

A Thorough Analysis of the Suppressed Fuzzy C-Means Algorithm

László Szilágyi^{1,2}, Sándor M. Szilágyi¹, and Zoltán Benyó¹

¹ Sapientia - Hungarian Science University of Transylvania,
Faculty of Technical and Human Science, Târgu-Mureş, Romania
lalo@ms.sapientia.ro

² Budapest University of Technology and Economics, Department of Control
Engineering and Information Technology, Budapest, Hungary

Abstract. Suppressed fuzzy c-means (s-FCM) clustering was introduced in [Fan, J. L., Zhen, W. Z., Xie, W. X.: Suppressed fuzzy c-means clustering algorithm. *Patt. Recogn. Lett.* 24, 1607–1612 (2003)] with the intention of combining the higher speed of hard c-means (HCM) clustering with the better classification properties of fuzzy c-means (FCM) algorithm. They added an extra computation step into the FCM iteration, which created a competition among clusters: lower degrees of memberships were diminished according to a previously set suppression rate, while the largest fuzzy membership grew by swallowing all the suppressed parts of the small ones. Suppressing the FCM algorithm was found successful in the terms of accuracy and working time, but the authors failed to answer a series of important questions. In this paper we attempt to clarify the view upon the optimality and the competitive behavior of s-FCM via analytical computations and numerical analysis.

Keywords: fuzzy c-means algorithm, suppressed fuzzy c-means algorithm, competitive clustering, alternating optimization.

1 Introduction

The fuzzy c-means (FCM) and hard c-means (HCM) algorithms are probably the most popular clustering approaches, due to their fine separation abilities and easy implementation. It is a well-known diagnosis, that FCM outperforms HCM in the terms of partition quality, at the cost of a slower convergence.

Several researchers have studied the convergence speed of FCM and tried to introduce modified algorithms with improved characteristics [3,6]. Another trend was to combine competitive and FCM clustering techniques. In this order, several new algorithms appeared which tried to fuzzify Kohonen's learning vector quantization (LVQ) algorithm [9,10]. Karayiannis and Bezdek [7] introduced an integrated approach to fuzzy LVQ and FCM, based on the notion of generalized mean of real numbers. These algorithms, even if their main goal wasn't to improve the convergence of accurate clustering techniques, they had a considerable contribution.

Recently, Wei and Xie [8] proposed a technique called rival checked FCM to modify the fuzzy membership values given by FCM, in order to improve the performance against the clock. Their solution, which rose the highest degree of membership at the detriment of the second largest, led to unacceptable results in some cases. In order to confront this problem, Fan et al. [4] introduced the suppressed fuzzy c -means algorithm (s-FCM). The authors stated that, by prizing the highest membership and suppressing all others, the modification does not disturb the original order among clusters. They also remarked, that setting the suppression rate $\alpha = 0$ makes s-FCM and HCM identical, while $\alpha = 1$ reduces the algorithm to the conventional FCM. The s-FCM algorithm was found successful based on some numerical analysis, but unfortunately, the authors left several issues wide open:

1. The extra step of s-FCM was inspired by the basis of competitive learning [4], but the authors failed to give any evidence of its competitive behavior.
2. The question of optimality did not even emerge.
3. The authors failed to provide any strategy to choose the suppression rate α . This was already pointed out by Hung et al. [5], who formulated a criterion for α based on considerations regarding cluster validity.

These are the main topics to be investigated in the present paper, which is structured as follows. Section 2 presents the background works that have impact on our investigations. Section 3 contains the analytical computations performed in order to reveal some properties of s-FCM. Section 4 gives a numerical analysis of the properties of s-FCM. Conclusions are given in the last section.

2 Background

The traditional FCM algorithm partitions a set of object data into a number of c clusters based on the minimization of a quadratic objective function: $J_{\text{FCM}} = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|\mathbf{x}_k - \mathbf{v}_i\|^2 = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d_{ik}^2$, where \mathbf{x}_k represents the input data ($k = 1 \dots n$), \mathbf{v}_i represents the prototype or centroid or representative element of cluster i ($i = 1 \dots c$), $u_{ik} \in [0, 1]$ is the fuzzy membership function showing the degree to which datum k belongs to cluster i , $m > 1$ is the fuzzyfication parameter, and d_{ik} is the distance between datum \mathbf{x}_k and prototype \mathbf{v}_i . According to the definition of fuzzy sets, for any datum \mathbf{x}_k , we have $\sum_{i=1}^c u_{ik} = 1$. The minimization of the objective function is reached by alternately applying the optimization of J_{FCM} over $\{u_{ik}\}$ with \mathbf{v}_i fixed, $i = 1 \dots c$, and the optimization of J_{FCM} over $\{\mathbf{v}_i\}$ with u_{ik} fixed, $i = 1 \dots c$, $k = 1 \dots n$ [2]. During each cycle, the optimal values are computed from the zero gradient conditions, and obtained as follows:

$$u_{ik}^* = \frac{d_{ik}^{-2/(m-1)}}{\sum_{j=1}^c d_{jk}^{-2/(m-1)}} \quad \text{and} \quad \mathbf{v}_i^* = \frac{\sum_{k=1}^n u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^n u_{ik}^m}, \quad (1)$$

According to the alternating optimization (AO) scheme, the formulae shown in (1) are alternately applied, until cluster prototypes stabilize.

The main difference between HCM and FCM clustering is the membership logic. While in case of fuzzy clustering, the degree of membership of a datum k to cluster i , denoted by u_{ik} , can take any value between 0 and 1, hard clustering uses as degrees of membership only two values: $h_{ik} \in \{0, 1\}$. Although the objective function of HCM is the same as the one of FCM, using $m = 1$, the optimization formula differs from Eq. (1): for any $k = 1 \dots n$, h_{ik} is set to 1 whenever \mathbf{v}_i is the closest cluster prototype viewed from \mathbf{x}_k , and 0 in any different case. Ties are resolved arbitrarily. The computation of cluster prototypes is performed according to the weighted averaging formula given in Eq. (1).

The suppressed fuzzy c-means algorithm was introduced by Fan et al. [4], having the declared goal of improving the convergence speed of FCM, while keeping its good classification accuracy. They inserted an extra computational step between the formulae given in Eq. (1): for each datum \mathbf{x}_k , after having obtained its new optimal fuzzy membership values u_{ik} , they search for the highest one $u_{w_k k}$, and declare cluster $w_k \in \{1, 2, \dots, c\}$ the winner. The fuzzy memberships are then modified such a way, that all non-winner values are decreased via multiplying by a so-called suppression rate α , ($0 \leq \alpha \leq 1$): $\mu_{ik} = \alpha u_{ik}$ if $i \neq w_k$, and the winner membership is increased such a way, that the probabilistic constraint remains fulfilled by the modified memberships: $\mu_{w_k k} = 1 - \alpha + \alpha u_{w_k k}$. Here μ_{ik} , $i = 1 \dots c$, $k = 1 \dots n$, represent the fuzzy memberships obtained with the modification introduced by s-FCM. Cluster prototypes are computed similarly to FCM, but using the modified degrees of membership.

3 What Is the Suppressed Fuzzy C-Means Algorithm?

3.1 Competition by Suppression

We will start the investigation from FCM’s fuzzy membership update formula shown in (1). When the new degrees of membership of datum \mathbf{x}_k are computed, everything depends on the distances d_{ik} , $i = 1 \dots c$. If we change the scale of the metric such a way, that distances are lengthened or shortened proportionally, the obtained memberships remain the same. In other words, the ratio of two membership values, say u_{ik}/u_{jk} does not depend on the above mentioned scale.

Conversely, when memberships to non-winner clusters are proportionally suppressed via multiplying them by α , it can be interpreted as their distances from vector \mathbf{x}_k remain constant, but as the winner cluster receives a higher degree of membership, prototype \mathbf{v}_{w_k} is *counted as it were closer* to \mathbf{x}_k than it really is. Again, in other words, when new cluster prototypes \mathbf{v}_i are computed using the weighted averaging formula in Eq. (1), the vectors \mathbf{x}_k whose competition the currently computed prototype has won, are taken into consideration as they were *at a reduced distance* $d'_{w_k k} < d_{w_k k}$, giving those winner vectors a higher impact than in FCM. This reduced distance can be characterized with a *learning rate* defined as:

$$\eta_{s-FCM} = 1 - \frac{d'_{w_k k}}{d_{w_k k}} \quad , \quad (2)$$

which will be computed in the followings. We will use the notations: $\delta_{ik} = d_{ik}^{-2/(m-1)} \forall i = 1 \dots c, \forall k = 1 \dots n$, and $\delta'_{w_k k} = \gamma \delta_{w_k k}$, where we expect $\gamma \geq 1$. Using these new notations, we can rewrite the left part of (1) for both winner and non-winner clusters. For the winner cluster we have:

$$\mu_{w_k k} = \frac{\gamma \delta_{w_k k}}{\gamma \delta_{w_k k} + \sum_{j=1, j \neq w_k}^c \delta_{jk}} = \frac{\gamma \delta_{w_k k}}{(\gamma - 1) \delta_{w_k k} + \sum_{j=1}^c \delta_{jk}} \quad (3)$$

while the non-winner clusters receive the memberships:

$$\mu_{ik} = \frac{\delta_{ik}}{\gamma \delta_{w_k k} + \sum_{j=1, j \neq w_k}^c \delta_{jk}} = \frac{\delta_{ik}}{(\gamma - 1) \delta_{w_k k} + \sum_{j=1}^c \delta_{jk}} \quad (4)$$

Now we can compare the suppressed memberships obtained in (3) and (4) with their definition given by Fan *et al.* From the latter one we get:

$$\frac{\delta_{ik}}{(\gamma - 1) \delta_{w_k k} + \sum_{j=1}^c \delta_{jk}} = \alpha \frac{\delta_{ik}}{\sum_{j=1}^c \delta_{jk}} \quad (5)$$

As δ_{ik} is never zero, this equation can be restructured as follows:

$$\sum_{j=1}^c \delta_{jk} = \alpha (\gamma - 1) \delta_{w_k k} + \alpha \sum_{j=1}^c \delta_{jk} \quad (6)$$

and so

$$\gamma = 1 + \frac{1 - \alpha}{\alpha u_{w_k k}} = \frac{\mu_{w_k k}}{\alpha u_{w_k k}} = \frac{\mu_{w_k k}}{\mu_{w_k k} - (1 - \alpha)} \quad (7)$$

If we started from (3), we would obtain the same formula for γ . Although this is not yet the learning rate, we should discuss about the possible singularities:

- The degree of membership assigned by FCM to the winner class, $u_{w_k k}$, cannot be zero, because then all other u_{ik} values would be zero, and that contradicts the probability constraint of FCM.
- The suppression rate α can be zero, but that would reduce s-FCM to HCM, which is a trivial case with strict winner-takes-all competition.
- As the suppression rate α and the winner cluster’s fuzzy membership are both in the interval $[0, 1]$, we indeed have $\gamma \geq 1$. Equality holds when $\alpha = 1$, that is, there is no suppression.

We can conclude, that Eq. (7) is valid if $0 < \alpha \leq 1$. Under these circumstances, the learning rate of the s-FCM is:

$$\eta_{s\text{-FCM}} = 1 - \frac{d'_{w_k k}}{d_{w_k k}} = 1 - \gamma^{(1-m)/2} = 1 - \left(1 + \frac{1 - \alpha}{\alpha u_{w_k k}} \right)^{(1-m)/2} \quad (8)$$

It was expected, that the fuzzyfication parameter m and the suppression rate α influence the learning rate. In addition, another factor is present, namely the fuzzy membership value of the winner cluster, $u_{w_k k}$. Some graphical representations of the learning rate vs. suppression rate are shown in Fig. 1 (a) and (b). So far we can conclude, that s-FCM has a quasi-competitive behavior with a variable learning rate given in Eq. (8).

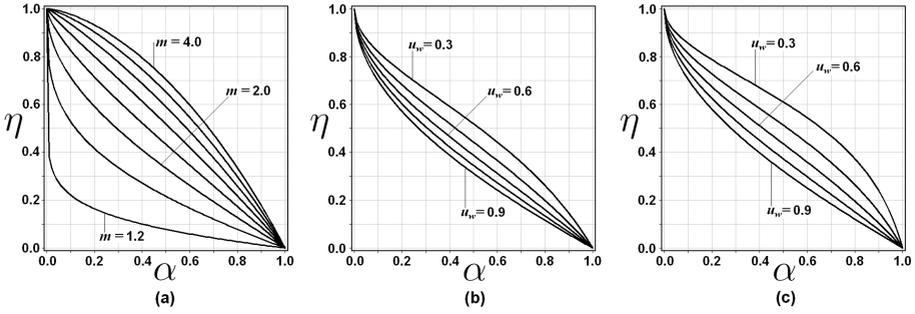


Fig. 1. Effect of suppression on the learning rate of s-FCM: (a) at $u_w = 0.8$ and different values of m , (b) at $m = 2$ and different values of winner membership u_w ; (c) learning rate of Os-FCM at $m = 2$ and different values of u_w

3.2 Is s-FCM Optimal?

Let us consider from the beginning, that $0 < \alpha < 1$, as the two extreme cases are trivial anyway and need no investigation. We can call s-FCM optimal if we find an objective function, whose AO minimization gives the optimization formulae:

$$\mu_{ik} = \alpha \cdot \frac{d_{ik}^{-2/(m-1)}}{\sum_{j=1}^c d_{jk}^{-2/(m-1)}} \quad \forall k = 1 \dots n \quad \forall i \neq w_k, \quad (9)$$

$$\mu_{w_k k} = 1 - \alpha + \alpha \cdot \frac{d_{w_k k}^{-2/(m-1)}}{\sum_{j=1}^c d_{jk}^{-2/(m-1)}} \quad \forall k = 1 \dots n, w_k = \arg \min_i \{d_{ik}\}, \quad (10)$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^n \mu_{ik}^m \mathbf{x}_k}{\sum_{k=1}^n \mu_{ik}^m} \quad \forall i = 1 \dots c. \quad (11)$$

Unfortunately, this kind of analytical function is not likely to exist. There exists at least one function, namely

$$J_{\text{not s-FCM}} = \sum_{k=1}^n \sum_{i=1}^c [\mu_{ik} - (1 - \alpha)h_{ik}]^m d_{ik}^2, \quad (12)$$

which satisfies the first two conditions for any $\alpha \in (0, 1)$, but the generated cluster prototypes coincide with the ones of FCM. Let us now propose a novel approach for an optimal suppression and name it optimally suppressed fuzzy c -means algorithm (Os-FCM):

$$J_{\text{Os-FCM}} = \sum_{k=1}^n \sum_{i=1}^c [\alpha u_{ik}^m + (1 - \alpha)h_{ik}] d_{ik}^2, \quad (13)$$

where α is a parameter that is intended to mix fuzzy and hard c -means clustering, similarly to s-FCM, but creating a pure mixture of FCM and HCM. It is obvious,

that there are two values of α , where s-FCM and Os-FCM coincide: 0 and 1, corresponding to HCM and FCM, respectively. In case of any other α , s-FCM and Os-FCM differ. The AO iteration formulae of Os-FCM are easy to obtain via the well-known technique of Lagrange multipliers. The update criteria we obtain for u_{ik} and h_{ik} are the same as in case of FCM and HCM algorithms, respectively. However, the update formula for cluster prototypes becomes:

$$\mathbf{v}_i = \frac{\sum_{k=1}^n [\alpha u_{ik}^m + (1 - \alpha)h_{ik}] \mathbf{x}_k}{\sum_{k=1}^n [\alpha u_{ik}^m + (1 - \alpha)h_{ik}]} \quad \forall i = 1 \dots c \quad (14)$$

In the followings, we will compare the s-FCM algorithm with our newly proposed optimal clustering model, from two different points of view: (1) we will compute the quasi-competitive learning rate of Os-FCM and compare it with the one of s-FCM; (2) we will analyze the behavior of both algorithms by employing them to cluster the IRIS data [1] with several different settings.

Using the notations defined at the computation of the learning rate of s-FCM, and based on Eq. (14), we can write the following equation in these new circumstances:

$$\alpha \left(\frac{\delta_{w_k k}}{\sum_{j=1}^c \delta_{jk}} \right)^m + (1 - \alpha) = \left(\frac{\gamma \delta_{w_k k}}{(\gamma - 1)\delta_{w_k k} + \sum_{j=1}^c \delta_{jk}} \right)^m, \quad (15)$$

which represents the weighting coefficient received by a vector \mathbf{x}_k whose competition was won by cluster prototype \mathbf{v}_i . According to Eq. (1), we can transcribe the following equation as:

$$\alpha u_{w_k k}^m + (1 - \alpha) = \left(\frac{\gamma u_{w_k k}}{1 + (\gamma - 1)u_{w_k k}} \right)^m, \quad (16)$$

which is an expression that is difficult to solve analytically. Let us ease the circumstances by setting $m = 2$, which then yields:

$$\gamma^2 u_{w_k k}^2 = [1 - \alpha + \alpha u_{w_k k}^2][1 + (\gamma - 1)u_{w_k k}]^2. \quad (17)$$

This leads to the following second order equation in γ :

$$\alpha u_{w_k k}^2 (1 + u_{w_k k}) \gamma^2 - 2u_{w_k k} [1 - \alpha + \alpha u_{w_k k}^2] \gamma - (1 - u_{w_k k}) [1 - \alpha + \alpha u_{w_k k}^2] = 0, \quad (18)$$

which has only one acceptable (non-negative) solution:

$$\gamma = \frac{1 - \alpha + \alpha u_{w_k k}^2 + \sqrt{1 - \alpha + \alpha u_{w_k k}^2}}{\alpha u_{w_k k} (1 + u_{w_k k})}. \quad (19)$$

Let us verify the extreme values: if $\alpha \rightarrow 0$, then $\gamma \rightarrow \infty$, which is what we expected. Also, if we set $\alpha \rightarrow 1$, we get $\gamma \rightarrow \frac{u_{w_k k}^2 + u_{w_k k}}{u_{w_k k} (1 + u_{w_k k})} = 1$, which means zero learning rate, corresponding to FCM. So the learning rate is given by:

$$\eta_{\text{Os-FCM}} = 1 - \gamma^{\frac{1-m}{2}} = 1 - \sqrt{\frac{\alpha u_{w_k k} (1 + u_{w_k k})}{1 - \alpha + \alpha u_{w_k k}^2 + \sqrt{1 - \alpha + \alpha u_{w_k k}^2}}}, \quad (20)$$

its graphical representation is shown in Fig. 1(c). The curves on in this graph are quite similar but not identical with the ones shown in Fig. 1(b). This is not yet a proof for similar behavior of the two algorithms, this is what was possible to show the analytical way. The comparison of s-FCM and Os-FCM should continue with a numerical analysis.

4 Numerical Analysis

In the followings we will present some numerical analysis of the functional characteristics of the suppressed FCM algorithm. These tests are performed using the IRIS data, which consist of 150 labeled feature vectors of four dimensions.

A series of numerical tests targeted the clustering accuracy. It is well-known that the IRIS data cannot be perfectly classified without using the labels for supervised learning. In case of unsupervised clustering, a deterministic misclassification rate of 10% represents fine accuracy. In case of FCM, HCM, or s-FCM, this accuracy is reached only if an intelligent cluster prototype initialization scheme is applied. If the initial prototypes are not properly chosen, the algorithms might fail. Generally FCM is less sensitive to initialization than HCM.

We have tested the clustering accuracy and robustness, intentionally using a least smart initialization: randomly chosen input vectors, differing from each other, were set as initial cluster prototypes. The left columns in Fig. 2(left) show the failure percentage of s-FCM vs. the suppression rate α . These results suggest, that s-FCM requires the proper initialization as much as HCM does. If the initial prototypes are chosen randomly, but choosing exactly one from each label class, then the confusion rate reduces to zero for all FCM, HCM, and s-FCM.

The right columns of in Fig. 2(left) show the number of necessary iterations vs. suppression rate, using the same threshold value in the stopping criterion. These values are in full accordance to those found in [4]: a suitably chosen suppression rate may reduce 2-3 times the necessary computation cycles. Taking into consideration, that an efficient implementation uses a few percent more

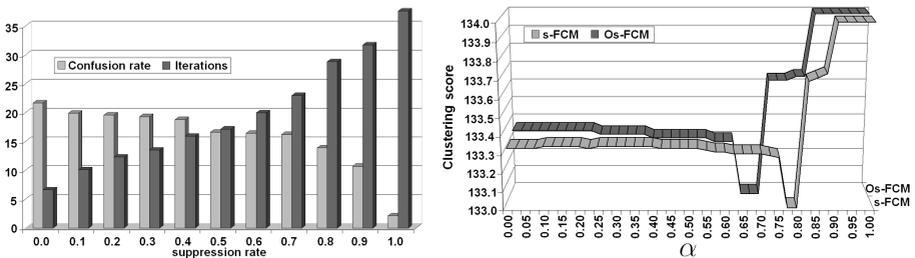


Fig. 2. (left) Failure rate (percentage) in case of random initialization, and the number necessary iterations to reach a strict convergence criterion, represented vs. the suppression rate; (right) Clustering score of s-FCM and Os-FCM, represented vs. the suppression rate.

computations to perform the fuzzy membership modification, we can conclude that suppressing the FCM algorithms is a useful idea.

Further tests have been executed to establish the differences and similarities between the behavior of FCM and s-FCM. We have found that both need practically the same amount of iterations to reach a given level of convergence at a predefined value of α , and the graphical plot of their classification score not only shows the same numbers, but also their shape is similar (see Fig. 2(right)).

5 Conclusions

In this paper we assumed to study the properties of the suppressed fuzzy c-means algorithm. Based on analytical computations, we have exploited the quasi-competitive behavior of s-FCM. On the other hand, using numerical analysis, we have shown, that s-FCM succeeded to inherit the quick convergence of HCM, and the accuracy of FCM, but also has the disadvantage of being almost as sensitive to prototype initialization as HCM is. We have proposed an optimal version of the suppressed FCM algorithm and compared their properties via analytical computations and numerical tests. We found only slight differences between their performance: Os-FCM should be favored mostly because of its guaranteed optimality.

References

1. Anderson, E.: The IRISes of the Gaspé peninsula. *Bull. Amer. IRIS Soc.* 59, 2–5 (1935)
2. Bezdek, J.C.: *Pattern recognition with fuzzy objective function algorithms*. Plenum, New York (1981)
3. Cannon, R.L., Dave, J.V., Bezdek, J.C.: Efficient implementation of the fuzzy c-means clustering algorithms. *IEEE Trans. Patt. Anal. Machine Intell.* 8, 248–255 (1986)
4. Fan, J.L., Zhen, W.Z., Xie, W.X.: Suppressed fuzzy c-means clustering algorithm. *Patt. Recogn. Lett.* 24, 1607–1612 (2003)
5. Hung, W.L., Yang, M.S., Chen, D.H.: Parameter selection for suppressed fuzzy c-means with an application to MRI segmentation. *Patt. Recogn. Lett.* 27, 424–438 (2006)
6. Kamel, M.S., Selim, S.Z.: New algorithms for solving the fuzzy clustering problem. *Patt. Recogn.* 27, 421–428 (1994)
7. Karayiannis, N.B., Bezdek, J.C.: An integrated approach to fuzzy learning vector quantization and fuzzy c-means clustering. *IEEE Trans. Fuzzy Syst.* 5, 622–628 (1997)
8. Wei, L.M., Xie, W.X.: Rival checked fuzzy c-means algorithm. *Acta Electr. Sin.* 28, 63–66 (2000)
9. Tsao, E.C.K., Bezdek, J.C., Pal, N.R.: Fuzzy Kohonen clustering networks. *Patt. Recogn.* 27, 757–764 (1994)
10. Yair, E., Zeger, K., Gersho, A.: Competitive learning and soft competition for vector quantization design. *IEEE Trans. Sign. Proc.* 40, 294–309 (1992)