A review on suppressed fuzzy c-means clustering models

László SZILÁGYI
Sapientia Hungarian University of Transylvania,
Cluj-Napoca, Romania
Dept. of Electrical Engineering, Târgu Mureş
Óbuda University, Budapest, Hungary
University Research, Innovation and Service Center
email: lalo@ms.sapientia.ro
szilagyi.laszlo@nik.uni-obuda.hu

László LEFKOVITS
Sapientia Hungarian University of Transylvania,
Cluj-Napoca, Romania
Dept. of Electrical Engineering, Târgu Mureş
email: lefkolaci@ms.sapientia.ro

David ICLANZAN
Sapientia Hungarian University of Transylvania,
Cluj-Napoca, Romania
Dept. of Mathematics-Informatics, Târgu Mureş
email: iclanzan@ms.sapientia.ro

Abstract. Suppressed fuzzy c-means clustering was proposed as an attempt to combine the better properties of hard and fuzzy c-means clustering, namely the quicker convergence of the former and the finer partition quality of the latter. In the meantime, it became much more than that. Its competitive behavior was revealed, based on which it received two generalization schemes. It was found a close relative of the so-called fuzzy

c-means algorithm with generalized improved partition, which could improve its popularity due to the existence of an objective function it optimizes. Using certain suppression rules, it was found more accurate and efficient than the conventional fuzzy c-means in several, mostly image processing applications. This paper reviews the most relevant extensions and generalizations added to the theory of fuzzy c-means clustering models with suppressed partitions, and summarizes the practical advances these algorithms can offer.

1   Introduction

C-means clustering algorithms represent a subset of the objective function optimizer clustering methods, which group a set of object data into a set of predefined number of clusters. Chronologically, the first c-means clustering algorithm is hard c-means (HCM) having its origins in the works of Steinhaus [34] and McQueen [23], also known as k-means, which uses bivalent (crisp) logic to represent the created partition, namely it assigns each object to a single cluster. The fuzzy c-means (FCM) algorithm was introduced by Dunn [7], and generalized by Bezdek [2]. FCM uses a probabilistic partition to create the clusters: for any object, the sum of the fuzzy memberships with respect to all clusters is always one.

Both HCM and FCM have certain limitations. HCM converges quickly but it is very sensitive to initialization [1], and frequently gives mediocre partitions because it gets stuck in local minima of the objective function. On the other hand, FCM has a slower convergence, which becomes a problem when the input data is huge. Despite these limitations, HCM and FCM are very popular algorithms, having lots of applications in various research domains.

Several solutions have been proposed to reduce the runtime of the FCM algorithm, without damaging the quality of the provided partition. Early solutions generally turned to data approximation, e.g. Cannon [3] et al. and Kamel et al. [18] implemented FCM using only computations on integer values. Cheng et al. [6] deployed a random sampling of the input data, thus achieving a fast approximative FCM clustering. Later, data reduction schemes were introduced, aggregating similar input data before proceeding to clustering. Eschrich et al. [8] accelerated FCM this way by an order of magnitude. Data aggregation was also employed in image segmentation: clustering gray intensity levels instead of individual pixel intensities can speed up FCM by up to two orders of magnitude [29]. Szilágyi et al. [31] extended this pixel aggregation scheme to color images, thus achieving an efficient color reduction procedure. Alternately, Lázaro et

al. [21] proposed a parallel hardware implementation to FCM, and deployed it successfully in signal processing. Kolen and Hutcheson [20] proposed an FCM implementation that does not need to store the partition matrix, which is a relevant step toward clustering unloadable amounts of data. Further remarkable FCM solutions specialized for clustering huge data sets were introduced by Hathaway and Bezdek [12], and Havens et al. [13].

FCM has a main governing parameter called fuzzy exponent, usually denoted by m, and generally constrained by $m > 1$. The value of m has a strong impact on the fuzzyness of the created partition, and on the convergence speed as well. Large values of m reduce the ability of FCM to distinguish the input data: above a certain limit value all clusters merge together at the grand mean of the input data. However, this limit value is unknown, it strongly depends on the data. If m approaches its lower limit, the fuzzy partition tends toward the crisp one. In the limit case $m \to 1_+$, FCM becomes HCM.

Let us consider an FCM algorithm that uses fuzzy exponent m_0. If we do not like its convergence speed and the partition it makes, we may reduce the fuzzy exponent to m (where $1 < m < m_0$), which is a step towards the behavior of the HCM algorithm. The algorithm we obtain with this change is still FCM. The suppressed fuzzy c-means (s-FCM) algorithm, introduced by Fan et al. [9] in 2003, also makes a step towards HCM determined by the so-called suppression rate $\alpha \in [0, 1]$, but a different way, without staying in the bounds of the FCM algorithm. The s-FCM proved to converge in less iterations than FCM when used with the same fuzzy exponent m, and it provided fine partitions in all tested cases. However, the authors left several questions open, including (1) how to choose the value of the suppression rate α, or (2) is s-FCM an optimal algorithm? Since its introduction, the theory of the s-FCM algorithm evolved a lot, some of the open questions were answered and further open questions emerged. For example, Szilágyi et al. [28] explained the effect of the partition suppression in a comparative study with competitive learning [19], based on which later they introduced several generalized suppression rules for the fuzzy partition [30]. Several other works [15, 16, 17, 24, 27, 36] proposed minor modifications of the original s-FCM algorithm, providing parameter selection schemes for the suppression rate, and successfully applying s-FCM in various image processing tasks.

This paper proposes to provide an inventory of the theoretical advances regarding the s-FCM algorithm, and the successful applications that emerged since its introduction. The rest of this paper is structured as follows: Section 2 enumerates the foundations of the s-FCM algorithm, the c-means clustering models s-FCM relies on. Section 3 presents the details of the original s-FCM

algorithm, analyses its competitive behavior, and shows some of its generalization schemes. Section 4 explores the relation between suppressed c-means clustering models and the so-called FCM with generalized improved partition, giving some hints about the optimality of suppressed FCM clustering models. Section 5 relates on suppression parameter setting techniques found in the literature. Section 6 discusses the advantages and disadvantages of suppressed FCM clustering algorithms, while Section 7 concludes this study.

2 Background

2.1 The fuzzy and hard c-means algorithms

The conventional c-means clustering algorithms partition a set of object data into a predefined number of c clusters, through minimizing of a quadratic objective function. The objective function of FCM is:

$$J_{FCM} = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m \|\mathbf{x}_k - \mathbf{v}_i\|^2 = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m d_{ik}^2 \; , \tag{1}$$

where

- $\mathbf{x}_k$ stands for the input data $(k = 1, 2, \ldots, n)$,

- $\mathbf{v}_i$ represents the prototype (or centroid or representative element) of cluster i $(i = 1, 2, \ldots, c)$,

- $u_{ik} \in [0, 1]$ is the fuzzy membership function describing the degree to which input vector $\mathbf{x}_k$ belongs to cluster i,

- $m > 1$ is the fuzzy exponent ($m = 2$ in the version of Dunn [7]),

- and d_{ik} represents the distance between vector $\mathbf{x}_k$ and cluster prototype $\mathbf{v}_i$.

FCM uses a probabilistic partition, meaning that for any input vector $\mathbf{x}_k$ we have

$$\sum_{i=1}^{c} u_{ik} = 1 \; . \tag{2}$$

The objective function J_{FCM} is minimized by alternately applying the optimization of J_{FCM} over $\{u_{ik}\}$ with $\mathbf{v}_i$ fixed, $i = 1, 2, \ldots, c$, and the optimization
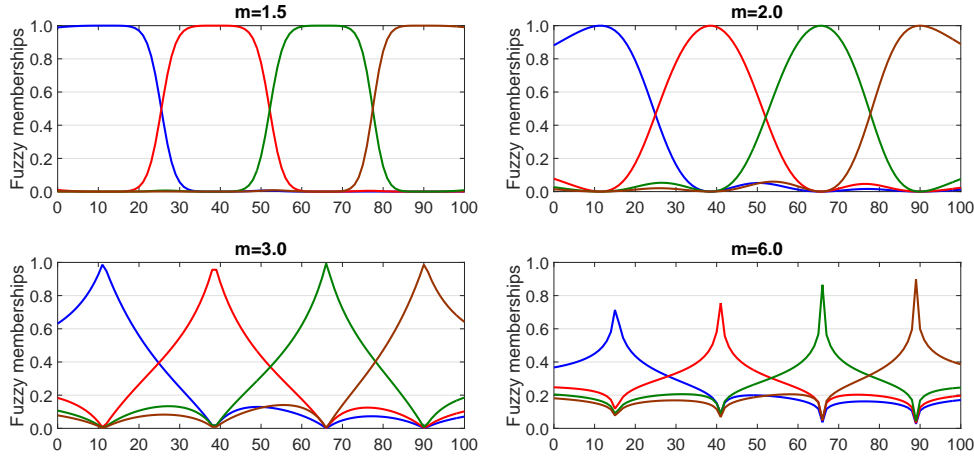
Figure 1: Fuzzy membership functions provided by FCM in a 1D problem, with random integer input values between 0 and 100. In case of $m = 6$, the peak of the membership functions also reach the maximum value 1, but not at integer valued inputs.

of J_{FCM} over $\{\mathbf{v}_i\}$ with u_{ik} fixed, $i = 1, 2, \ldots, c$; $k = 1, 2, \ldots, n$ [2]. The optimization formulas are deduced from the zero gradient conditions of J_{FCM} and Lagrange multipliers, and obtained as follows:

$$u_{ik}^{\star} = \frac{d_{ik}^{-2/(m-1)}}{\sum\limits_{j=1}^{c} d_{jk}^{-2/(m-1)}} \qquad \begin{array}{l} \forall\, i = 1, 2, \ldots, c \\ \forall\, k = 1, 2, \ldots, n \end{array}, \qquad (3)$$

$$\mathbf{v}_i^{\star} = \frac{\sum\limits_{k=1}^{n} u_{ik}^m \mathbf{x}_k}{\sum\limits_{k=1}^{n} u_{ik}^m} \qquad \forall\, i = 1, 2, \ldots, c \ . \qquad (4)$$

According to the optimization scheme of the FCM, Eqs. (3) and (4) are alternately applied, until cluster prototypes converge.

HCM is a limit case of FCM, which uses $m \to 1_+$, and thus the memberships are obtained by the winner-takes-all rule:

$$u_{ik}^{\star} = \begin{cases} 1 & \text{if } i = \arg\min\limits_{j}\{d_{jk}, j = 1, \ldots, c\} \\ 0 & \text{otherwise} \end{cases} \ . \qquad (5)$$

The propotype of each cluster in HCM is the average of the input vectors assigned to it, according to Eq. (4) using $m = 1$.

2.2 FCM versions with improved partition

An undesired property of the fuzzy memberships provided by FCM is their multimodality. Figure 1 presents some membership functions obtained by FCM in a single dimension problem with $c = 4$ clusters at various values of the fuzzy exponent. Each of the four fuzzy membership functions has a maximum at the cluster prototype, having the value of 1. Because of the probabilistic constraint, all other fuzzy membership functions are zero at these points. But, for example the fuzzy membership function represented in blue can have elevated values for input data situated very far from the cluster prototype. This is the multimodality, which gets stronger as the value of the fuzzy exponent m grows. So it is obvious that the multimodality gets suppressed if we reduce the value of m. Alternately, intending to suppress the multimodality, Höppner and Klawonn [14] proposed the so-called FCM with improved partition (IFP-FCM), which adds a rewarding term to the objective function of FCM:

$$J_{\text{IFP-FCM}} = \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik}^{m} d_{ik}^{2} - \sum_{k=1}^{n} a_{k} \sum_{i=1}^{c} (\mu_{ik} - 1/2)^{2} \ , \tag{6}$$

where parameters denoted by a_k are supposed to be positive numbers. The second term pushes the fuzzy membership values u_{ik}, $i = 1, 2, \ldots, c$; $k = 1, 2, \ldots, n$ towards the limits situated at 0 and 1, while preserving the probabilistic constraint. A generalized version of this algorithm (GIFP-FCM) was introduced by Zhu et al. [40], by replacing the rewarding term as follows:

$$J_{\text{GIFP-FCM}} = \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik}^{m} d_{ik}^{2} + \sum_{k=1}^{n} a_{k} \sum_{i=1}^{c} \mu_{ik} (1 - \mu_{ik}^{m-1}) \ . \tag{7}$$

The partition update formula derived by zero gradient conditions of the objective function is

$$\mu_{ik}^{\star} = \frac{(d_{ik}^{2} - a_{k})^{-1/(m-1)}}{\sum\limits_{j=1}^{c} (d_{jk}^{2} - a_{k})^{-1/(m-1)}} \qquad \begin{matrix} \forall\, i = 1, 2, \ldots, c \\ \forall\, k = 1, 2, \ldots, n \end{matrix} \ . \tag{8}$$

Fuzzy membership functions obtained with Eq. (8) can be interpreted as FCM partition memberships obtained by applying virtually reduced distances between cluster prototypes and input data. Each distance d_{ik} is replaced by

$\delta_{ik} = \sqrt{d_{ik}^2 - a_k}$. The authors also proposed a formula for a_k:

$$a_k = \omega \min_i \{d_{ik}^2, i = 1, 2, \ldots, c\} \ , \tag{9}$$

where $\omega \in [0.9, 0.99]$. Setting $\omega = 1$ would reduce GIFP-FCM to HCM, while $\omega = 0$ to FCM.

It is important to remark that both of the above improved clustering models kept FCM's prototype update formula given in Eq. (4), but applied to fuzzy membership functions μ_{ik} instead of u_{ik}, as described in Eq. (11).

Lately, these improved FCM clustering models were involved in several applications [4, 5].

3 The suppressed FCM algorithm

3.1 The original suppressed FCM

The suppressed fuzzy c-means algorithm was introduced by Fan et al. [9], declaring the goal to propose an algorithm with better convergence speed and reduced execution time than FCM, but providing the same partition quality. The s-FCM algorithm manipulates the optimization scheme of FCM, by inserting an extra step in each iteration, between partition updating via Eq. (3) and prototype updating via Eq. (4). This new step deforms the partition given by FCM according to the following rule:

$$\mu_{ik} = \begin{cases} 1 - \alpha + \alpha u_{ik} & \text{if } i = \arg \max_j \{u_{jk}\} \\ \alpha u_{ik} & \text{otherwise} \end{cases} \ , \tag{10}$$

where μ_{ik} $(i = 1, 2, \ldots, c; k = 1, 2, \ldots, n)$ represents the fuzzy memberships obtained after suppression. The cluster prototype update formula of s-FCM becomes:

$$\mathbf{v}_i^\star = \frac{\sum\limits_{k=1}^n \mu_{ik}^m \mathbf{x}_k}{\sum\limits_{k=1}^n \mu_{ik}^m} \qquad \forall i = 1, 2, \ldots, c \ . \tag{11}$$

Just like in case of competitive clustering [19], s-FCM sets up a competition for each input vector $\mathbf{x}_k$ in each iteration, which is won by the cluster whose prototype is situated at shortest distance from $\mathbf{x}_k$. Fuzzy memberships of $\mathbf{x}_k$ with respect to any non-winner cluster is proportionally suppressed ($\mu_{ik} = \alpha u_{ik}$), while all suppressed parts are given to the winner cluster to preserve
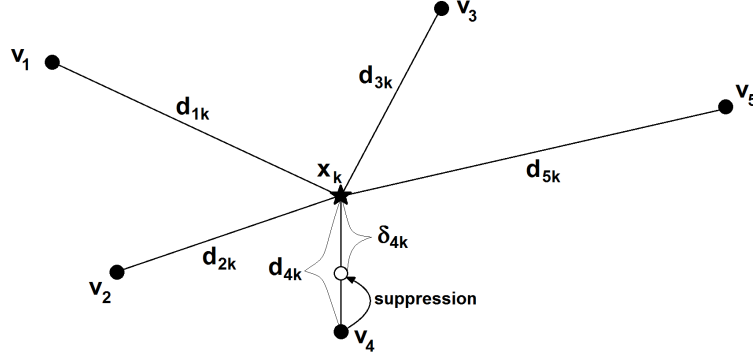
Figure 2: Shortened distance caused by suppression.

the probabilistic constraint: $\mu_{wk} = 1 - \alpha + \alpha u_{wk}$. Let w stand for the winner class index in the current competition for input vector $\mathbf{x}_k$. Actually it should be denoted by w_k, as it is specific to input vector $\mathbf{x}_k$, but for the sake of simplicity, the index of w is neglected in all formulas.

In the original article that introduced s-FCM, Fan et al. did not give any recipe or hint how to choose a suppression rate that would be optimal in any sense, or suitable for general or any specific purpose. They set the suppression rate to the middle of its range ($\alpha = 0.5$). Fan et al. validated s-FCM with some toy problems and found it insensitive to the fuzzy exponent m [9].

3.2 Suppression, regarded as a competition

Szilágyi et al. [28] showed that the proportional suppression of the FCM partition, the multiplication of all non-winner fuzzy memberships with a suppression rate $\alpha \in [0, 1]$, is mathematically equivalent with a reduction of the distance between the input vector and the closest cluster prototype. In any stage of the algorithm at partition updating, for any input vector $\mathbf{x}_k$ and its winner class with index w, there exists a virtually reduced distance $\delta_{wk} < d_{wk}$, which fulfils all partition update formulas of s-FCM, namely:

$$\mu_{wk} = \frac{\delta_{wk}^{\frac{-2}{m-1}}}{\delta_{wk}^{\frac{-2}{m-1}} + \sum_{j=1, j \neq w}^{c} d_{jk}^{\frac{-2}{m-1}}} = 1 - \alpha + \alpha \frac{d_{wk}^{\frac{-2}{m-1}}}{\sum_{j=1}^{c} d_{jk}^{\frac{-2}{m-1}}} \qquad (12)$$
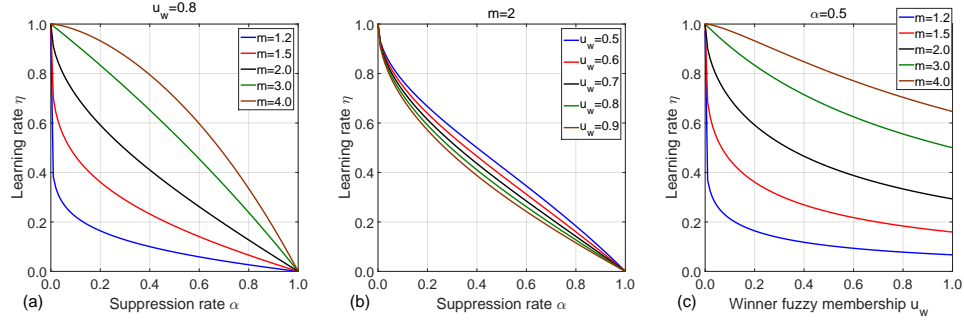
Figure 3: Some characteristics of the s-FCM algorithm: (a) learning rate η plotted against suppression rate α in case of $u_w = 0.8$, at various constant values of fuzzy exponent m; (b) learning rate η plotted against suppression rate α in case of $m = 2$, at various constant values of winner fuzzy membership u_w; (c) learning rate η plotted against winner fuzzy membership u_w in case of $\alpha = 0.5$, at various constant values of fuzzy exponent m.

and

$$\mu_{ik} = \frac{d_{ik}^{\frac{-2}{m-1}}}{\delta_{wk}^{\frac{-2}{m-1}} + \sum_{j=1, j \neq w}^{c} d_{jk}^{\frac{-2}{m-1}}} = \alpha \frac{d_{ik}^{\frac{-2}{m-1}}}{\sum_{j=1}^{c} d_{jk}^{\frac{-2}{m-1}}} \qquad \forall i \neq w \ . \qquad (13)$$

This virtual distance reduction is exhibited in Fig. 2. For the sake of clarity we need to remark that although the competition among clusters seems to be the same as in case of conventional competitive algorithms, these algorithms radically differ in the sense that conventional competitive algorithms do not work with quadratic objective functions [26].

Szilágyi et al. [28] also defined a quasi learning rate η of the s-FCM algorithm, in a similar way to the learning rate of competitive algorithms, and deduced its formula:

$$\eta(m, \alpha, u_{wk}) \equiv 1 - \frac{\delta_{wk}}{d_{wk}} = 1 - \left(1 + \frac{1-\alpha}{\alpha u_{wk}}\right)^{\frac{1-m}{2}} , \qquad (14)$$

where u_{wk} stands for the winner fuzzy membership value of vector $\mathbf{x}_k$, obtained without suppression. The learning rate η depends on both parameters of the s-FCM algorithm (fuzzy exponent m and suppression rate α), but also on the winner fuzzy membership function (u_{wk}) of the given input vector $\mathbf{x_k}$. This

means that η can hardly be a constant in any s-FCM scenario, unless defined as such.

3.3 Generalized suppressed FCM algorithm

The FCM algorithm with generalized suppression (gs-FCM) was introduced by Szilágyi et al. [30]. The s-FCM algorithm, as proposed by Fan et al. [9], uses a constant suppression rate α. There are three possible ways to give the suppression rate some variation:

1. *Time variant suppression* means to employ a suppression rate α_t that changes according to the iteration counter t. This sort of variation was used for example by Hung et al. [15].

2. *Context sensitive* or *data sensitive suppression* means to introduce time invariant rules of suppression, which provide dedicated suppression rate α_k for each input vector $\mathbf{x}_k$, depending on the current distances d_{ik}, $k = 1, 2, \ldots, n$.

3. *Time and context variant suppression* means to combine both previous variation versions in a unique suppression rule.

Szilágyi et al. [30] focused on context sensitive suppression rules only. Aiming at achieving quicker convergence they did not consider changing the suppression rule in every iteration. Their suppression rules were define according to two different schemes, presented in the following sections.

3.3.1 Learning rate defined as a function of the winner fuzzy membership

The first generalization scheme of the s-FCM algorithm uses a learning rate defined as a function of winner fuzzy membership u_w: $\eta = f(u_w)$, where $f : [0, 1] \to [0, 1]$ is a continuous function. Using Eq. (14), the context dependent suppression rate generally becomes

$$\alpha_k = \left[1 - u_w + u_w (1 - f(u_w))^{\frac{2}{1-m}} \right]^{-1} . \tag{15}$$

Some special cases using the above definition are:

- θ-type gs-FCM (gs$_\theta$-FCM) that uses constant learning rate $\eta = f(u_w) = \theta$ with parameter $\theta \in [0, 1]$, which leads to suppression rate

$$\alpha_k = \left[1 - u_w + u_w (1 - \theta)^{\frac{2}{1-m}} \right]^{-1} , \tag{16}$$

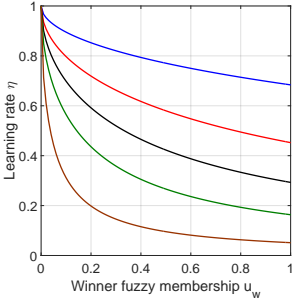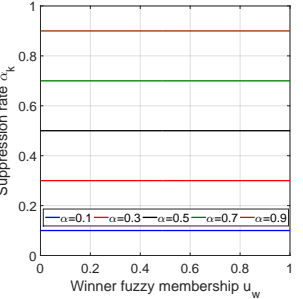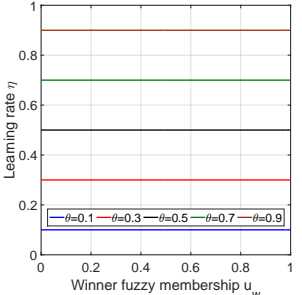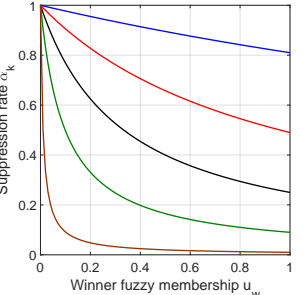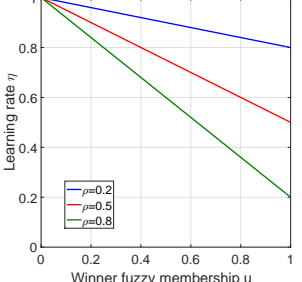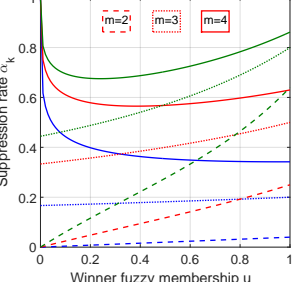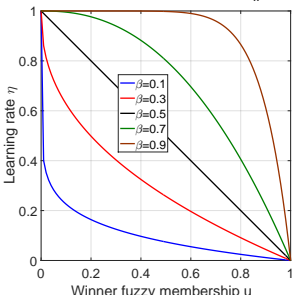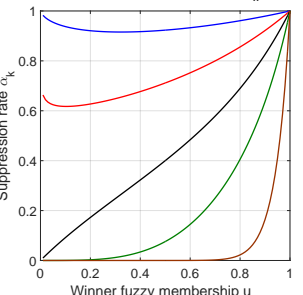| Definition	η vs. u_w	α_k vs. u_w
Original s $-$ FCM $\alpha_k = \alpha$ $\alpha \in [0,1]$		
$gs_\theta-$FCM $\eta = \theta$ $\theta \in [0,1]$		
$gs_\rho-$FCM $\eta = 1 - \rho u_w$ $\rho \in [0,1]$		
$gs_\beta-$FCM $\eta = 1 - u_w^{\beta/(1-\beta)}$ $\beta \in [0,1)$		

Table 1: Original s-FCM and generalized s-FCM algorithms defined with the first scheme ($\eta = f(u_w)$): definitions and context dependent suppression rates.

Definition	η vs. u_w	α_k vs. u_w
$\mathrm{gs}_\tau-\mathrm{FCM}$ $\mu_w = \frac{u_w+\tau}{1+u_w\tau}$ $\tau \in [0,1]$		
$\mathrm{gs}_\sigma-\mathrm{FCM}$ $\mu_w = u_w^\sigma$ $\sigma \in [0,1]$		
$\mathrm{gs}_\xi-\mathrm{FCM}$ $\mu_w = (\sin\frac{\pi u_w}{2})^\xi$ $\xi \in [0,1]$		

Table 2: Generalized s-FCM algorithms defined with the second scheme ($\mu_w = g(u_w)$): learning rates and context dependent suppression rates.

- ρ-type gs-FCM (gs$_\rho$-FCM) that uses learning rate linearly decreasing with the winner fuzzy membership $\eta = f(u_w) = 1 - \rho u_w$ with parameter $\rho \in [0,1]$, which leads to suppression rate

$$\alpha_k = \left[1 - u_w + \rho^{\frac{2}{1-m}} u_w^{\frac{3-m}{1-m}} \right]^{-1} , \qquad (17)$$

- β-type gs-FCM (gs$_\beta$-FCM) that uses learning rate decreasing with the winner fuzzy membership according to the exponential rule $\eta = f(u_w) = 1 - u_w^{\frac{\beta}{1-\beta}}$ with parameter $\beta \in [0, 1)$, which leads to suppression rate

$$\alpha_k = \left[1 + u_w \left(u_w^{\frac{2\beta}{(1-m)(1-\beta)}} - 1 \right) \right]^{-1} \ . \tag{18}$$

3.3.2 Direct formula between μ_w and u_w

The second generalization scheme is defined by a direct formula between the winner fuzzy membership values before and after suppression. In a general form, it is defined as $\mu_w = g(u_w)$ with $g : [0, 1] \rightarrow [0, 1]$ and $g(x) \geq x$ $\forall x \in [1/c, 1]$. Using Eq. (14), the context dependent suppression rate generally becomes

$$\alpha_k = \frac{1 - g(u_w)}{1 - u_w} \qquad \forall u_w \in [1/c, 1) \ . \tag{19}$$

For the special case when $u_w = 1$, the suppression rate is irrelevant, as nonwinner memberships are zero valued, so there is nothing to suppress. Some special cases using the above definition are:

- τ-type gs-FCM (gs$_\tau$-FCM) that is inspired by the relativistic speed addition formula $\mu_w = (u_w + \tau)/(1 + u_w \tau)$ with parameter $\tau \in [0, 1]$, which leads to suppression rate

$$\alpha_k = \frac{1 - \tau}{1 + u_w \tau} \qquad \forall u_w \in [1/c, 1] \ , \tag{20}$$

- σ-type gs-FCM (gs$_\sigma$-FCM) that uses the relation $\mu_w = u_w^\sigma$ with parameter $\sigma \in [0, 1]$, which leads to suppression rate

$$\alpha_k = \frac{1 - u_w^\sigma}{1 - u_w} \qquad \forall u_w \in [1/c, 1) \ , \tag{21}$$

- ξ-type gs-FCM (gs$_\xi$-FCM) that uses learning rate decreasing with the winner fuzzy membership according to the rule $\mu_w = (\sin \pi u_w/2)^\xi$ with parameter $\xi \in [0, 1]$, which leads to suppression rate

$$\alpha_k = \frac{1 - \left(\sin \frac{\pi u_w}{2} \right)^\xi}{1 - u_w} \qquad \forall u_w \in [1/c, 1) \ . \tag{22}$$

Algorithm 1: The gs-FCM algorithm

Data: Input data $\mathbf{x}_k$, $k = 1, 2, \ldots, n$

Data: Fuzzy exponent $m > 1$, suppression scheme and parameter, limit ε

Result: Cluster prototypes $\mathbf{v}_i$, $i = 1, 2, \ldots, c$

Initialize cluster prototypes $\mathbf{v}_i$, $i = 1, 2, \ldots, c$ with random input vectors, $\mathbf{v}_i \neq \mathbf{v}_j$ $\forall i \neq j$

repeat

> **for** $k = 1, 2, \ldots, n$ **do**
>
> > Compute fuzzy memberships u_{ik}, $i = 1, 2, \ldots, c$ with Eq. (3).
> >
> > Compute suppression rate α_k, according to the chosen suppression scheme and parameter value with one of the Eqs. (16)-(22).
> >
> > Compute suppressed fuzzy memberships μ_{ik}, $i = 1, 2, \ldots, c$ with Eq. (10), using suppression rate α_k.
>
> **end**
>
> **for** $i = 1, 2, \ldots, c$ **do**
>
> > $\mathbf{v}_i^{(\mathrm{old})} \leftarrow \mathbf{v}_i$
> >
> > Update cluster prototype $\mathbf{v}_i$ with Eq. (11).
>
> **end**

until $\sum\limits_{i=1}^{c} \|\mathbf{v}_i^{(\mathrm{old})} - \mathbf{v}_i\| < \varepsilon$;

3.3.3   The gs-FCM algorithm

The previous sections presented six generalized suppression rules, each regulated by a suppression parameter that can take an infinite number of different values. A limited number of these parameter values (up to two) reduce the generated algorithm to either FCM or HCM, while all other value define new clustering algorithms, different from HCM and FCM, or the original s-FCM. The suppression rules introduced above are summarized in Tables 1 and 2. Anyone can define further suppression rules by following the recipe given in Sections 3.3.1 and 3.3.2, by proposing a function $\eta = f(u_w)$ or $\mu_w = g(u_w)$, different from the ones exhibited in Tables 1 and 2. The gs-FCM algorithm is summarized in Algorithm 1.

4 The relation between s-FCM and GIFP-FCM clustering models

Now let us investigate the similarities and differences in the main loop of gs-FCM variants (let us call in this section the original s-FCM a variant of gs-FCM) and GIFP-FCM:

- All gs-FCM variants and GIFP-FCM use the same formula to update the cluster prototypes, given in Eq. (11).

- Compared to the original FCM, all gs-FCM variants and GIFP-FCM use modified d_{ik} distances. GIFP-FCM changes all distances by subtracting the same amount from the square of all d_{ik} values, $i = 1, 2, \ldots, c$, while gs-FCM variants reduce only the shortest distance d_{wk} to $\delta_{wk} = d_{wk}(1 - \eta_k)$, where η_k is the learning rate applied to vector x_k.

- GIFP-FCM makes two optimal steps in each loop, by executing its partition update and cluster prototype update formula. However, it changes the cost function in every loop by establishing the winner cluster for each vector x_k and adjusting the a_k values accordingly. Szilágyi [32] showed that gs-FCM variants can act the same way, the only difference is in the a_k terms, which are changed to s_{ik} and thus made dependent on cluster index i.

Szilágyi [32] introduced a unification theory for gs-FCM variants and GIFP-FCM. The new clustering model has the objective function very similar to the one of GIFP-FCM, but the so-called rewarding term, now denoted by s_{ik}, has a double indexing.

$$J_U = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m d_{ik}^2 + \sum_{k=1}^{n} s_{ik} \sum_{i=1}^{c} u_{ik}(1 - u_{ik}^{m-1}) \ . \tag{23}$$

Obviously, we can make this clustering model act like GIFP-FCM by setting $s_{ik} = a_k \ \forall i = 1, 2, \ldots, c$, where a_k is the rewarding term of GIFP-FCM defined in Eq. (9).

On the other hand, if we wish this new clustering model act like a certain gs-FCM variant, it is necessary to set:

$$s_{ik} = \begin{cases} d_{wk}^2(1 - \eta_k^2) & \text{if } i = w \equiv \arg\min_j\{d_{jk}, j = 1, \ldots, c\} \\ 0 & \text{otherwise} \end{cases} \ , \tag{24}$$

Algorithm	Parameter	Formula of s_{wk}
s-FCM	$\alpha \in [0,1]$	$d_{wk}^2 \left[1 - \left(\frac{\alpha u_w}{1-\alpha+\alpha u_w} \right)^{m-1} \right]$
gs_θ-FCM	$\theta \in [0,1]$	$d_{wk}^2 \theta(2-\theta)$
gs_ρ-FCM	$\rho \in [0,1]$	$d_{wk}^2 (1 - \rho^2 u_w^2)$
gs_β-FCM	$\beta \in [0,1)$	$d_{wk}^2 \left(1 - u_w^{2\beta/(1-\beta)} \right)$
gs_τ-FCM	$\tau \in [0,1]$	$d_{wk}^2 \left[1 - \left(\frac{u_w(1-\tau)}{u_w+\tau} \right)^{m-1} \right]$
gs_σ-FCM	$\sigma \in [0,1]$	$\begin{cases} d_{wk}^2 \left[1 - \left(\frac{u_w - u_w^{\sigma+1}}{u_w^\sigma - u_w^{\sigma+1}} \right)^{m-1} \right] & \text{if } u_w < 1 \\ 0 & \text{if } u_w = 1 \end{cases}$
gs_ξ-FCM	$\xi \in [0,1]$	$\begin{cases} d_{wk}^2 \left[1 - \left(\frac{u_w}{1-u_w}((\sin \frac{\pi u_w}{2})^{-\xi} - 1) \right)^{m-1} \right] & \text{if } u_w < 1 \\ 0 & \text{if } u_w = 1 \end{cases}$

Table 3: The definition of s_{wk} rewarding term for gs-FCM algorithm variants, u_w stands for the highest fuzzy membership provided by FCM for the input vector $\mathbf{x}_k$.

where η_k represents the learning rate applied to the current vector $\mathbf{x}_k$ according to the chosen suppression scheme, suppression parameter, and d_{ik} distances with $i = 1, 2, \ldots, c$. Table 3 exhibits s_{wk} rewarding terms for various gs-FCM algorithms.

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. The partition update formula of the unified clustering model is:

$$\mu_{ik}^\star = \frac{(d_{ik}^2 - s_{ik})^{-1/(m-1)}}{\sum\limits_{j=1}^{c} (d_{jk}^2 - s_{jk})^{-1/(m-1)}} \qquad \begin{array}{l} \forall i = 1, 2, \ldots, c \\ \forall k = 1, 2, \ldots, n \end{array}, \qquad (25)$$

while the cluster prototype update formula is the one given in Eq. (11). The unified clustering algorithm that integrates all suppressed clustering models and the GIFP-FCM algorithm is exhibited in Algorithm 2. This version of the algorithm is not the recommended one to implement GIFP-FCM or gs-FCM model. It is only the proof of their similar structure. GIFP-FCM runs optimally as described in [40], while gs-FCM variants as described in Algorithm 1.

Algorithm 2: The unified algorithm

Data: Input data $\mathbf{x}_k$, $k = 1, 2, \ldots, n$

Data: Fuzzy exponent $m > 1$, limit ε, the chosen algorithm and its
        parameter (one of $\{\alpha, \theta, \rho, \beta, \tau, \sigma, \xi, \omega\}$)

Result: Cluster prototypes $\mathbf{v}_i$, $i = 1, 2, \ldots, c$

Initialize cluster prototypes $\mathbf{v}_i$, $i = 1, 2, \ldots, c$ with random
 input vectors, $\mathbf{v}_i \neq \mathbf{v}_j$ $\forall i \neq j$

repeat

   for $k = 1, 2, \ldots, n$ **do**

      Find the index of the winner cluster
      $w = \arg \min_i \{d_{ik}, i = 1, 2, \ldots, c\}$.

      Compute fuzzy memberships u_{ik}, $i = 1, 2, \ldots, c$ with Eq.
       (3).

      if *algorithm is GIFP-FCM* **then**
      $\quad | \quad s_{ik} \leftarrow \omega d_{wk}^2$, $i = 1, 2, \ldots, c$
      end

      else
      $\quad |$ Set s_{ik} values according to Eq. (24) and Table 3.
      end

      Compute suppressed fuzzy memberships μ_{ik}, $i = 1, 2, \ldots, c$
       with Eq. (25).

   end

   for $i = 1, 2, \ldots, c$ **do**
   $\quad \mathbf{v}_i^{(old)} \leftarrow \mathbf{v}_i$
      Update cluster prototype $\mathbf{v}_i$ with Eq. (11).
   end

until $\sum_{i=1}^{c} \|\mathbf{v}_i^{(old)} - \mathbf{v}_i\| < \varepsilon$;

---

5 Parameter selection

Fan et al. [9] did not give any recipe how to choose the value of the suppression
rate α. They set the suppression rate to the middle of its definition interval,
to be well in between HCM and FCM. Several further works, including the
gs-FCM clustering models of Szilágyi et al. [28, 30], set the parameter values
experimentally and showed that there are various settings that make s-FCM
and gs-FCM models work better than FCM or HCM in various applications.

However, there are some application papers [15, 16, 17, 24, 27, 36] in the literature that give recipes for the choice of the suppression parameter.

5.1 Constant suppression rate based on input data

Fan et al. [10] proposed a constant suppression rate based on the distribution of the input data, defined as

$$\alpha = \frac{\sum\limits_{j=1}^{n} \sum\limits_{r=1}^{n} \|\mathbf{x}_j - \mathbf{x}_r\|}{n \sum\limits_{j=1}^{n} \|\mathbf{x}_j - \overline{\mathbf{x}}\|} - 1 \ , \tag{26}$$

where $\overline{\mathbf{x}}$ stands for the grand mean of the input vectors: $\overline{\mathbf{x}} = n^{-1} \sum_{i=1}^{n} \mathbf{x}_i$. The authors proved that the value of α defined by Eq. (26) is always in the interval $[0, 1]$. The value of α should be evaluated once as an initialization step of the algorithm, and applied as constant compression rate through all optimization loops. Obviously, this only works with the original s-FCM algorithm, not with gs-FCM models.

5.2 Time variant suppression rate based on partition entropy

Li et al. [22] proposed a time variant suppression rate based on the entropy of the partition provided by the FCM algorithm, defined with the formula

$$\alpha_{Li} = \frac{1}{\log c} \left(-\frac{1}{n} \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} \log(u_{ik}) \right) \ . \tag{27}$$

This formula is evaluated in every optimization loop, after having applied the partition update formula given in Eq. (3) and before starting the partition suppression using Eq. (10). The authors found their method successful in image segmentation problems, despite this entropy based suppression rate fully suppresses the crisp partition and applies no change in the completely ambiguous situation described by $u_{ik} = 1/c \ \forall i = 1, 2, \ldots, c$ and $\forall k = 1, 2, \ldots, n$. In our opinion, it would be more useful setting the suppression rate to $1 - \alpha_{Li}$.

5.3 Time variant suppression rate based on current cluster prototypes

There is a set of works that apply time variant suppression rate, which changes from iteration to iteration according to the current cluster prototypes. As

prototypes converge, the suppression rate also stabilizes. The foundation of all these recipes is the formula of the Xie-Beni cluster validity index [38]:

$$\mathsf{XB} = \frac{\sum\limits_{i=1}^{c} \sum\limits_{k=1}^{n} u_{ik}^2 \|\mathbf{x}_k - \mathbf{v}_i\|^2}{n \left(\min_{1 \le i \ne j \le c} \|\mathbf{v}_i - \mathbf{v}_j\|^2\right)} \ , \tag{28}$$

which indicates fine cluster quality at low values of XB. Knowing that well separable clusters are best partitioned by HCM, while FCM handles overlapping clusters better, it seems a good idea to apply stronger suppression when the minimum distance between cluster prototypes is higher. In this order, Hung et al. [15] proposed a time variant suppression rate defined as

$$\alpha = \exp\left(-\frac{1}{\beta} \min_{1 \le i \ne j \le c} \|\mathbf{v}_i - \mathbf{v}_j\|^2\right) \ , \tag{29}$$

where

$$\beta = \frac{1}{n} \sum_{j=1}^{n} \|\mathbf{x}_j - \overline{\mathbf{x}}\|^2 \ , \tag{30}$$

and $\overline{\mathbf{x}} = n^{-1} \sum_{i=1}^{n} \mathbf{x}_i$ is the grand mean of input vectors. In the application of Hung et al. [15], α is evaluated at the beginning of each optimization loop, and applied to suppress the fuzzy memberships provided by the FCM partition update formula. This suppression was found successful in an ophthalmology image segmentation problem, similarly to the alternative one introduced by the same authors in [16]:

$$\alpha = \left(1 + \min_{1 \le i \ne j \le c} \frac{\|\mathbf{v}_i - \mathbf{v}_j\|^2}{\beta}\right)^{-1} \ , \tag{31}$$

that also uses the formula of β given in Eq. (30). Tsai et al. [36] also introduced a kernel-based suppressed FCM version, using the suppression formula derived from Eq. (29).

A very similar formulation of the suppression rate formula was given by Nyma et al. [24]:

$$\alpha = \exp\left(- \min_{1 \le i \ne j \le c} \frac{\|\mathbf{v}_i - \mathbf{v}_j\|^2}{m}\right) \ , \tag{32}$$

where m is the fuzzy exponent. In spite of being applied successfully in medical image segmentation by the authors, we find this an ill-posed formula, as it advises a different suppression rate if we replace all input vectors by a constant $\kappa \ne 1$.

6 Discussion

All fuzzy c-means algorithms with suppressed or improved partitions use an extra parameter compared to FCM, which regulates the alteration of the FCM partition. In case of the GIFP-FCM algorithm we cannot talk about suppression, but the effect is similar: although the parameter ω is recommended to be chosen from the interval $[0.9, 0.99]$, we need to remark that $\omega = 1$ reduces the GIFP-FCM to HCM, while $\omega = 0$ means fully FCM behavior for GIFP-FCM. Some of the suppression schemes, namely the original s-FCM and the gs-FCM algorithm of type σ act like HCM when the value of the parameter (α, σ) is 0, and as FCM at the other extremum 1. On the other hand, the gs-FCM algorithms of type θ, β, τ act like FCM when the value of the parameter (θ, β, τ) is 0, and as HCM at the other extremum 1. Any other value of the suppression parameters defines a clustering model that is different from HCM or FCM of any fuzzy exponent $m > 1$.

Suppressed FCM clustering algorithms have a moderate popularity, because they are not optimal, as they do not minimize the objective function of FCM or any other known objective function. The unification theory with the GIFP-FCM algorithm revealed that all suppressed clustering models can be considered optimal to the same extent as GIFP-FCM. Despite this disadvantage, several applications showed that s-FCM and gs-FCM clustering models can perform better than HCM or FCM, and run in less time than FCM. Although the quality of the clustering outcome should be characterized by cluster validity indexes (CVI), several image processing applications showed that suppressed FCM clustering models can capture better the underlying structure of the input data than FCM or HCM, leading to better segmentation quality. Szilágyi et al. [30] employed the Xie-Beni [38], the extended Xie-Beni [25], and the Fukuyama-Sugeno [11] CVIs to prove the ability of gs-FCM models to produce valid partitions. The authors also showed that suppressed FCM clustering is substantially less sensitive to imbalanced cluster sizes. Hung et al. [15, 16] deployed suppressed FCM in an ophthalmologic MRI image segmentation problem and found it more effective than the classical FCM. Zhao et al. [39] reported improved image processing accuracy achieved via suppressing the FCM partition in a general purpose image processing environment. Improvement in performance against the clock achieved via suppressing the FCM partition were reported by Szilágyi et al. [31, 33] in a color reduction application. In a very recent paper, Wu et al. [37] combined the suppressed FCM with the so-called picture fuzzy clustering method [35] that has type

II fuzzy background. The resulting clustering model performed better than previous methods in terms of both accuracy and efficiency.

7   Conclusion

This paper presents the short history of the suppressed fuzzy c-means algorithm, focusing on the most important theoretical advances and providing a short summary of practical achievements. Applying suppressed partitions in clustering models derived from fuzzy c-means currently have a moderate popularity, which may rise in the future due to the recent successful extensions and applications.

Acknowledgements

References

[1] D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, *Proc. 18th Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, LA USA*, 2007, pp. 1027–1035. ⇒303

[2] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*, Plenum, New York (1981) ⇒303, 306

[3] 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. ⇒303

[4] A. Celikyilmaz, I. B. Türkşen, Enhanced fuzzy system models with improved fuzzy clustering algorithm, *IEEE Trans. Fuzzy Syst.* **16** (2008) 779–794. ⇒308

[5] 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. ⇒308

[6] T. W. Cheng, D. B. Goldgof, L. O. Hall, Fast fuzzy clustering, *Fuzzy Sets Syst.* **93** (1998) 49–56. ⇒303

[7] 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. ⇒303, 305

[8] 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 ⇒303

[9] J. L. Fan, W. Z. Zhen, W. X. Xie, Suppressed fuzzy c-means clustering algorithm. *Patt. Recogn. Lett.* **24** (2003) 1607–1612. ⇒304, 308, 309, 311, 318

[10] J. L. Fan, J. Li, A fixed suppressed rate selection method for suppressed fuzzy c-means clustering algorithm, *Appl. Math.* **5** (2014) 1275–1283. ⇒319

[11] Y. Fukuyama, M. Sugeno, A new method of choosing the number of clusters for the fuzzy c-means method (in Japanese), *Proc. 5th Fuzzy Systems Symposium*, Japan, 1989, pp. 247–250. ⇒321

[12] R. J. Hathaway, J. C. Bezdek, Extending fuzzy and probabilistic clustering to very large data sets. *Comput. Stat. Data Anal.* **51** (2006) 215–234. ⇒304

[13] 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. ⇒304

[14] F. Höppner, F. Klawonn, Improved fuzzy partition for fuzzy regression models, *Int. J. Approx. Reason.* **5** (2003) 599–613. ⇒307

[15] W. L. Hung, M. S. Yang, D. H. Chen, Parameter selection for suppressed fuzzy c-means with an application to MRI segmentation, *Patt. Recogn. Lett.* **27** (2006) 424–438. ⇒304, 311, 319, 320, 321

[16] W. L. Hung, Y. C. Chang, A modified fuzzy c-means algorithm for differentiation in MRI of ophtalmology, *Int'l Conference on Modeling Decisions for Artificial Intelligence*, Tarragona, Spain, *LNCS* **3885** (2006) 340–350. ⇒304, 319, 320, 321

[17] 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. ⇒304, 319

[18] M. S. Kamel, S. Z. Selim, New algorithms for solving the fuzzy clustering problem, *Patt. Recogn.* **27** (1994) 421–428. ⇒303

[19] T. Kohonen, The self-organizing map, *Proc. IEEE* **78** (1990) 1474–1480. ⇒304, 308

[20] J. F. Kolen, T. Hutcheson, Reducing the time complexity of the fuzzy c-means algorithm. *IEEE Trans. Fuzzy Syst.* **10** (2002) 263–267. ⇒304

[21] J. Lázaro, J. Arias, J. L. Martín, C. Cuadrado, A. Astarloa, Implementation of a modified fuzzy c-means clustering algorithm for real-time applications, *Microproc. & Microsyst.* **29** (2005) 375–380. ⇒304

[22] J. Li, J. L Fan, Parameter selection for suppressed fuzzy c-means clustering algorithm based on fuzzy partition entropy, *11th Int. Conf. on Fuzzy Systems and Knowledge Discovery*, Xiamen, China, 2014, pp. 82–87. ⇒319

[23] S. McQueen, Some methods for classification and analysis of multivariate observations, *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, 1967, pp. 281–297. ⇒303

[24] A. Nyma, M. Kang, Y. K. Kwon, C. H. Kim, J. M. Kim, A hybrid technique for medical image segmentation, *J. Biomed. Biotechnol.* **2012** (2012) 830252. ⇒304, 319, 320

[25] N. R. Pal, J. C. Bezdek, On cluster validity for the fuzzy c-means model, *IEEE Trans. Fuzzy Syst.* **3** (1995) 370–379. ⇒321

[26] N. R. Pal, J. C. Bezdek, R. J. Hathaway, Sequential competitive learning and the fuzzy c-means clustering algorithms, *Neural Networks* **9** (1996) 787–796. ⇒ 310

[27] M. F. Saad, A. M. Alimi, Improved modified suppressed fuzzy c-meanss, *Proc. 2nd Int'l Conference on Image Processing Theory, Tools and Applications*, Paris, 2010, pp. 313–318. ⇒304, 319

[28] 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, *Soft. Comput.* **14** (2010) 495–505. ⇒304, 309, 310, 318

[29] L. Szilágyi, S. M. Szilágyi, B. Benyó, Efficient inhomogeneity compensation using fuzzy c-means clustering models, *Comput. Meth. Prog. Biol.* **108** (2012) 80–89. ⇒303

[30] L. Szilágyi, S. M. Szilágyi, Generalization rules for the suppressed fuzzy c-means clustering algorithm, *Neurocomput.* **139** (2014) 298–309. ⇒304, 311, 318, 321

[31] L. Szilágyi, G. Dénesi, S. M. Szilágyi, Fast color reduction using approximative c-means clustering models, *Proc. IEEE Int. Conference on Fuzzy Systems*, Beijing, 2014, pp. 194–201. ⇒303, 321

[32] L. Szilágyi, A unified theory of fuzzy c-means clustering models with improved partition, *Proc. Int'l Conference on Modeling Decisions for Artificial Intelligence*, Skövde, Sweden, *LNCS* **9321** (2015) 129–140. ⇒316

[33] L. Szilágyi, G. Dénesi, C. Enăchescu, Fast color quantization via fuzzy clustering, *Proc. 23rd International Conference on Neural Information Processing*, Kyoto, 2016, *LNCS* 9950 (2016) 95–103. ⇒321

[34] H. Steinhaus, Sur la division des corps matériels en parties, *Bulletin de l'Académie Polonaise des Sciences* **C1** III.(IV) (1956) 801–804. ⇒303

[35] P. H. Thong, L. H. Son, Picture fuzzy clustering: a new computational intelligence method, *Soft. Comput.* **20** (2016) 3549–3562. ⇒321

[36] 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, *Proc. Int'l Conference on Artificial Intelligence and Computational Intelligence*, Chengdu, China, *LNCS* 7530 (2012) 744–754. ⇒304, 319, 320

[37] C. M. Wu, N. Liu, Suppressed robust picture fuzzy clustering for image segmentation, *Soft. Comput.* doi: 10.1007/s00500-020-05403-8. ⇒321

[38] X. L. Xie, G. A. Beni, Validity measure for fuzzy clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* **3** (1991) 841–846. ⇒320, 321

[39] F. Zhao, J. L. Fan, H. Q. Liu, Optimal-selection-based suppressed fuzzy c-means clustering algorithm with self-tuning non local spatial information for image segmentation, *Expert Syst. Appl.* **41** (2014) 4083–4093. ⇒321

[40] L. Zhu, F. L. Chung, S. Wang, Generalized fuzzy c-means clustering algorithm with improved fuzzy partitions, *IEEE Trans. Syst. Man Cybern. B.* **39** (2009) 578–591. ⇒307, 317