ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

Median Variant of Fuzzy c-Means

Tina Geweniger[1,*] Dietlind Zühlke[2] and Barbara Hammer[3] and Thomas Villmann[4]

[1]University of Leipzig - Med. Dep., Computational Intelligence Group
Semmelweisstrasse 10, 04103 Leipzig - Germany

[2]Fraunhofer Institute for Applied Information Technology - Life Science Informatics
Schloss Birlinghoven, 53229 Sankt Augustin - Germany

[3]Clausthal University of Technology, Institute of Computer Science,
Clausthal-Zellerfeld - Germany

[4]University of Applied Sciences Mittweida
Department of Mathematics/Physics/Informatics, Computational Intelligence Group
Technikumplatz 17, 09648 Mittweida - Germany

Abstract. In this paper we introduce Median Fuzzy C-Means (M-FCM). This algorithm extends the Median C-Means (MCM) algorithm by allowing fuzzy values for the cluster assignments.

To evaluate the performance of M-FCM, we compare the results with the clustering obtained by employing MCM and Median Neural Gas (MNG).

1 Introduction

Clustering of objects is a main task in machine learning. Thereby one can distinguish between hard and soft variants. For both strategies advanced methods are developed. Popular approaches favor prototype based adaptive algorithms. Usually, these algorithms belong to the class of vector quantizers, requiring the data objects and the prototypes to be embedded in a metric vector space. Recent designs relax this last condition only demanding the existence of a similarity relation between the data objects given by a data similarity matrix **D** [1],[2]. Yet, these approaches are only crisp quantizer.

The aim of this paper is to extend the standard *fuzzy c-means* (FCM, [3]) algorithm to handle data objects, if only similarities between them are given, i.e. we propose the median variant of FCM. For this purpose, we demonstrate that the algorithm follows a gradient descent on a cost function which converges in a limited number of time steps. Finally, we give examplary applications of the algorithm for artificial and real world data.

2 The Median Fuzzy c-Means (M-FCM)

First, to derive the M-FCM, we briefly reconsider *median c-means*, which only takes the distances $d\left(x^j, x^i\right)$ between the objects itself into account.

* *corresponding author, email: tg@sonowin.de*

2.1 Median c-means

Median c-means (M-CM) is a variant of classic *c-means* [1],[3]. The cost function for M-CM is given by

$$E = \sum_{i=1}^{p} \sum_{j=1}^{n} X_{I(x^j)}(i) \cdot d\left(x^j, w^i\right) \tag{1}$$

where p is the cardinality of the set $W = \left\{w^k\right\}$ of the prototypes and n the number of data points $x_i \in X = \{x_1, ..., x_n\}$. $X_{I(x^j)}(i)$ is the characteristic function of the winner index $I(x^j)$, which refers to the index of the prototype with minimum distance to x^j (winner).

E is optimized by iteration through the following two adaptation steps until convergence is reached.

1. determine the winner $I(x^j)$ for each data point x^j
2. Since for proximity data only the distance matrix is available the new prototype i has to be chosen from the set X of data points with $w^i = x^l$ where
$$l = \underset{l'}{\operatorname{argmin}} \sum_{j=1}^{p} X_{I(x^j)}(l) \cdot d(x^j, x^{l'}). \tag{2}$$

2.2 Median Fuzzy c-Means

We now turn to the *median fuzzy c-means* (M-FCM) which merges M-CM and FCM. It allows fuzzy assigments of the objects to the cluster prototypes like in FCM, which are restricted to be objects itself according to M-CM. In the following, first a general derivation of the algorithm is given and afterwards two special object metrics - euclidian metric and β-metric - are examined.

The cost function of the M-FCM is exactly that one of the standard FCM:

$$E = \sum_{i=1}^{p} \sum_{j=1}^{n} \left(\Psi_j(x_i)\right)^m d(x_i, w_j) \tag{3}$$

where $\Psi_j(x_i)$ are the fuzzy assignments specified later. The similarity between the data is given by the matrix $\mathbf{D}$, with $D_{ij} = d(x_i, x_j)$. The exponent m determines the fuzzyness and is commonly set to $m \geq 1.5$.

There exist two variants of the cost function depending on the fuzzy assignments $\Psi_j(x_i)$:

- probabilistic: $\Psi_j(x_i) \geq 0$ with $\sum_j \Psi_j(x_i) = 1$
- possibilistic: $\Psi_j(x_i) \geq 0$

We restrict ourself to the probabilistic variant. Starting from a random assignment of prototypes the M-FCM algorithm performs a two-step iteration to minimize the cost function by adapting the prototypes as well as the assigments Ψ as known from M-CM and FCM:

1. $\Psi_j(x_i) = \frac{\sqrt[m-1]{d(x_i,w_j)^{-1}}}{\sum_k \sqrt[m-1]{d(x_i,w_k)^{-1}}}$

2. $w_j = x_l \qquad\qquad l = \arg\min_k[\sum_i \Psi_j(x_i)^m d(x_i, x_k)]$

We have to show that this algorithm follows a gradient descent on (3). For this purpose, we have to analyze the derivatives showing them to be non-negative. We utilize the Lagrangian method

$$L(\boldsymbol{\lambda}) = E + \sum_q \lambda_q(\sum_k \Psi_k(x_q) - 1) \tag{4}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{p} \Psi_j(x_i)^m d(x_i, w_j) + \sum_{i=1}^{n} \lambda_i(\sum_{j=1}^{p} \Psi_j(x_i) - 1), \tag{5}$$

whereby we plugged (3) into (4) for the second line. Using the first derivatives

$$\begin{aligned}\frac{\partial L}{\partial \Psi_j(x_i)} &= m\Psi_j(x_i)^{m-1} d(x_i, w_j) + \lambda_i \\ &= L_{j,i}\end{aligned}$$

the second derivatives according to the assignments are obtained as

$$\frac{\partial L_{j,i}}{\partial \Psi_{j'}(x_{i'})} = \begin{cases} 0 & i \neq i', j \neq j' \\ m(m-1)\Psi_{j'}(x_{i'})^{m-2} d(x_{i'}, w_{j'}) & \text{sonst} \end{cases}.$$

and, hence, are non-negative. Looking further at

$$\begin{aligned}\frac{\partial L}{\partial w_j} &= \sum_i \Psi_j(x_i)^m \; \frac{\partial d(x_i, w_j)}{\partial w_j} \\ &= L^j\end{aligned}$$

we derive

$$\frac{\partial L^j}{\partial w_{j'}} = \begin{cases} 0 & j \neq j' \\ \sum_{i,j} \Psi_{j'}(x_i)^m \frac{\partial^2 d(x_i, w_{j'})}{\partial w_j \partial w_{j'}} & \text{otherwise} \end{cases}. \tag{6}$$

Since $\sum_{i,j} \Psi_{j'}(x_i)^m$ is always non-negative, it has to be assured that $d(x_i, w_j)$ and $\frac{\partial^2 d(x_i, w_j)}{\partial w_j \partial w_{j'}}$ respectively are positive. Then the convergence follows from gradient descent and the always non-negative cost function.

Usually, the similarity measure d is not known. However, to demonstrate the flexibility of the algorithm to follow different metrics (although unknown but implicitely given by the matrix $\mathbf{D}$), we illustrate the approach for two similarity measures. For the Euclidean metric $d(x, y) = ||x - y||^2$ it is obvious. More challenging is the second example

$$d(x, y) = 1 - e^{(-\beta||x-y||^2)}$$

ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

denoted as β-metric with $\beta > 0$ [4][1]. The second derivate needed in (6) yields

$$\begin{aligned} \frac{\partial^2 d(x,y)}{\partial^2 y} &= e^{(-\beta||x-y||^2)}((2\beta(x-y))+1)+2\beta \\ &\geq 0. \end{aligned}$$

Hence, the β-distance fulfills the above requirements.

3 Experiments

We perform several exemplary experiments to show the abilities of the new M-FCM. The first experiments are clustering of overlapping Gaussians generating the distance matrix using the Euclidean metric. Subsequently, we apply the algorithm to cluster text fragments of dialogs from psychotherapy sessions.

3.1 Measures for comparing results

To measure the agreement of two different cluster solutions, we applied Fuzzy Cohen's Kappa $\kappa_C \in [-1, 1]$ [5]. This measure yields the statement that, if the result is greater than zero, the cluster agreements are not random but systematic. The maximum value is perfect agreement [6].

3.2 Overlapping Gaussians

The dataset for the first experiment is created from five overlapping Gaussian distributions. We used two different settings: in the first setting there is a substantial overlap whereas for the second run the Gaussians are separated more clearly. As shown in Fig. 1[2], a fuzzy cluster assignment reflecting the diffuse separation of the initial clusters was achieved by M-FCM for both settings.

Additionally, we compare the result with clusterings obtained by FCM. The comparison is done in terms of the κ -agreement. The κ-agreement between the fuzzy clusterings of FCM and M-FCM is 0.76209 indicating a substantial agreement [7].

3.3 Psychotherapy session transcripts

In a real world experiment, we analyze text transcripts of a series of 37 psychotherapy session dialogs from a psychodynamic therapy. It is known that the therapy was a two-phase process with the culminating point around session 17 [8]. This fact is based on the evaluation of several clinical therapy measures [8]. The similarity between the transcripts is determined using the universal distance description length (*Kolmogorov complexity)* estimated by the file length of the compressed texts, for details see [9].

As shown in [9], algorithms like M-CM and Median Neural Gas (M-NG) find similar cluster solutions separating the transcripts into two groups reflecting the

[1] Although it is not a metric in the strong mathematical sense we will denote as metric. To be a metric one has to put $d(x, y) = \sqrt{1 - e^{(-\beta||x-y||^2)}}$.

[2] Colored versions of the images can be obtained from the corresponding author on request.

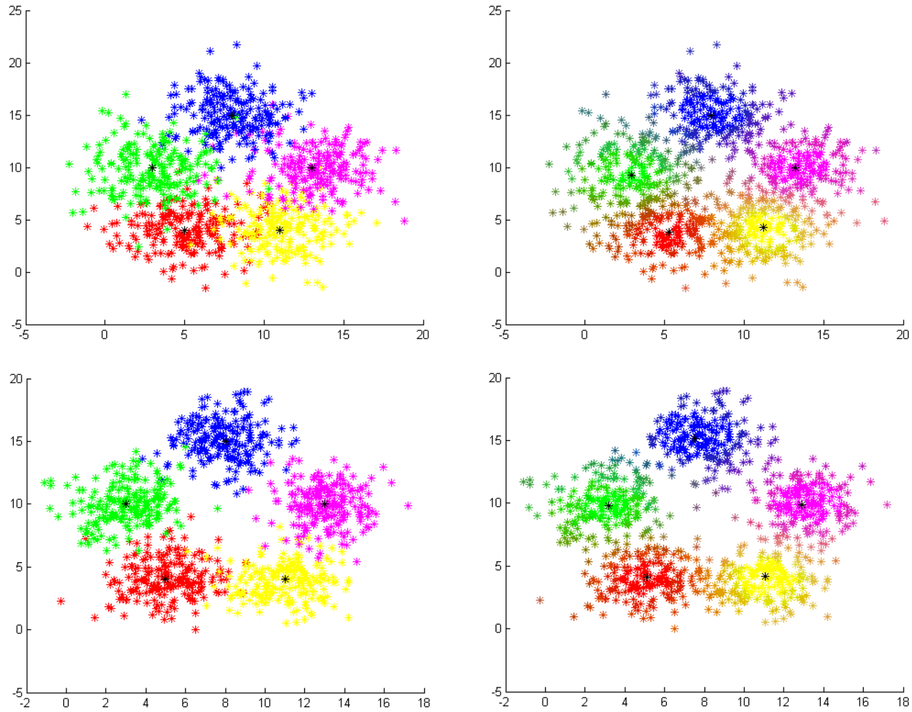

Fig. 1: Original Gaussians (left) - moderate overlapping (top) and slightly overlapping (bottom). On the right, fuzzy classifications according to M-FCM. The fuzzy assignment values are color coded. The black stars are the cluster centers. (The colored figure can be obtained from the corresponding author on request.)

break through in therapy. This concordance led to the hypothesis that narratives of the psychotherapy can be related to the therapeutic process [8].

Clustering the same data using fuzzy assignments as with M-FCM shows a smooth transition from phase one to phase two, see Fig.2. However, a more fine tuned decision can be observed.

4 Conclusion

We developed a median variant of *fuzzy c-means* which is obtained by a combination of FCM with classical *median c-means*. We have shown that the algorithm follows a gradient descent on a cost function with convergence. Exemplary applications demonstrate the abilities of the new approach.

References

[1] M. Cottrell, B. Hammer, A. Hasenfuß, and T. Villmann. Batch and median neural gas. *Neural Networks*, 19:762–771, 2006.

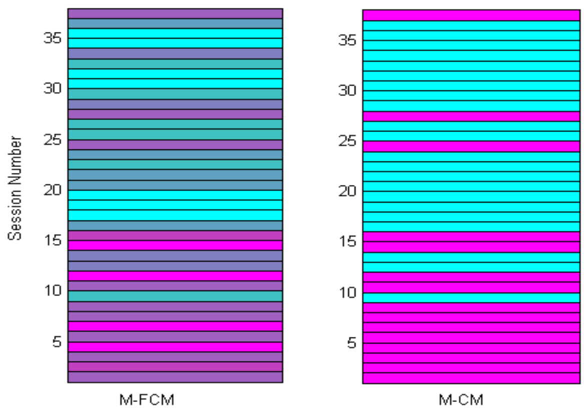

Fig. 2: Transition of psychotherapy sessions from phase 1 to phase 2. The fuzzy assignments are color coded.[2] In agreement with clinical findings there is a transition around the 17th therapy session. (The colored figure can be obtained from the corresponding author on request.)

[2] R.J. Hathaway and J.C. Bezdek. NERF c-means: Non-Euclidean relational fuzzy clustering. *Pattern recognition*, 3(27):429–437, 1994.

[3] J.C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum, New York, 1981.

[4] K.L. Wu and M.S. Young. Alternative c-means clustering algorithms. *Pattern recognition*, 35:2267–2278, 2002.

[5] W. Dou, Y. Ren, Q. Wu, S. Ruan, Y. Chen, and D. Bloyet AndJ.-M. Constans. Fuzzy kappa for the agreement measure of fuzzy classifications. *Neurocomputing*, 70:726–734, 2007.

[6] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46, 1960.

[7] L. Sachs. *Angewandte Statistik*. Springer Verlag, 7-th edition, 1992.

[8] T. Villmann, C. Liebers, B. Bergmann, A. Gumz, and M. Geyer. Investigation of psychophysiological interactions between patient and therapist during a psychodynamic therapy and their relation to speech in terms of entropy analysis using a neural network approach. *New Ideas in Psychology*, 26:309–325, 2008.

[9] T. Geweniger, F.-M. Schleif, A. Hasenfuss, B. Hammer, and T. Villmann. Comparison of cluster algorithms for the analysis of text data using kolmogorov complexity. In *Proc. of the International Conferens on Neural Information Processing (ICONIP'2008)*, page in press. Springer, 2009.