Hoppner and Klawonn [17] called the FCM partition multimodal. They also pointed out the fact that the multimodality could be suppressed by reducing the fuzzy exponent m, but in most cases that is not desired because it also influences the fuzzyness of the algorithm. To avoid or at least suppress this multimodality, they introduced the so-called FCM with improved partition (IFP-FCM), which is derived from an objective function that additionally contains a rewarding term:

$$J_{\mathrm{IFP-FCM}} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m d_{ik}^2 - \sum_{k=1}^{n} a_k \sum_{i=1}^{c} (u_{ik} - 1/2)^2, \quad (5)$$

where parameters a_k are positive numbers. The second term has the effect of pushing the fuzzy membership values u_{ik}, $i = 1\ldots c$, $k = 1\ldots n$ towards 0 or 1, while maintaining the probabilistic constraint.

Later, Zhu et al. [42] introduced a generalized version of this algorithm, derived from the objective function:

$$J_{\mathrm{GIFP-FCM}} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m d_{ik}^2 + \sum_{k=1}^{n} a_k \sum_{i=1}^{c} u_{ik}(1 - u_{ik}^{m-1}), \quad (6)$$

whose optimization leads to the partition update formula

$$u_{ik}^\star = \frac{(d_{ik}^2 - a_k)^{-1/(m-1)}}{\sum_{j=1}^{c} (d_{jk}^2 - a_k)^{-1/(m-1)}} \quad \forall\ i = 1\ldots c,\ \forall\ k = 1\ldots n. \quad (7)$$

Eq. (7) explains to us the behavior of GIFP-FCM: for any input vector $\mathbf{x}_k$, the square of its distances measured from all cluster prototypes is virtually reduced by a constant positive value a_k, and

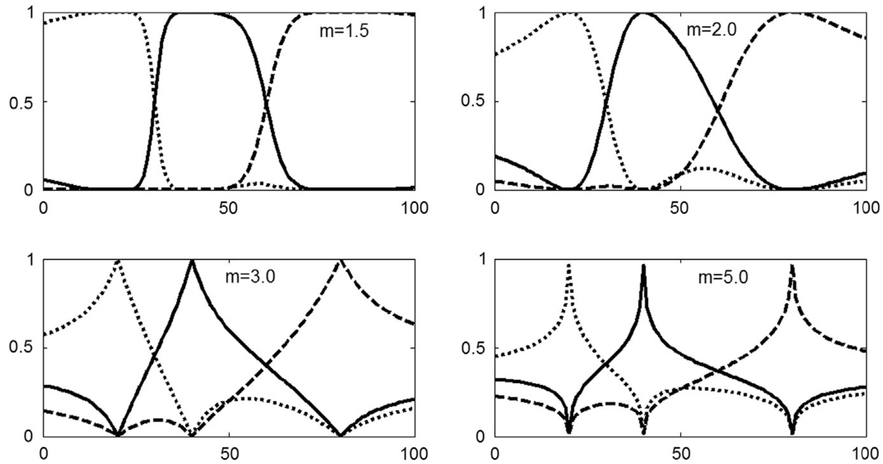

Fig. 1. Multimodal fuzzy membership functions in a single dimensional problem, at various values of the fuzzy exponent m. The phenomenon escalates as the fuzzy exponent grows.

then the partition is updated with the formula of the standard FCM. Zhu et al. [42] also proposed a formula for the choice of a_k:

$$a_k = \omega \min_i \{d_{ik}^2, i = 1 \ldots c\}, \tag{8}$$

with $\omega \in [0.9, 0.99]$, thus keeping the square of all distorted distances positive. Using $\omega = 1$ would reduce GIFP-FCM to HCM.

It is important to remark that both of the above improved clustering models kept FCM's prototype update formula given in Eq. (4).

Lately, these improved FCM clustering models were involved in several applications [6,7].

2.3. The suppressed FCM algorithm

The suppressed fuzzy c-means algorithm was introduced in [12], having the declared goal of reducing the execution time of FCM by improving the convergence speed, while preserving its good classification accuracy. The s-FCM algorithm does not minimize J_{FCM}. Instead of that, it manipulates with the AO scheme of FCM, by inserting an extra computational step in each iteration, placed between the partition update formula (3) and prototype update formula (4). This new step deforms the partition (fuzzy membership functions) according to the following rule:

$$\mu_{ik} = \begin{cases} 1 - \alpha + \alpha u_{ik} & \text{if } i = \arg\max_j \{u_{jk}\} \\ \alpha u_{ik} & \text{otherwise,} \end{cases} \tag{9}$$

where μ_{ik} ($i = 1 \ldots c$, $k = 1 \ldots n$) represents the fuzzy memberships obtained after suppression. During the iterations of s-FCM, these suppressed membership values μ_{ik} will replace u_{ik} in Eq. (4).

Suppression can be explained in words as follows: in each iteration, clusters compete for each input vector, and the prototype situated closest wins the competition. Fuzzy memberships of the given vector with respect to any non-winner class are proportionally suppressed via multiplication with the previously defined value of α, while the winner fuzzy membership is increased such that the modified membership values μ_{ik} still fulfil the probabilistic constraint (2).

In paper [37] we have shown that the proportional suppression of non-winner fuzzy memberships is mathematically equivalent with a virtual reduction of the distance between the winner cluster's prototype and the given input vector. There we proved that in any iteration, for any input vector $\mathbf{x}_k$ and its winner class with index w there exists a virtually reduced distance $\delta_{wk} < d_{wk}$ for

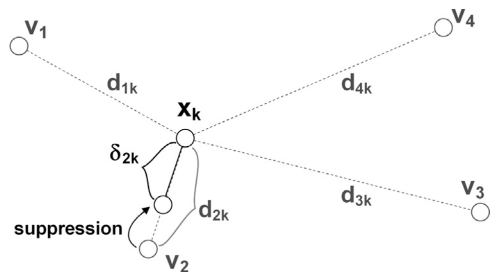

Fig. 2. The effect of suppression: cluster w_k=2 is the winner here, as $\mathbf{v}_{2k}$ is the closest prototype from vector $\mathbf{x}_k$. The virtually reduced distance provides an increased membership degree to the winner cluster, while all non-winner memberships will be proportionally suppressed.

which

$$\mu_{wk} = \frac{\delta_{wk}^{-2/(m-1)}}{\delta_{wk}^{-2/(m-1)} + \sum_{j=1, j \neq w}^{c} d_{jk}^{-2/(m-1)}} \tag{10}$$

and

$$\mu_{ik} = \frac{d_{ik}^{-2/(m-1)}}{\delta_{wk}^{-2/(m-1)} + \sum_{j=1, j \neq w}^{c} d_{jk}^{-2/(m-1)}} \quad \forall i \neq w. \tag{11}$$

This virtual reduction of distance is explained in Fig. 2.

We also defined a quasi-learning rate η of the s-FCM algorithm, in an analogous way to the learning rate of competitive algorithms [37], and deduced its formula:

$$\eta(m, \alpha, u_{wk}) = 1 - \frac{\delta_{wk}}{d_{wk}} = 1 - \left(1 + \frac{1-\alpha}{\alpha u_{wk}}\right)^{(1-m)/2}. \tag{12}$$

The learning rate η depends on two parameters of the s-FCM algorithm (namely m and α), and the winner fuzzy membership function (u_{wk}) of the given input vector $\mathbf{x_k}$.

The literature has shown some of the possible advantages of suppression [12,19–21,32,35,39], motivating us to employ Eq. (12) to define a variety of novel ways of suppression for the FCM algorithm.

3. Proposed methodology

The suppressed FCM algorithm, as proposed by Fan et al. [12], works with a constant suppression rate. Varying the suppression rate could possibly occur in three different ways:

1. *Time variant suppression* means to apply a suppression rate that varies from iteration to iteration as a function of the

iteration count. This was applied, for example, by Hung et al. [19].

2. *Context sensitive* or *data sensitive suppression* means to define a time invariant rule of suppression, which provides a dedicated suppression rate α_k to each input vector $\mathbf{x}_k$.
3. *Time and context variant suppression* means to combine both previous variation models in a single suppression rule.

Considering the fact that suppression aims at achieving a quicker convergence and provides improved partitioning compared to FCM, it

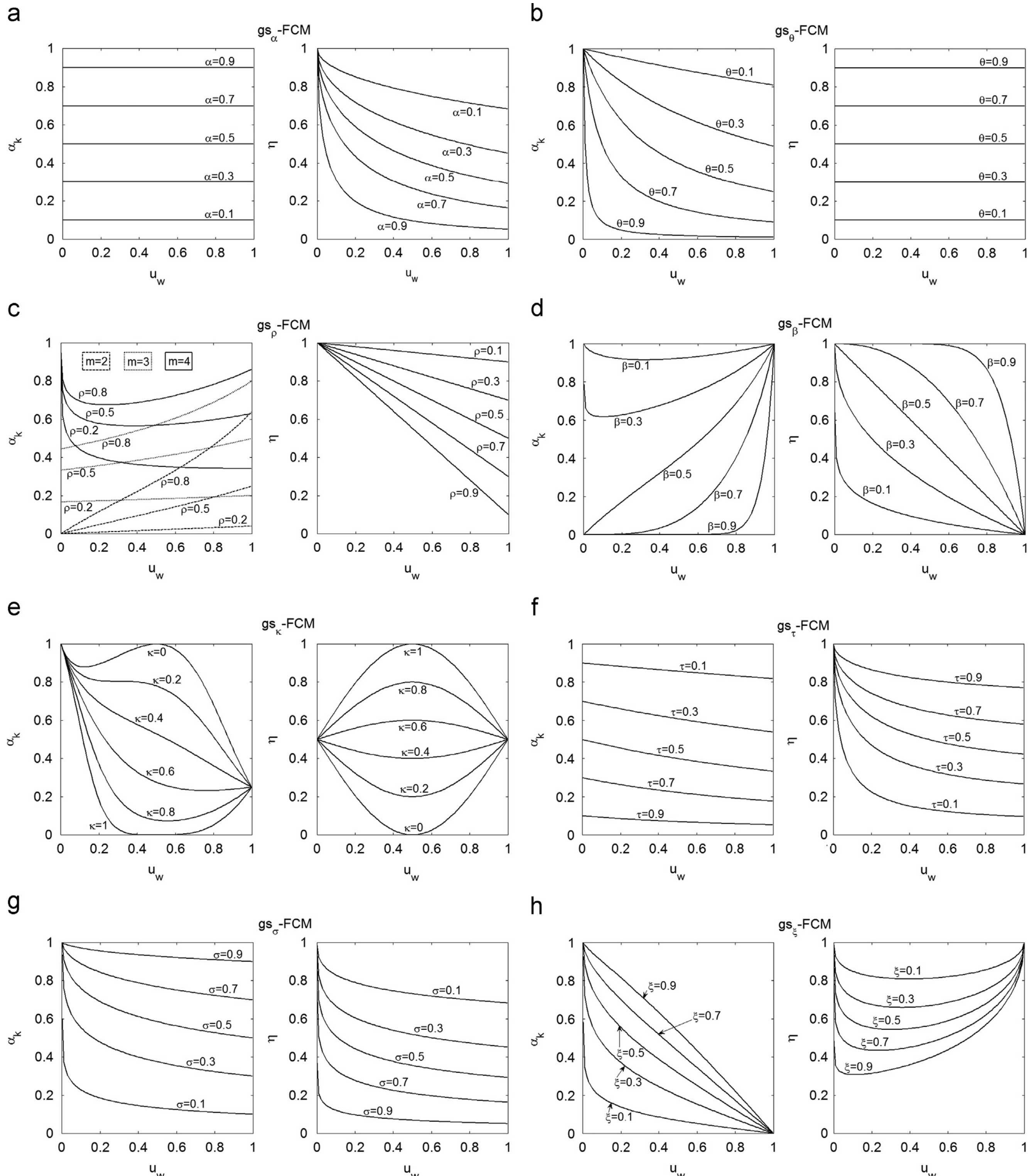

Fig. 3. Characteristics of eight suppression schemes: the evolution of the context dependent suppression rate (α_k) and the quasi-learning rate (η), both plotted against the winner fuzzy membership u_w. Unless otherwise specified, these curves represent the $m=2$ case. Presented suppression schemes are (a) gs$_\alpha$-FCM or conventional s-FCM [12], having constant suppression rate α; (b) gs$_\theta$-FCM uses constant learning rate θ; (c)–(e) suppression schemes with learning rate defined as a function of winner fuzzy membership u_w, namely (c) gs$_\rho$-FCM uses $\eta = 1-\rho u_w$ (left diagram plots characteristics for m=2,3,4 cases), (d) gs$_\beta$-FCM uses $\eta = 1-u_w^{\beta/(1-\beta)}$, (e) gs$_\kappa$-FCM uses $\eta=[1+(2\kappa-1)\sin(\pi u_w)]/2$; (f)–(h) suppression schemes defined by a direct formula between μ_w and u_w, namely (f) gs$_\tau$-FCM uses $\mu_w=(u_w+\tau)/(1+u_w\tau)$, (g) gs$_\sigma$-FCM uses $\mu_w=u_w^\sigma$, (h) gs$_\xi$-FCM uses $\mu_w=\sin^\xi(\pi u_w/2)$.

is not advisable to change the suppression rule in every iteration. This is why in the following, we will exploit the possibilities of the second generalization way, namely we will define some specific suppression rules and will apply them until the convergence is achieved. All algorithms proposed in this section will be called generalized suppressed fuzzy c-means (gs-FCM) clustering, but there will be several types of them.

Throughout this section, we will study algorithms that apply context dependent suppression rules, which employ a dedicated suppression rate α_k for each input vector $\mathbf{x}_k$. The value of α_k never depends on other input vectors $\mathbf{x}_l$, $k \neq l$. That is why, for the sake of simplicity, we will denote by u_w (instead of u_{wk} (or the even more precise $u_{w_k k}$, as w also depends on k)) the largest fuzzy membership function of vector, $\mathbf{x}_k$, before suppression.

Fig. 3 presents the characteristics of eight types of suppressed FCM algorithms, the first of which—shown in Fig. 3(a) is the conventional version introduced by Fan et al. [12]. Each of Fig. 3 (a)–(h) contains two subplots: on the left side the suppression rate α_k is plotted against the winner fuzzy membership u_w, while on the right side the dependency between the quasi-learning rate η and winner fuzzy membership function u_w is indicated. The algorithms presented in Fig. 3(b)–(h) will be described in the following.

3.1. Learning rate defined as a function of the winner fuzzy membership

In this section we will suppose that the QLR varies according to a function of the winner membership: $\eta = f(u_w)$, where $f : [0,1] \to [0,1]$ is a continuous function. In this case, the context dependent suppression rate becomes

$$\eta = f(u_w) \Rightarrow 1 - \left(1 + \frac{1-\alpha_k}{\alpha_k u_w}\right)^{(1-m)/2} = f(u_w)$$
$$\Rightarrow \frac{1-\alpha_k}{\alpha_k u_w} = (1 - f(u_w))^{2/(1-m)} - 1$$
$$\Rightarrow \alpha_k = [1 - u_w + u_w(1 - f(u_w))^{2/(1-m)}]^{-1}. \tag{13}$$

For the special case of constant learning rate we use $\eta = f(u_w) = \theta$ with $\theta \in [0,1]$. A learning rate that linearly decreases with the winner fuzzy membership can be defined by $f(u_w) = 1 - \rho u_w$ with $0 \leq \rho \leq 1$. Let us further define two more such functions: $f(u_w) = 1 - u_w^{\beta/(1-\beta)}$ with parameter $\beta \in [0,1)$ and $f(u_w) = \frac{1}{2}[1 + (2\kappa - 1)\sin(\pi u_w)]$ with parameter $\kappa \in [0,1]$. Each of these functions deduces a suppression rule from Eq. (13). The obtained suppression rules are indicated in the first four rows of Table 1. The algorithms derived from these four suppression rules will be referred to as generalized fuzzy c-means of type θ, ρ, β, and κ.

Fig. 3(b)–(e) exhibits the characteristics of the above-described suppression schemes. On the right side of each figure, we have the variation of the learning rate against the winner fuzzy membership for various values of the parameter showing the definition of the given generalized suppression rule. On the left side of each figure, the variation of the obtained suppression rate α_k against the winner fuzzy membership u_w is displayed, under various constraints. Although they stem from the same equation, the suppression rules give different characteristics in this graphical representation.

3.2. Direct formula between μ_w and u_w

According to this approach, we may formulate a direct dependence rule between the winner fuzzy membership before and after suppression. In the general case, we may write $\mu_w = g(u_w)$ with $g : [0,1] \to [0,1]$ and $g(x) \geq x \forall x \in [1/c, 1]$. Eq. (9) allows us to write

$$\mu_w = g(u_w) \Rightarrow 1 - \alpha_k + \alpha_k u_w = g(u_w)$$
$$\Rightarrow \alpha_k(1 - u_w) = 1 - g(u_w)$$
$$\Rightarrow \alpha_k = \frac{1 - g(u_w)}{1 - u_w}. \tag{14}$$

Let us write a few such direct relations. For example, let us employ $\mu_w = u_w^\sigma$ with parameter $\sigma \in [0,1]$, and $\mu_w = [\sin(\pi u_w/2)]^\xi$ with parameter $\xi \in [0,1]$. The obtained suppression rules are indicated in the last two rows of Table 1. The algorithms derived from these suppression rules will be referred to as generalized fuzzy c-means of type σ and ξ. Further on, let us derive the generalized fuzzy c-means of type τ from the following formula inspired by the relativistic speed addition:

$$\mu_w = \frac{u_w + \tau}{1 + u_w \tau} \Rightarrow 1 - \alpha_k + \alpha_k u_w = \frac{u_w + \tau}{1 + u_w \tau}$$
$$\Rightarrow \alpha_k(1 - u_w) = 1 - \frac{u_w + \tau}{1 + u_w \tau}$$
$$\Rightarrow \alpha_k = \frac{1 + u_w \tau - u_w - \tau}{(1 - u_w)(1 + u_w \tau)}$$
$$\Rightarrow \alpha_k = \frac{1 - \tau}{1 + u_w \tau}, \tag{15}$$

with parameter τ ranging between 0 and 1. The above formula holds for any $u_w < 1$. When $u_w = 1$, the suppression rate is irrelevant. Fig. 3(f)–(h) shows the characteristics of the three different suppression rules defined along direct formulas between μ_w and u_w.

3.3. The gs-FCM algorithm

So far we have introduced seven generalized suppression rules, each governed by a parameter that can take an infinite number of different values. A limited number of these values (one or two of them) reduce the generated algorithm to FCM or HCM, while every other parameter value defines a new clustering model.

Table 1
Proposed generalized suppression rules.

Algorithm	Parameter	Definition	Suppression formula
gs$_\theta$-FCM	$\theta \in [0,1]$	$\eta = \theta$	$\alpha_k = [1 - u_w + u_w(1-\theta)^{2/(1-m)}]^{-1}$
gs$_\rho$-FCM	$\rho \in [0,1]$	$\eta = 1 - \rho u_w$	$\alpha_k = [1 - u_w + \rho^{2/(1-m)} u_w^{(3-m)/(1-m)}]^{-1}$
gs$_\beta$-FCM	$\beta \in [0,1)$	$\eta = 1 - u_w^{\beta/(1-\beta)}$	$\alpha_k = [1 + u_w(u_w^{2\beta/(1-m)(1-\beta)} - 1)]^{-1}$
gs$_\kappa$-FCM	$\kappa \in [0,1]$	$\eta = \frac{1}{2}[1 + (2\kappa - 1)\ \sin(\pi u_w)]$	$\alpha_k = \left[1 - u_w + u_w\left(\frac{1}{2} - \frac{2\kappa - 1}{2}\ \sin(\pi u_w)\right)^{2/(1-m)}\right]^{-1}$
gs$_\tau$-FCM	$\tau \in [0,1]$	$\mu_w = \frac{u_w + \tau}{1 + u_w \tau}$	$\alpha_k = \frac{1-\tau}{1 + u_w \tau}$
gs$_\sigma$-FCM	$\sigma \in [0,1]$	$\mu_w = u_w^\sigma$	$\alpha_k = \frac{1 - u_w^\sigma}{1 - u_w}$
gs$_\xi$-FCM	$\xi \in [0,1]$	$\mu_w = \left(\sin\frac{\pi u_w}{2}\right)^\xi$	$\alpha_k = \frac{1 - \left(\sin\frac{\pi u_w}{2}\right)^\xi}{1 - u_w}$

The suppression rules introduced above are summarized in Table 1. Further suppression rules can be defined similar to the ways presented in Sections 3.1 and 3.2.

The proposed algorithm can be summarized as follows:

1. Set the number of clusters c, and the value of the fuzzy exponent $m>1$.
2. Initialize cluster prototypes according to some intelligent principles (or by selecting random input vectors for each prototype).
3. Choose suppression rule and set the value of its parameter, according to Table 1.
4. Compute fuzzy membership with the conventional formula of FCM, Eq. (3).
5. For each input vector $\mathbf{x_k}$, find the winner cluster, set w equal to the index of the winner cluster, and compute the suppression rate α_k according to the suppression rule, with the corresponding formula from the last column of Table 1.
6. For each input vector $\mathbf{x_k}$, compute suppressed fuzzy memberships with the conventional suppression formula given in Eq. (9), using the suppression rate α_k.
7. Update cluster prototypes using the suppressed fuzzy memberships, as in the original suppressed FCM algorithm.
8. Repeat steps 4–7 until the norm of the variation of the cluster prototypes reduces under a predefined constant ε.

4. The relation between suppressed FCM and GIFP-FCM

In order to unify the theory of FCM with various kinds of improved partitions and the theory of suppressed FCM, let us introduce an objective function, which slightly differs from $J_{\text{GIFP-FCM}}$. The difference consists in the double indexing of reward term's parameter:

$$J_{\text{U}} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m d_{ik}^2 + \sum_{k=1}^{n} a_{ik}\sum_{i=1}^{c} u_{ik}(1-u_{ik}^{m-1}). \tag{16}$$

This objective function allows us to define the reward term in such a way that parameter a_{ik} varies according to the referred cluster i and input vector with index k. The optimization formulas of J_{U} deduced from zero gradient conditions using Lagrange multipliers are

$$u_{ik}^{\star} = \frac{(d_{ik}^2 - a_{ik})^{-1/(m-1)}}{\sum_{j=1}^{c}(d_{jk}^2 - a_{jk})^{-1/(m-1)}} \quad \forall\ i=1\ldots c, \forall\ k=1\ldots n. \tag{17}$$

for partition updating, while the cluster prototypes are updated with the standard FCM's formula given in Eq. (4).

Obviously, J_{U} reduces to $J_{\text{GIFP-FCM}}$ if we choose $a_{1k}=a_{2k}=\cdots=a_{ck}=a_k$ for any $k=1\ldots n$.

On the other hand, we showed that s-FCM and every version of gs-FCM only change one of the distances $d_{1k}, d_{2k}, \ldots, d_{ck}$, for any $k=1\ldots n$. They all virtually reduce the distance between the input vector and the closest (winner) cluster prototype, namely d_{wk} where $w=\arg\min_i\{d_{ik}, i=1\ldots c\}$. Consequently, if we choose

$$a_{ik} = \begin{cases} d_{wk}^2\left[1-\left(1+\frac{1-\alpha_k}{\alpha_k u_{wk}^{(\text{FCM})}}\right)^{1-m}\right] & \text{if } i=w \text{ and } \alpha_k>0 \\ d_{wk}^2 & \text{if } i=w \text{ and } \alpha_k=0 \\ 0 & \text{otherwise} \end{cases}, \tag{18}$$

where $u_{wk}^{(\text{FCM})}$ stands for the winner fuzzy membership value of vector $\mathbf{x}_k$ obtained using the standard FCM partition update formula, then the optimization of J_{U} will perform exactly the same iterations as gs-FCM (or s-FCM) leading to the same results.

Consequently we can affirm that s-FCM and gs-FCM are optimal algorithms to the same extent as GIFP-FCM, as they all optimize J_{U}.

5. Results and discussion

5.1. Evaluation criteria

A series of experimental tests was performed on standard test data sets, which contained ground truth information for evaluation purposes. Our aim was to find those suppression rules, which provide the highest number of correct decisions and best value of cluster validity indexes (CVI). We employed cluster validity

Table 2
Accuracy test results using the WINE data set: for each algorithm and each fuzzy exponent, the average number of correct decisions (out of 178) is exhibited, maximized with respect to the suppression parameter. Algorithms are ranked according to their observed accuracy.

Algorithm	Fuzzy exponent m					
	1.4	1.7	2.0	2.4	2.8	3.2
FCM	170.000	170.000	169.000	166.000	163.000	159.000
gs$_\xi$-FCM	170.072	171.448	171.602	172.022	172.476	172.488
gs$_\kappa$-FCM	170.554	171.480	171.494	171.444	171.538	171.606
gs$_\theta$-FCM	170.510	170.966	170.988	170.694	171.548	171.606
gs$_\tau$-FCM	170.500	170.508	170.532	170.586	170.626	170.704
gs$_\sigma$-FCM	170.502	170.508	170.524	170.578	170.608	170.638
gs$_\alpha$-FCM	170.496	170.486	170.514	170.534	170.584	170.610
gs$_\beta$-FCM	170.486	170.000	170.000	169.828	169.826	169.848
gs$_\rho$-FCM	168.804	168.792	168.790	168.840	169.716	169.770

Table 3
Accuracy test results using the WINE data set: for each algorithm and each fuzzy exponent, the count of parameter values (out of 101) is indicated for which gs-FCM was found more accurate than FCM at the given value of m. Algorithms are ranked according to their observed accuracy.

Algorithm	Fuzzy exponent m					
	1.4	1.7	2.0	2.4	2.8	3.2
gs$_\tau$-FCM	41	41	84	99	99	100
gs$_\xi$-FCM	5	32	84	101	101	101
gs$_\alpha$-FCM	29	30	77	100	100	100
gs$_\sigma$-FCM	28	30	75	100	100	100
gs$_\kappa$-FCM	16	35	74	101	100	101
gs$_\theta$-FCM	15	26	71	99	99	100
gs$_\beta$-FCM	7	0	32	99	99	99
gs$_\rho$-FCM	0	0	0	101	101	101

Table 4
Accuracy test results using the WINE data set: for each algorithm and each fuzzy exponent, the count of parameter values (out of 101) is indicated for which gs-FCM was found more accurate than FCM at its ideal setting. Algorithms are ranked according to their observed accuracy.

Algorithm	Fuzzy exponent m					
	1.4	1.7	2.0	2.4	2.8	3.2
gs$_\tau$-FCM	41	41	40	36	38	39
gs$_\xi$-FCM	5	32	44	51	54	55
gs$_\kappa$-FCM	16	35	44	47	50	56
gs$_\theta$-FCM	15	26	36	43	47	51
gs$_\sigma$-FCM	28	30	23	25	26	27
gs$_\alpha$-FCM	29	30	18	17	19	22
gs$_\beta$-FCM	7	0	0	0	0	0
gs$_\rho$-FCM	0	0	0	0	0	0

indexes like the Xie–Beni index [40] defined as

$$I_{\mathrm{XB}} = \frac{\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{2} \|\mathbf{x}_k - \mathbf{v}_i\|^2}{n(\min_{i \neq j} \|\mathbf{v}_i - \mathbf{v}_j\|)}, \tag{19}$$

its extended version proposed by Pal and Bezdek [33]:

$$I_{\overline{\mathrm{XB}}} = \frac{\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} \|\mathbf{x}_k - \mathbf{v}_i\|^2}{n(\min_{i \neq j} \|\mathbf{v}_i - \mathbf{v}_j\|)}, \tag{20}$$

and the Fukuyama–Sugeno index [13] computed as

$$I_{\mathrm{FS}} = \frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}(\|\mathbf{x}_k - \mathbf{v}_i\|^2 - \|\mathbf{v}_i - \overline{\mathbf{x}}\|^2), \tag{21}$$

where $\overline{\mathbf{x}} = (1/n)\sum_{k=1}^{n}\mathbf{x}_k$ stands for the grand mean of the input vectors. The lowest value of these CVIs indicates the most valid clusters.

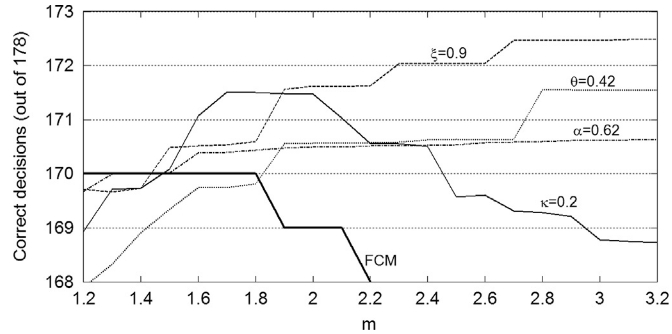

Fig. 5. Results of clustering the normalized WINE data set using various suppression rules. The number of correct decisions (out of 178) is plotted against the value of fuzzy exponent m, in case of some chosen suppression rules and given values of parameters.

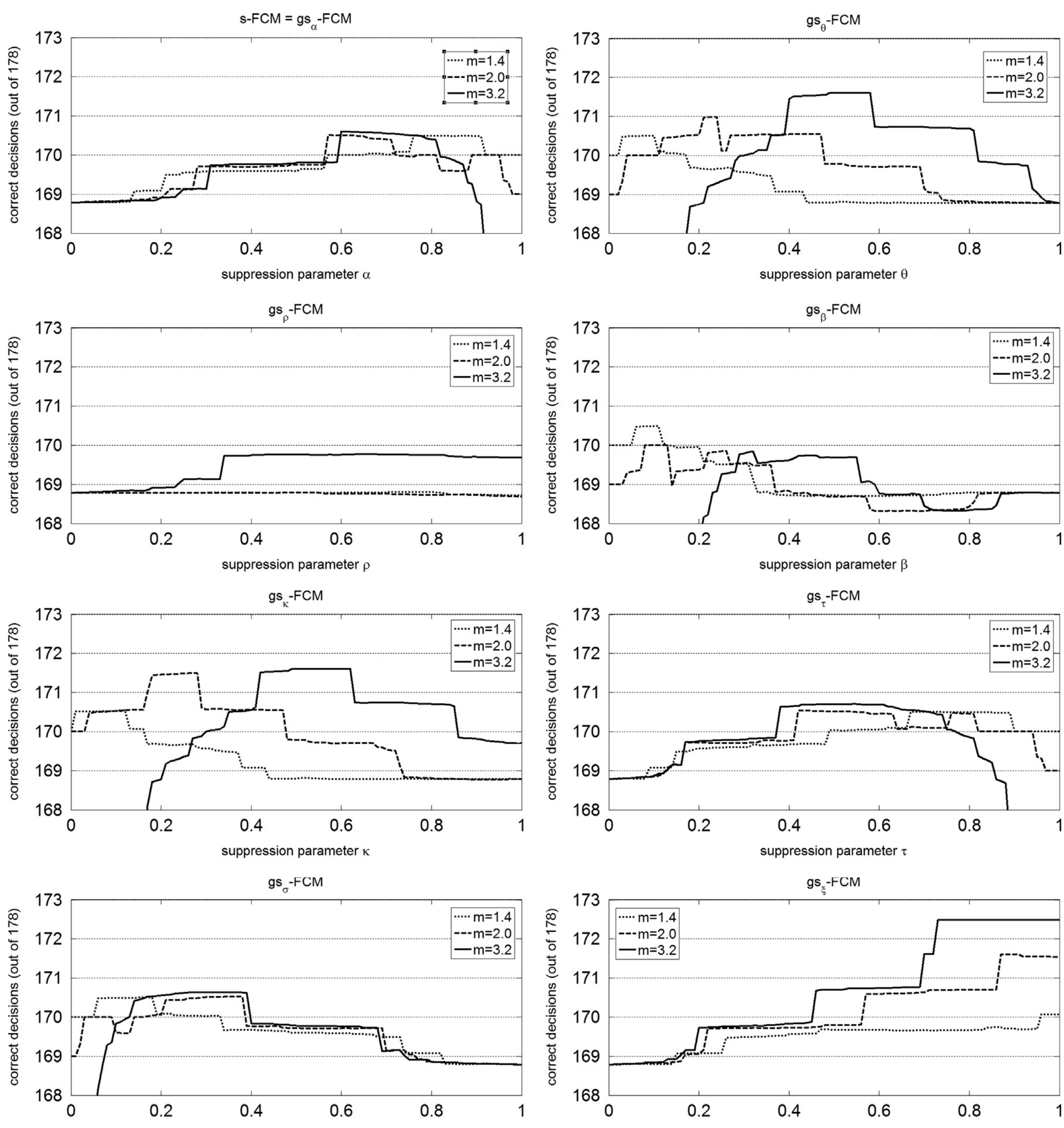

Fig. 4. Results of clustering the normalized WINE data set using various suppression rules. Each image plots the average number of correct decisions (out of 178 input vectors) against the value of the suppression parameter (α, θ, ρ, β, κ, τ, σ, or ξ). Each image contains three graphs, showing the performance of the given gs-FCM algorithm for $m \in \{1.4, 2.0, 3.2\}$.

5.2. Tests on the WINE data set

The proposed generalized suppression schemes underwent detailed numerical tests involving the 13-dimensional WINE data set [2], which contains 178 input vectors, grouped into three classes of a different cardinality. The data set also contains a label for each vector, which is used as ground truth at the evaluation of accuracy. The WINE data vectors were linearly normalized in each dimension into the [0,1] interval.

The first criterion to evaluate accuracy will be the number of correct decisions out of 178. The conventional FCM algorithm, which we use as reference, gives 170 correct decisions for fuzzy exponent $m < 1.85$; as m rises beyond this threshold, the accuracy of FCM drops quickly.

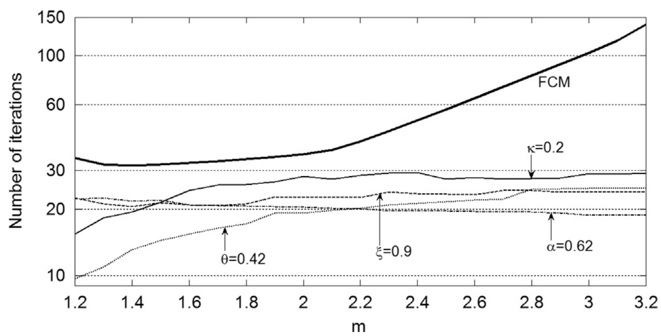

Fig. 6. Convergence speed of various gs-FCM algorithms, compared to FCM. The number of necessary iterations to achieve a predefined convergence level is plotted against the fuzzy exponent m. The vertical axis is logarithmic.

All seven proposed generalized suppression schemes, together with the s-FCM proposed by Fan et al. [12], were tested in various circumstances:

1. various values of the fuzzy exponent $m \in [1.2, 3.2]$, covering the range recommended in [18];
2. suppression parameter ranging from 0 to 1 in steps of 0.01;
3. 500 different, randomly chosen initialization sets of the cluster prototypes, but the same ones for each algorithm and each parameter value. Initialization vectors in all 500 cases were chosen randomly, but the distance between any two vectors in the same set—in the normalized 13-dimensional space—was at least 1.

For each algorithm and each suppression parameter value, we computed the average number of correct decisions (ANCD) obtained from the 500 different runs. Then we selected the maximum value of ANCD for each algorithm, along the domain of definition of the suppression parameter. These values are indicated in Table 2, separately for six different values of the fuzzy exponent. This table ranks the applied suppression rules by their best achieved accuracy. With the exception of suppression rule of type ρ, all proposed algorithms performed better than FCM, and five out of seven generalized suppression rules proved more accurate than the conventional s-FCM.

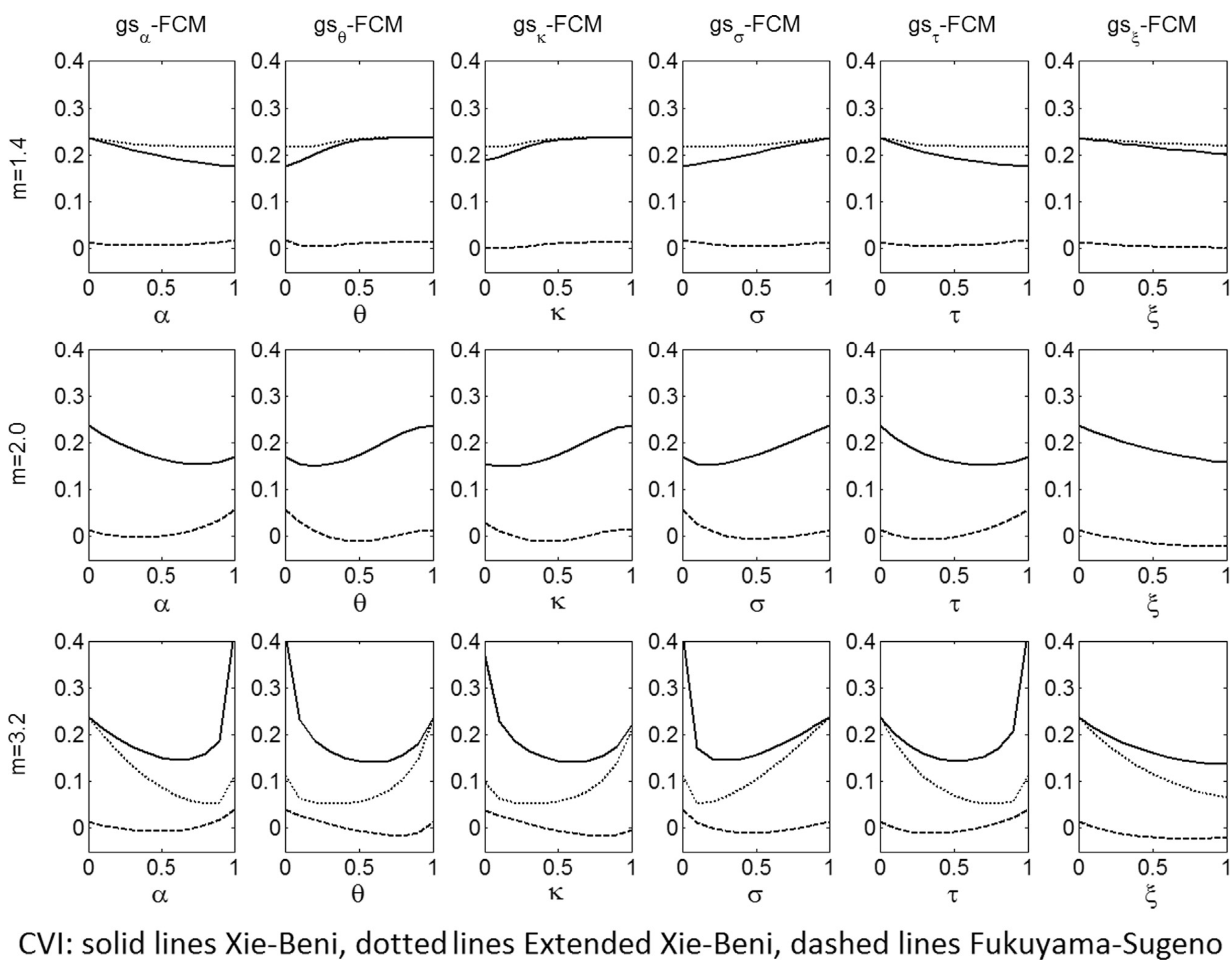

Fig. 7. WINE data clustering results: cluster validity indexes of the six most accurate suppression rules, for various values of fuzzy exponent m, plotted against the suppression parameter of each approach. Most suppression rules have wide ranges of their suppression parameter, for which the clusters are more valid than in FCM.

Tables 3 and 4 summarize the comparison of the accuracy of each suppression rule with the performance of FCM. For each algorithm, the suppression parameter varied from 0 to 1 in steps of 0.01, and the number of correct decisions averaged along the 500 different random initializations was compared with the number of FCM's correct decisions. Table 3 indicates in how many cases (out of 101) these suppression rules proved more accurate than FCM at the given fuzzy exponent m. Table 4 indicates in how many cases (out of 101) these suppression rules had more correct decisions than 170, the highest score reached by FCM under any circumstances. Both these tables rank the suppression rules according to their clustering accuracy.

The s-FCM algorithm, as proposed in [12], proved to be more accurate than FCM, at constant suppression rate $0.6 < \alpha < 0.8$. The difference is most relevant in case of $m > 2$.

The generalized suppression schemes of type τ and σ performed quite similar to s-FCM. Suppressions of type κ, θ and ξ proved to have relevant advantage in accuracy over s-FCM and FCM as well. The highest achieved average number of correct decisions was given by suppression rule of type ξ at $m \approx 3$ and $\xi > 0.75$. The 172.5 correct decisions represent approximately 30% less misclassifications than in case of FCM in its best setting.

On the other hand, suppression rules of type ρ and β led to worse accuracy that FCM, showing that not every generalized suppression rule is useful (Fig. 4).

Fig. 5 exhibits some of the suppression rules and parameter values with the highest accuracy achieved on the WINE data set. These graphs plot the average number of correct decisions of the chosen algorithms against the value of the fuzzy exponent m. FCM visibly becomes weak above $m > 2$, while most of the chosen gs-FCM algorithms have their strongest accuracy in this area.

Fig. 6 relates on the efficiency of the presented algorithms, showing the number of necessary iterations to reach a certain level of convergence ($\varepsilon = 10^{-10}$). One iteration of any gs-FCM is

Table 5
Accuracy of clustering the Breast cancer data: besides the outcome of FCM, we recorded the highest number of correct decisions achieved by each algorithm, along with the intervals where they were achieved, and where they performed better than FCM.

Algorithm	$m=1.5$			$m=2.0$			$m=2.5$			$m=3.0$		
	Correct decisions	Best	Better than FCM	Correct decisions	Best	Better than FCM	Correct decisions	Best	Better than FCM	Correct decisions	Best	Better than FCM
FCM	667			666			667			668		
gs_α-FCM	670	$\alpha < 0.18$	$\alpha < 0.89$	670	$\alpha < 0.16$	$\alpha < 0.98$	670	$\alpha < 0.22$	$\alpha < 0.86$	670	$\alpha < 0.30$	$\alpha < 0.77$
gs_θ-FCM	670	$\theta > 0.48$	$\theta > 0.06$	670	$\theta > 0.60$	$\theta > 0.02$	670	$\theta > 0.80$	$\theta > 0.22$	670	$\theta > 0.90$	$\theta > 0.49$
gs_ρ-FCM	670	$\rho < 0.84$	Any ρ	670	$\rho < 0.77$	Any ρ	670	$\rho < 0.63$	Any ρ	670	$\rho < 0.44$	Any ρ
gs_β-FCM	670	$\beta > 0.58$	$\beta > 0.10$	671	$[0.75, 0.85]$	$\beta > 0.03$	671	$[0.68, 0.86]$	$\beta > 0.18$	671	$[0.70, 0.86]$	$\beta > 0.10$
gs_κ-FCM	670	$\kappa > 0.47$	Any κ	670	$\kappa > 0.60$	Any κ	671	$\kappa > 0.95$	$\kappa > 0.34$	671	$\kappa > 0.94$	$\kappa > 0.52$
gs_σ-FCM	670	$\sigma > 0.74$	$\sigma > 0.07$	670	$\sigma > 0.75$	$\sigma > 0.02$	670	$\sigma > 0.70$	$\sigma > 0.07$	670	$\sigma > 0.67$	$\sigma > 0.17$
gs_τ-FCM	670	$\tau < 0.14$	$\tau < 0.87$	670	$\tau < 0.12$	$\tau < 0.97$	670	$\tau < 0.15$	$\tau < 0.87$	670	$\tau < 0.17$	$\tau < 0.64$
gs_ξ-FCM	670	$\xi < 0.33$	Any ξ	670	$\xi < 0.28$	Any ξ	670	$\xi < 0.22$	Any ξ	670	$\xi < 0.23$	$\xi < 0.85$

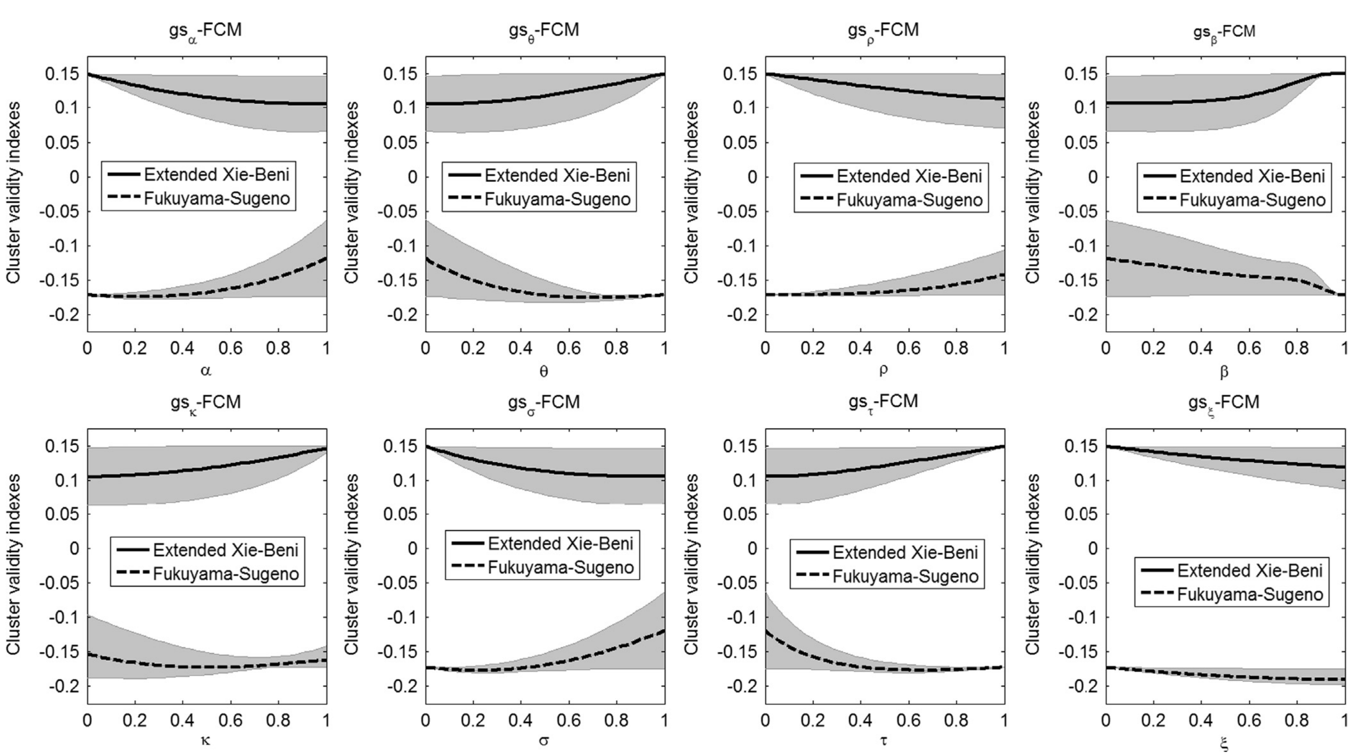

Fig. 8. Cluster validity indexes obtained from Breast cancer data clustering: for each studied suppression rule, black curves represent the average value of CVI computed along the valid range of fuzzy exponent m, while the gray area indicates the minimum–maximum range of each CVI.

computationally more costly than in FCM, because they perform extra computations to suppress the partition. However, the 30–80% less iterations performed by the gs-FCM schemes assure a shorter overall execution time for the proposed algorithms, especially for fuzzy exponents $m > 2$.

Fig. 7 is a repository of CVIs obtained for the 6 best performing suppression rules and three values of the fuzzy exponent m, along the whole range of each suppression parameter. Each of the 18 subplots contains three curves: the solid, dotted, and dashed lines indicate the values of Xie–Beni, extended Xie–Beni, and Fukuyama–Sugeno indexes, respectively. For $m=2.0$ the two Xie–Beni indexes coincide. Clusters are more valid when the CVI value is low. The most important thing reflected by Fig. 7 is that FCM (equivalent to $\alpha = 1$, $\theta = 0$, $\sigma = 1$ or $\tau = 0$) produces the least valid clusters. Further on, suppressing FCM apparently has a beneficial effect on cluster validity, mostly at larger values of m. The best overall cluster validity belongs to the gs_ξ–FCM clustering model, but it does not have a significant lead over the others.

5.3. Tests on breast cancer data

We also tested the proposed generalized suppression schemes to cluster the 9-dimensional vectors of the Breast Cancer Wisconsin data set [2], which contains 699 input vectors, grouped into two classes of a different cardinality. Although in each dimension, the vectors range between 1 and 10, we applied linear normalization to the [0,1] interval. All proposed suppression schemes were tested in various circumstances: fuzzy exponent varying in range 1.2–3.0 with steps of 0.1, suppression parameters covering their whole range in steps of 0.01, and 10 different random initialization scenarios chosen in the same way as in WINE data sets.

The conventional FCM algorithm used as reference gives 666–668 correct decisions out of 699 depending on the value of the fuzzy exponent. Table 5 summarizes the accuracy report of each suppression rule, giving the best achieved number of correct decisions, the range of parameter value where this best score was achieved, and the interval where the suppressed algorithm performed better than FCM. Apparently all tested algorithms have wide ranges of their suppression parameter where they show better accuracy than FCM. The lowest misclassification rates are achieved by suppressions of type β and κ.

Fig. 8 reflects the measure of cluster validity for all tested suppression rules, obtained from the analysis of breast cancer data. For each algorithm and each suppression parameter value, we extracted the minimum, maximum, and average values of the CVIs, obtained for various fuzzy exponent values in range 1.2–3. Again in this case, the most valid clusters were produced by suppression rule ξ. Apparently all suppression schemes lead to clusters with comparable validity to those of FCM.

5.4. Tests on large amount of data

The proposed gs-FCM clustering models underwent some tests using a specially constructed artificial data set. These tests are intended to reveal some advantages of suppression.

The data set contains $n = 10{,}000$ two-dimensional vectors, having normal distribution around three fixed vectors: n_1 vectors around $\nu_{10} = (0, -2)$, n_2 vectors around $\nu_{20} = (0, 2)$, and n_3 vectors around $\nu_{30} = (10, 0)$. We used $n_3 \ll n_1 = n_2$. The input vectors are presented in Fig. 9(a) for the $n_3 = 20$ case. These vectors will be grouped into $c=3$ clusters, using FCM and the proposed algorithms.

Fuzzy c-means was applied first to show the adverse effect of the multimodal fuzzy membership functions. Vectors situated on the boundary between clusters 1 and 2 have a considerable (non-zero) membership with respect to the third cluster. Because

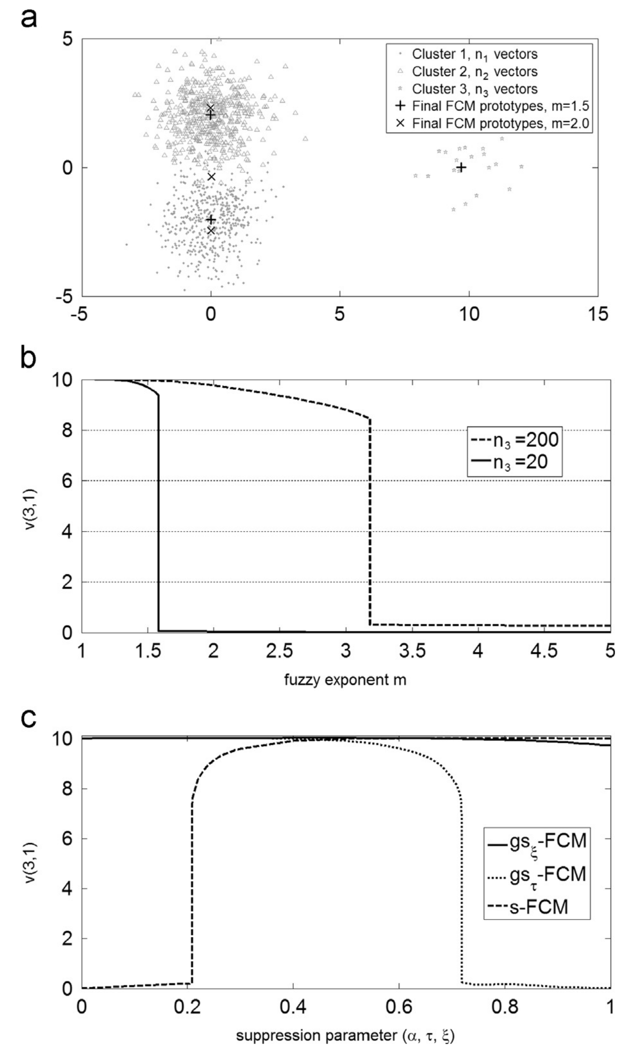

Fig. 9. 2D clustering problem with $n = 10^4$ input vectors, two large and overlapping clusters and a third, distant and small one: (a) FCM clustering fails to create $c=3$ correct clusters above a certain limit of the fuzzy exponent m. (b) The abscissa of the final third cluster prototype, given by FCM, plotted against fuzzy exponent m. The two curves clearly show under what constraints of m can FCM distinguish a small cluster of $n_3 = 20$ or $n_3 = 200$ elements from the other two, large clusters. (c) The abscissa of the final third cluster prototype, given by three different gs-FCM variants, at fuzzy exponent $m=5$. Generalized suppressed FCM of type ξaccurately solves the problem for any value of its suppression parameter.

there are several such input vectors, in a number comparable or possibly greater than n_3, they attract the prototype of the third cluster. The studied parameter will be the final horizontal coordinate of the third cluster's prototype ν_3, which in ideal conditions should be 10. FCM was employed using several values of the fuzzy exponent. Results are exhibited in Fig. 9(b). When $n_3 = 200$ vectors are in the distant, smaller cluster, FCM can distinguish this third cluster from the other two as long as the fuzzy exponent is $m < 3.183$. Even below this limit, cluster prototype ν_3 drifts towards the other two as m grows. Above this limit, all three cluster prototypes will be in the left sided spot of $n - n_3 = 9800$ vectors. When we choose $n_3 = 20$, the limit value of the fuzzy exponent is 1.585: Fig. 9(a) indicates the final cluster prototypes given by FCM at $m=1.5$ and $m=2$.

The situation is completely different if we suppress the FCM partition. In any suppressed partition, the effects of the multimodality are reduced. Fig. 9(c) shows the horizontal position of the third cluster obtained by three different suppression rules at extremely high value of the fuzzy exponent $m=5$. For this given problem, s-FCM (gs_α-FCM) can distinguish the three clusters if the suppression rate is set to $\alpha<0.718$, gs_τ-FCM does the same at $\tau>0.209$, while gs_ξ-FCM creates fine partition in the whole range of its suppression parameter. This way suppression helps small separated groups to form an individual cluster, instead of being merged with other larger groups.

5.5. Suppressed c-means partitions in image segmentation

Willing to demonstrate the partitions and fuzzy membership functions that we can obtain, the proposed gs-FCM variants were also tested for the segmentation of single channel intensity images. An image whose intensity histogram contained three relevant peaks was chosen as input data. Fig. 10 shows the variety of fuzzy membership functions achieved with the tested algorithms, for pixel intensities in range of 0–255. Fig. 10(a) and (b) shows the outcome of FCM, and s-FCM with constant $\alpha=0.5$. The largest membership function in s-FCM always has a granted lead of $1-\alpha$ above any non-winner membership. Non-winner memberships are reduced proportionally and no other deformations of the curves are present. However, with the introduction of the gs-FCM algorithms, a wide variety of membership functions become available. Two of them are demonstrated in Fig. 10(c) and (d). The choice of the suppression rule and the value of the suppression parameter have great influence on the shape of the curves. The multimodal behavior of fuzzy membership functions is effectively reduced by each suppression rule.

The set of suppression rules introduced in this paper is hardly complete. We have only introduced a handful of suppression rules using two different recipes. Anyone can define further suppression rules, the recipe being given in Sections 3.1 and 3.2. In fact, some variants of the gs-FCM algorithm are already present in the literature [28], mostly without being identified as such. Details in this matter we have reported in [38].

5.6. Further remarks

The clustering models introduced in this paper are fully compatible with all extensions and applications of the fuzzy c-means algorithm, as long as the alternate optimization scheme of FCM is broken only by the suppression. This includes, for example, solutions like

1. efficient implementation via aggregation of similar data [5,36,14], or without storing the partition matrix [27];
2. fuzzy c-shell variants [9,29] and derivations [4,16] for the detection of clusters with certain shapes.

6. Conclusion

In this paper we introduced several generalization schemes for the suppressed fuzzy c-means algorithm. Each scheme has a suppression parameter defined in the range between 0 and 1, and each of these values leads to a different clustering algorithm. We demonstrated the advantages of certain suppression models in a multidimensional environment, both in terms of accuracy and efficiency of the partitioning. We also demonstrated how the generalized suppressed FCM combats the multimodality of FCM's fuzzy partition. We proved the correctness of the results using several cluster validity indexes. We identified existing, reportedly accurate gs-FCM clustering models employed in image processing problems, which had been introduced along a completely different rationale. We introduced a generalized objective function using which we unified the theories of FCM algorithms with improved partition [17,42] and the suppressed FCM clustering models [12,37].

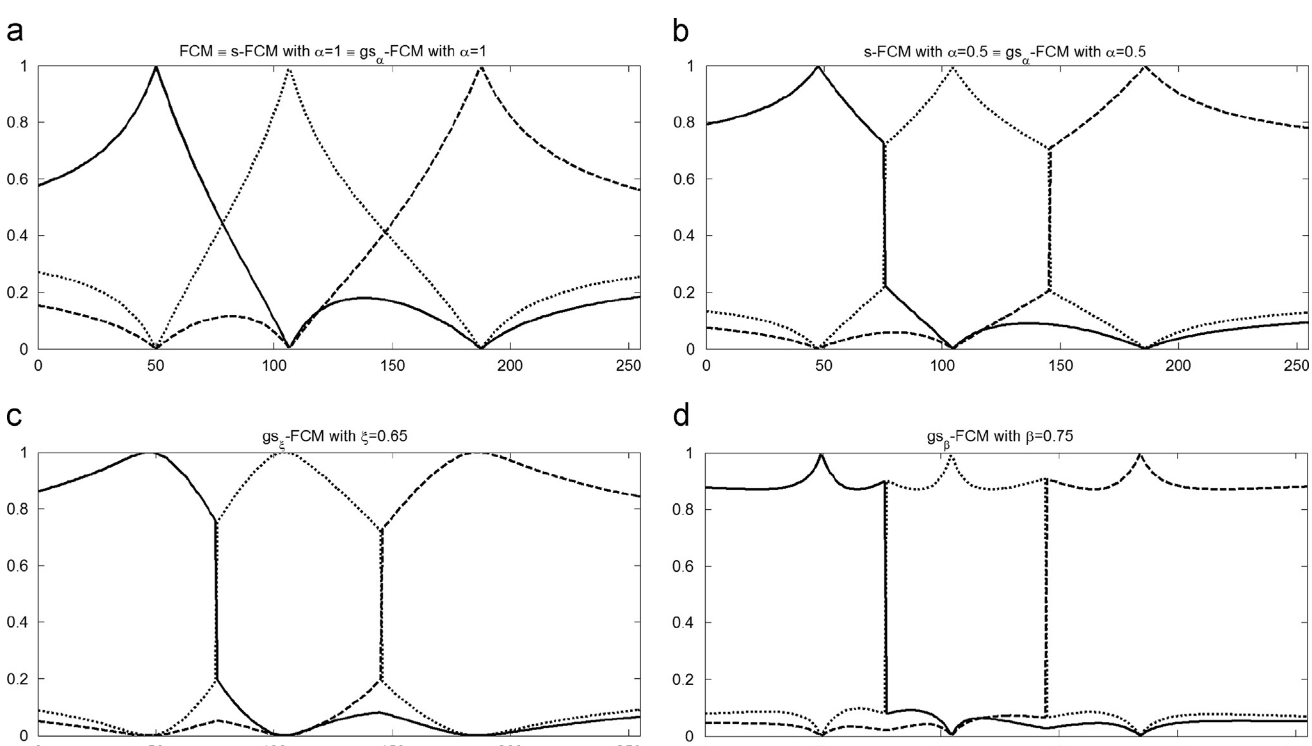

Fig. 10. Fuzzy membership functions given by suppressed FCM variants on a single channel intensity image: (a) FCM; (b) s-FCM at $\alpha=0.5$; (c) gs-FCM of type ξ at $\xi=0.65$; (d) gs-FCM of type β at $\beta=0.75$. All these curves were obtained using fuzzy exponent $m=3$. All suppression rules subdue the multimodality of the fuzzy membership functions.

This novel objective function also serves as evidence for the optimality of the suppressed FCM algorithm variants. The easily implementable optimal suppressed FCM variants will surely gain several applications in the coming years.

Acknowledgments

The work of L. Szilágyi was funded by UEFISCDI under Grant no. PD667, Contract no. 28/05.08.2010, with the exception of Sections 2.2 and 4, which are the products of Grant no. PD103921 funded by the Hungarian National Research Fund (OTKA). The work of S.M. Szilágyi was funded by the project "Transnational Network for Integrated Management of Postdoctoral Research in Communicating Sciences. Institutional building (postdoctoral school) and fellowships program (CommScie)"- POSDRU/89/1.5/S/63663, financed under the Sectoral Operational Program Human Resources Development 2007–2013.

Appendix A. Supplementary material

Supplementary data associated with this paper can be found in the online version at http://dx.doi.org/10.1016/j.neucom.2014.02.027.

References

[1] D. Arthur, S. Vassilvitskii, *k*-means++: the advantages of careful seeding, in: Proceedings of Symposium on Discrete Algorithm, 2007, pp. 1027–1035.
[2] A. Asuncion, D.J. Newman, UCI Machine Learning Repository, available at: ⟨http://archive.ics.uci.edu/ml/datasets.html⟩.
[3] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York, NY, 1981.
[4] J.C. Bezdek, R.J. Hathaway, N.R. Pal, Norm induced shell prototype (NISP) clustering, Neural Parallel Sci. Comput. 3 (1995) 431–450.
[5] R.L. Cannon, J.V. Dave, J.C. Bezdek, Efficient implementation of the fuzzy *c*-means clustering algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 8 (1986) 248–255.
[6] A. Celikyilmaz, I.B. Türkşen, Enhanced fuzzy system models with improved fuzzy clustering algorithm, IEEE Trans. Fuzzy Syst. 16 (2008) 779–794.
[7] A. Celikyilmaz, I.B. Türkşen, R. Aktaş, M.M. Doganay, N.B. Ceylan, Increasing accuracy of two-class pattern recognition with enhanced fuzzy functions, Expert Syst. Appl. 36 (2009) 1337–1354.
[8] T.W. Cheng, D.B. Goldgof, L.O. Hall, Fast fuzzy clustering, Fuzzy Sets Syst. 93 (1998) 49–56.
[9] R.N. Davé, K. Bhaswan, Adaptive fuzzy c-shells clustering and detection of ellipses, IEEE Trans. Neural Netw. 3 (1992) 643–662.
[10] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters, J. Cybern. 3 (1974) 32–57.
[11] S. Eschrich, J. Ke, L.O. Hall, D.B. Goldgof, Fast accurate fuzzy clustering through data reduction, IEEE Trans. Fuzzy Syst. 11 (2003) 262–270.
[12] J.L. Fan, W.Z. Zhen, W.X. Xie, Suppressed fuzzy *c*-means clustering algorithm, Pattern Recognit. Lett. 24 (2003) 1607–1612.
[13] Y. Fukuyama, M. Sugeno, A new method of choosing the number of clusters for the fuzzy *c*-means method (in Japanese), in: Proceedings of the Fifth Fuzzy Systems Symposium, 1989, pp. 247–250.
[14] R.J. Hathaway, J.C. Bezdek, Extending fuzzy and probabilistic clustering to very large data sets, Comput. Stat. Data Anal. 51 (2006) 215–234.
[15] T.C. Havens, J.C. Bezdek, C. Leckie, L.O. Hall, M. Palaniswami, Fuzzy *c*-means algorithms for very large data, IEEE Trans. Fuzzy Syst. 20 (2012) 1130–1146.
[16] F. Hoppner, Fuzzy shell clustering algorithms in image processing: fuzzy c-rectangular and 2-rectangular shells, IEEE Trans. Fuzzy Syst. 5 (1997) 599–613.
[17] F. Hoppner, F. Klawonn, Improved fuzzy partitions for fuzzy regression models, Int. J. Approx. Reason. 32 (2003) 85–102.
[18] M. Huang, Z. Xia, H. Wang, Q. Zeng, Q. Wang, The range of the value for the fuzzifier of the fuzzy *c*-means algorithm, Pattern Recognit. Lett. 33 (2012) 2280–2284.
[19] W.L. Hung, M.S. Yang, D.H. Chen, Parameter selection for suppressed fuzzy *c*-means with an application to MRI segmentation, Pattern Recognit. Lett. 27 (2006) 424–438.
[20] W.L. Hung, Y.C. Chang, A modified fuzzy *c*-means algorithm for differentiation in MRI of ophtalmology, in: Lecture Notes in Computer Science (MDAI'06, Tarragona), vol. 3885, 2006, pp. 340–350.
[21] W.L. Hung, D.H. Chen, M.S. Yang, Suppressed fuzzy-soft learning vector quantization for MRI segmentation, Artif. Intell. Med. 52 (2011) 33–43.
[22] R. Huth, C. Beck, A. Philipp, M. Demuzere, Z. Ustrnul, M. Cahynová, J. Kyselý, O.E. Tveito, Classifications of atmospheric circulation patterns: recent advances and applications, Ann. N.Y. Acad. Sci. 1146 (2008) 105–152.
[23] H. Izakian, A. Abraham, Fuzzy *c*-means and fuzzy swarm for fuzzy clustering problem, Expert Syst. Appl. 38 (2011) 1835–1838.
[24] L.X. Jun, W.Y. Xin, Q.J. Guang, W.L. Hui, The application of fuzzy *c*-means clustering in macro-economic forecast, in: Proceedings of the Second International Symposium on Electronic Commerce and Security (ISECS2009, Nanchang, China), 2009, pp. 609–611.
[25] M.S. Kamel, S.Z. Selim, New algorithms for solving the fuzzy clustering problem, Pattern Recognit. 27 (1994) 421–428.
[26] T. Kohonen, The self-organizing map, Proc. IEEE 78 (1990) 1474–1480.
[27] J.F. Kolen, T. Hutcheson, Reducing the time complexity of the fuzzy *c*-means algorithm, IEEE Trans. Fuzzy Syst. 10 (2002) 263–267.
[28] S. Krinidis, V. Chatzis, A robust fuzzy local information *c*-means clustering algorithm, IEEE Trans. Image Process. 19 (2010) 1328–1337.
[29] R. Krishnapuram, O. Nasraoui, H. Frigui, A fuzzy *c* spherical shells algorithm: a new approach, IEEE Trans. Neural Netw. 3 (1992) 663–671.
[30] J. Lázaro, J. Arias, J.L. Martín, C. Cuadrado, A. Astarloa, Implementation of a modified fuzzy *c*-means clustering algorithm for real-time applications, Microprocess. Microsyst. 29 (2005) 375–380.
[31] S. McQueen, Some methods for classification and analysis of multivariate observations, in: The Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297.
[32] A. Nyma, M. Kang, Y.K. Kwon, C.H. Kim, J.M. Kim, A hybrid technique for medical image segmentation, J. Biomed. Biotechnol. (2012) 1–7 (Article ID 830252).
[33] N.R. Pal, J.C. Bezdek, On cluster validity for the fuzzy *c*-means model, IEEE Trans. Fuzzy Syst. 3 (1995) 370–379.
[34] E.H. Ruspini, A new approach to clustering, Inf. Control 16 (1969) 22–32.
[35] M.F. Saad, A.M. Alimi, Improved modified suppressed fuzzy *c*-means, in: Proceedings of the Second International Conference on Image Processing Theory Tool and Applications (IPTA2010, Paris), 2010, pp. 313–318.
[36] L. Szilágyi, Z. Benyó, S.M. Szilágyi, H.S. Adam, MR brain image segmentation using an enhanced fuzzy *c*-means algorithm, in: Proceedings of the 25th Annual International Conference of IEEE EMBC (Cancún), 2003, pp. 724–726.
[37] L. Szilágyi, S.M. Szilágyi, Z. Benyó, Analytical and numerical evaluation of the suppressed fuzzy *c*-means algorithm: a study on the competition in *c*-means clustering models, Softw. Comput. 14 (2010) 495–505.
[38] L. Szilágyi, Lessons to learn from a mistaken optimization, Pattern Recognit. Lett. 36 (2014) 29–35, http://dx.doi.org/10.1016/j.patrec.2013.08.027.
[39] H.S. Tsai, W.L. Hung, M.S. Yang, A robust kernel-based fuzzy *c*-means algorithm by incorporating suppressed and magnified membership for MRI image segmentation, in: Lecture Notes in Artificial Intelligence in Medicine, vol. 7530, 2012, pp. 744–754.
[40] X.L. Xie, G.A. Beni, Validity measure for fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 3 (1991) 841–846.
[41] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (1965) 338–353.
[42] L. Zhu, F.L. Chung, S. Wang, Generalized fuzzy *c*-means clustering algorithm with improved partitions, IEEE Trans. Syst. Man Cybern. B. 39 (2009) 578–591.

László Szilágyi received M.S.c in electrical engineering, in 1998, from Petru Maior University of Tîrgu Mureş, Romania, and Ph.D., in 2009, from Budapest University of Technology and Economics, Hungary. Currently he is an associate professor at Sapientia University of Transylvania, Tîrgu Mureş, Romania, where he teaches Image Processing and Pattern Recognition, Modeling and Simulation, and Parallel Algorithms. His research topics include medical image processing, artificial intelligence, and unsupervised learning methods in pattern recognition. He has published two books, 10+ journal papers, 20+ book chapters and 60+ conference papers. He is member of IEEE and IAPR's Hungarian Section. He was awarded the János Bolyai Fellowship Prize by the Hungarian Academy of Sciences, in 2010.

Sándor M. Szilágyi received M.Sc. in electrical engineering, in 1996, from Petru Maior University of Tîrgu Mureş, Romania, and Ph.D., in 2008, from Budapest University of Technology and Economics, Hungary. Currently he is an associate professor at Petru Maior University of Tîrgu Mures, Romania, where he teaches Software Engineering, Low-level Programming, and Biomedical Systems. His research topics include ECG signal processing and dynamic heart modeling, and various applications of pattern recognition. He has published one book, 10+ journal papers, 15+ book chapters and 50+ conference papers. He was awarded János Bolyai Fellowship Prize by the Hungarian Academy of Sciences, in 2011.