ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

An adjustable p-exponential clustering algorithm

Valmir Macario[12] and Francisco de A. T. de Carvalho[2] *

1- Universidade Federal Rural de Pernambuco - Deinfo
Rua Dom Manoel de Medeiros, s/n - Campus Dois Irmãos CEP: 52171-900 - Recife/PE - Brazil

2- Universidade Federal de Pernambuco - Cin
Av. Jornalista Anibal Fernandes, s/n - Cidade Universitária CEP: 50740-560 - Recife/PE - Brazil

Abstract. This paper proposes a new exponential clustering algorithm (XPFCM) by reformulating the clustering objective function with an additional parameter p to adjust the exponential behavior for membership assignment. The clustering experiments show that the proposed method assign data to the clusters better than other fuzzy C-means (FCM) variants.

1 Introduction

Clustering is the agglomeration of objects (or samples) into clusters so that objects within the same cluster are more similar, according to some similarity measures, while objects from different clusters have a lower similarity. The clustering goal is maximize the homogeneity of the objects in same cluster while maximizing the heterogeneity of objects in different clusters [1]. Fuzzy clustering [2] provides an additional conceptual enhancement by allowing the sample to be allocated to several clusters (classes) to various degrees (membership values). By this, patterns can be treated more realistically and the analysis is capable of identifying eventual outliers.

A number of methods have been proposed to improve the performance of fuzzy clustering algorithms [3]. For example, Myamoto [4] adds a regularized term based on a Shannon entropy [5] to guide the clustering process. This entropy regularized term is often used as a validity clustering criteria. Recently, Kannan et al. [6] propose new objective functions with quadratic entropy regularization and mean quadratic entropy regularization to enhance the flexibility in obtaining clusters with more noised data.

The entropy based FCM algorithm performs well with noise-free data, it obtains a rather poor result when having to deal with data corrupted by noise, and other artifacts, as is often the case with real-world dataset [6]. To address this issue, we propose in this work a new objective function with exponential behavior based on the previous work of Treerattanapitak and Jaruskulchai [7]. The exponential function has a quite aggressive attribution of pertinence degrees. In this work, we have an additional parameter "p" that multiplies the pertinence degree, that controls the influence of the exponential function to the award of the pertinence degree. Thus, we expect that for different values of these parameters, the objective function can escape from local minima. Furthermore, in this work we have used the Karush-Kuhn-Tucker (KKT) optimization, which avoids pertinence degree with negative values. With experiments on synthetic and real data sets, we have verified that our proposed method has obtained better performance than other clustering algorithms.

*The authors would like to thank CNPq and FACEPE (Brazilian agencies) for their financial support

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

The paper is organized as follows. In Section 2, we present the new exponential fuzzy clustering. In Section 3, we perform experiments to validate our proposed algorithm against other clustering algorithms. In Section 5, we draw a conclusion and make recommendations for the future work.

2 An adjustable p-exponential fuzzy clustering (XPFCM)

The idea consists of developing a clustering algorithm that handles an aggressive membership degree attribution using an exponential function. The behavior of the proposed objective function is to force a high value of a membership degree u_{ik} in matrix U be associated with a low value of a distance d_{ik} to a cluster prototype v_k from matrix V, where $(i = 1, \ldots, n; k = 1, \ldots, C)$, n is the number of items x_i, C is the number of clusters and $d_{ik} = (x_i - v_k)^T(x_i - v_k)$ is the Euclidean distance. In this situation, the objective function is optimized in a exponentially, only if the distance d_{ik} is low. However, the exponential performance is controlled by two parameters m and p. The parameter m is raised by the membership degree u_{ik}, multiplied by another parameter p. The parameter p will decreasing or increasing the u_{ik} exponentiation. Thus, the proposed objective function of the XPFCM algorithm is:

$$J_{XPFCM}(\mathbf{U},\mathbf{V}) = \sum_{k=1}^{C}\sum_{i=1}^{n}(m^{p*u_{ik}} - 1)d^2(x_i, v_k) \tag{1}$$

$$\tag{2}$$

where the parameter $p > 0$ adjusts the exponent of the exponential function. The membership degree functions are subject to the constraints

$$u_{ik} \geq 0 \; for \; i = 1, ..., n; k = 1, ..., C \tag{3}$$
$$\sum_{k=1}^{C} u_{ik} = 1 \; for \; i = 1, ..., n$$

The minimization of equation (1) can be performed by Lagrange equation L_{XPFCM}:

$$\begin{aligned} L_{XPFCM}(\mathbf{U},\mathbf{V}) &= \sum_{k=1}^{C}\sum_{i=1}^{n}(m^{p*u_{ik}} - 1)d^2(x_i, v_k) \\ &- \sum_{i}\lambda_i\left(\sum_{k} u_{ik-1}\right) - \sum_{i}\sum_{k}\psi_{ik}u_{ik} \end{aligned} \tag{4}$$

whereλ_i and ψ_{ik} are known as Lagrange multipliers. Using objective function (4) and the constraints (3) of the original minimization problem, we can write down the corresponding Karush-Kuhn-Tucker conditions:

$$\psi_{ik} \geq 0 \tag{5}$$
$$\frac{\partial L_{XPFCM}}{\partial u_{ik}} = 0 \tag{6}$$
$$u_{ik}\psi_{ik} = 0 \tag{7}$$

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

Considering the conditions (5)(6)(7), the relation can take one of two forms for each object x_i
1. $\psi_{ik} = 0$ for $k = 1, ..., C$ so that

$$u_{ik} = \frac{\frac{1}{C} * \left(p * \log(m) - C * \log(d_{ik}) + \sum_{l=1}^{C} \log(d_{il})\right)}{p * \log(m)} \tag{8}$$

and it is valid only if $u_{ik} \geq 0$ for all objects. If this condition is not fulfilled, we have to consider the alternative form:
2. $\psi > 0$ for at least some k. Because Eq. (3) must also be satisfied, it is clear that this solution is not valid to all k of one object i. Hence let us define the partition:

$$\begin{aligned} V- &= \{u_{ik} = 0\} \\ V+ &= \{u_{ik} > 0 \Rightarrow \psi_{ik} = 0\} \neq \emptyset \end{aligned} \tag{9}$$

To compute partition $V-$ we take in consideration the constraint (3), resulting in $\psi = 0$. Beside, if $k \in V+$, we have an equation analogous to the equation (8).
To compute the centroid, the resulting equation is

$$\mathbf{v}_k = \frac{\sum_{i=1}^{n}(m^{p*u_{ik}} - 1)x_i}{\sum_{i=1}^{n}(m^{p*u_{ik}} - 1)} \tag{10}$$

Now we can formulating the algorithm as follow:
SCHEMA OF THE XPFCM CLUSTERING ALGORITHM

1. **Initialization**

a. Fix the number C of clusters; Fix $MaxIter$ (maximum number of iterations); Fix $\epsilon >> 0$; Fix $s = 0$ (iteration count)

b. Randomly select C distinct prototypes $v_k^{(0)} \in X = x_1, x_2, \ldots, x_N (k = 1, \ldots, C)$;

c. Compute the membership degrees:
$u_i^{(0)} = (u_{i1}^{(0)}, \ldots, u_{iC}^{(0)})(i = 1, \ldots, N)$
with
$u_{ik}^{(0)} = \left[\sum_{l=1}^{C} \left(\frac{(d_{ik}^{(0)})^2}{(d_{il}^{(0)})^2}\right)\right]^{-1}$

Repeat :

2. **Representation step:**
Compute the prototypes $v_k (k = 1, \ldots, C)$ using equation (10)
3. **Allocation step:**
Compute the fuzzy membership degree u_{ik} of data point $x_i (i = 1, \ldots, n)$ into cluster v_k $(k = 1, \ldots, C)$ using equation (8)
If $u_{ik} <= 0 \Rightarrow V- = V - \cup\, v_k \Rightarrow u_{ik} = 0$
If $u_{ik} > 0 \Rightarrow V+ = V + \cup\, v_k \Rightarrow u_{ik} = u_{ik}$

Until $|U^{s+1} - U^s| \leq \epsilon$ **or** $s > MaxIter$

3 Experiments

To evaluate the new clustering algorithm we have performed clustering tasks in real and synthetic data sets. The real data sets are Iris and Ionosphere from UCI repository.

The iris data set contains 150 samples describing 3 types of iris plants. There are 4 attributes describing the length and width measures of petals and sepals of the plant. There are 50 samples to represent each plant.

The Ionosphere data set analysis the quality of a radar returns from ionosphere. The targets were free electrons in the ionosphere. "Good" radar returns are those showing evidence of some type of structure in the ionosphere. "Bad" returns are those that do not, their signals pass through the ionosphere. The data set consists of 351 instances described by 34 real-valued attributes, being 225 instances of good returns and 126 instances of bad radar returns.

The synthetic data set Cassini can be generated using the mlbench library in R language. The inputs are uniformly distributed on a 2-dimensional space within 3 structures. The 2 external structures (classes) are banana-shaped structures and in between them, the middle structure (class) is a circle. There are 500 total examples, being 200 from class 1 and 2 and 100 from class 3.

The result of the clustering algorithms in this work are a fuzzy partition into C fuzzy clusters. Then, a hard hard partition $R = (R_1, \ldots, R_C)$ is obtained from this fuzzy partition by defining the cluster $R_k(k = 1, \ldots, C)$ as: $R_k : \{x_i \in X : u_{ik} \geq u_{im}, m \in \{1, \ldots, C\} \ and \ m \neq k\}$. So, each example is associated to that cluster whose the membership degree is the greatest. So that, we create the final hard partition which contain all examples associated to one cluster.

To compare the clustering results furnished by the clustering methods, an external index - the corrected Rand index (CR) [8] - and the overall error rate of classification (OERC) [9] will be considered.

The purpose of the experiment applied in this work is evaluate the proposed clustering algorithm against other algorithms with similar characteristics, like the Fuzzy C-Means (FCM) proposed by Bezdek [2], and the exponential algorithm (XFCM) proposed by Treerattanapitak and Jaruskulchai [7].

In this experiment, we have ranged the parameter settings of each algorithm in order to get the best clustering acccuracy. For the algorithms FCM, XFCM and XPFCM we have ranged the fuzzifier value of m from 1.5 to 5, including the euler exponential number $e = 2.71\ldots$. The XPFCM algorithm have an additional parameter p that was ranged from 0.5 to 5, including e, for each value of m. We have performed 100 iterations with different centroid initializations. The selected iteration result is that whose objective function value is the lowest. We have performed experiments with k equal to the number of priori classes to each data set.

4 Results

This sections presents the results for the clustering algorithms in a clustering task. The tables present the CR and OERC results for the algorithms Fuzzy C-Means (FCM), exponential algorithm (XFCM), and the proposed algorithm (XPFCM). The fuzzifier

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

parameter m was ranged for all algorithms. For XPFCM algorithm, we show for each m configuration, the p value obtained by the lowest value of the objective function of the XPFCM algorithm.

Table 1 presents the results obtained for the synthetic Cassini data set. We can observe that variation of parameter m is not significant for the FCM algorithm. The exponential algorithms has obtained better performances. The XFCM obtained better performance with $m = 1.5$, and the XPFCM has obtained the best overall performance with $m = 2$ and $p = 1.5$. The performance of the proposed algorithm was better than other clustering algorithms.

	m	m=1.5	m=2	m=2.5	m=e	m=3	m=3.5	m=4	m=4.5	m=5
FCM	OERC	*13.33*	*13.33*	13.33	13.33	13.33	13.33	13.33	13.33	13.33
	CR	*0.534*	*0.534*	0.532	0.53	0.529	0.529	0.528	0.528	0.528
XFCM	OERC	*5.33*	5.6	10.8	11.47	11.07	13.33	13.33	13.33	13.33
	CR	*0.817*	0.804	0.677	0.664	0.671	0.533	0.533	0.534	0.534
	p	p=2.5	p=1.5	p=1.5	p=1	p=1	p=1	p=1	p=1	p=0.5
XPFCM	OERC	0.933	**0.8**	1.33	0.93	1.2	1.47	1.6	2.67	2.27
	CR	0.964	**0.969**	0.949	0.964	0.954	0.944	0.94	0.902	0.915

Table 1: Performance of the algorithms in Cassini dataset

The results for Iris data set are presented in Table 2. The variation of the m parameter does not change the performance of the XFCM algorithm. The performance of the FCM algorithm is better for $m \geq 4.0$. The best performance was obtained by the proposed algorithm for two configurations of the experiment.

	m	m=1.5	m=2	m=2.5	m=e	m=3	m=3.5	m=4	m=4.5	m=5
FCM	OERC	7.11	7.11	7.11	7.11	7.11	7.11	*6.67*	*6.67*	*6.67*
	CR	0.73	0.73	0.73	0.716	0.729	0.729	*0.743*	*0.743*	*0.743*
XFCM	OERC	7.11	7.11	7.11	7.11	7.11	7.11	7.11	7.11	7.11
	CR	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73
	p	p=4.5	p=3	p=2	p=2	p=1.5	p=1.5	p=1.5	p=1	p=1
XPFCM	OERC	**4.44**	6.22	**4.44**	5.78	6.22	5.33	6.22	6.22	6.67
	CR	**0.815**	0.753	**0.815**	0.768	0.749	0.783	0.753	0.749	0.734

Table 2: Performance of the algorithms in Iris dataset

The results for Ionosphere data set are presented in Table 3. The performance of the FCM algorithm decreased to $m \geq 4$. The performance of the XFCM algorithm is not influenced by the m parameter. The proposed algorithm has obtained the better overall performance for $m = 5$ and $p = e$.

	m	m=1.5	m=2	m=2.5	m=e	m=3	m=3.5	m=4	m=4.5	m=5
FCM	OERC	29.06	29.06	29.06	29.06	29.06	29.06	29.63	29.63	29.63
	CR	0.173	0.173	0.173	0.173	0.173	0.173	0.163	0.163	0.163
XFCM	OERC	29.63	29.63	29.63	29.63	29.63	29.63	29.63	29.63	29.63
	CR	0.163	0.163	0.163	0.163	0.163	0.163	0.163	0.163	0.163
		p=2	p=e	p=e	p=e	p=4	p=2.5	p=0.5	p=0.5	p=e
XPFCM	OERC	28.77	21.94	21.65	23.93	29.04	29.06	29.06	28.77	**20.51**
	CR	0.177	0.311	0.317	0.268	0.173	0.172	0.173	0.178	**0.343**

Table 3: Performance of the algorithms in Ionosphere dataset

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

As observed in the experiments, the parameters have influenced the clustering algorithms performance. The proposed algorithm is very influenced by the parameters m and p. However, the variation of these parameters, allows the algorithm to obtain better performance with notable difference compared to the evaluated algorithms.

5 Conclusion

The main contribution of this paper is a reformulated exponential clustering algorithm by adding an adjustable parameter p, and optimizing the objective function with the KKT conditions. The p parameter allows an adjustment in the aggressive exponential behavior of the algorithm, while the kkt optimization avoid negative pertinence degrees. As we can observe by the experiments, the additional parameter p, and the utilization of the KKT optimization in the proposed XPFCM algorithm, gives a performance advantage in clustering task regarding the evaluated algorithms. For future, we pretend study the behavior of the m and p parameters to propose a method aiming an optimal objective function optimization.

References

[1] A.K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31:651–666, 2010.

[2] J.C. Bezdek. *Pattern Recognition With Fuzzy Objective Function Algorithms*. Plenum Press, New York:, 1981.

[3] C. Bouveyron, B. Hammer, and T. Villmann. Recent developments in clustering algorithms. *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 447–458, 2012.

[4] S. Miyamoto and M. Mukaidono. Fuzzy c-means as a regularization and maximum entropy approach. 2:86–92, 1997.

[5] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. Journal*, 27:279–423, 1948.

[6] S.R. Kannan, R. Devi, S. Ramathilagam, and K. Takezawa. Effective fcm noise clustering algorithms in medical images. *Computers in Biology and Medicine*, 43:73–83, 2013.

[7] Kiatichai Treerattanapitak and Chuleerat Jaruskulchai. Exponential fuzzy c-means for collaborative filtering. *Journal of Computer Science and Technology*, 27(3):567–576, 2012.

[8] L. J. Hubbert and P. Arabie. Comparing partitions. *Journal of Classification*, 2:63–76, 1985.

[9] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen. *Classification and Regression Trees*. Chapman and Hall/CRC, Boca Raton, 1984.