

Adaptive Fuzzy Clustering Approach Based on Evolutionary Cat Swarm Optimization

Alina Shafronenko^{1[0000-0002-8040-0279]}, Yevgeniy Bodyanskiy^{1[0000-0001-5418-2143]}

¹Kharkiv National University of Radio Electronics, Nauky Ave., 14, Kharkiv, Ukraine
yevgeniy.bodyanskiy@nure.ua, alina.shafronenko@nure.ua

Abstract. Computational intelligence methods are widely used to solve many complex problems, including, of course, traditional: Data Mining and such new directions as Dynamic Data Mining, Data Stream Mining, Big Data Mining, Web Mining, Text Mining, etc. One of the main areas of computational intelligence are evolutionary algorithms that essentially represent certain mathematical models of biological organisms evolution. In the paper adaptive methods of fuzzy clustering using on evolutionary cat swarm optimization were proposed. Using proposed approach it's possible to solve clustering task in on-line mode.

Keywords: fuzzy clustering, learning rule, online probabilistic fuzzy clustering, online possibilistic fuzzy clustering, online credibilistic fuzzy clustering, cat swarm.

1 Introduction

The task of classification in the self-learning mode (clustering) of multidimensional data is an important part of traditional intellectual analysis such as Data Mining, Dynamic Data Mining, Data Stream Mining, Big Data Mining, Web Mining [1, 2].

One of the main areas of computational intelligence are the so-called evolutionary algorithms, which are mathematical models of biological organisms evolution. The problem connected with by vector-observations, clustering often occurs in many applications of data mining, and first of all in data fuzzy clustering when processing vector-observation with different levels of probability possibility, credibility etc. can belong to more than one class.

Very effective are Kohonen self-organizing maps [3] and the evolutionary algorithms that can improve data clusterization in case, when the data are processed sequentially in online mode.

The problem of fuzzy clustering of data arrays is considered in the conditions when the formed clusters arbitrarily overlap in the space of features. The source information for solving the problem is an array of multidimensional data vectors, formed by a set of vector-observations $X = (x(1), x(2), \dots, x(k), \dots, x(N)) \subset R^n$ where k in the general case the observation number in the initial array, $x(k) = (x_1(k), \dots, x_i(k), \dots, x_n(k))^T$. The result of clustering is the partition of this array on m overlapped classes with

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

prototypes-centroids $Cl_j \in R^n$, $j = 1, 2, \dots, m$, and computing of membership levels $0 \leq U_j(k) \leq 1$ of each vector-observation $x(k)$ to every cluster Cl_j .

2 Adaptive algorithm for probabilistic fuzzy clustering (APrFC)

The goal function of popular probabilistic clustering has the form [4]

$$E(U_j(k), Cl_j) = \sum_{k=1}^N \sum_{j=1}^m U_j^\beta(k) D^2(\tilde{x}_k, Cl_j) \quad (1)$$

where $\sum_{j=1}^m U_j(k) = 1$, $0 < \sum_{k=1}^N U_j(k) < N$. Solving the nonlinear programming problem, we obtain the probabilistic fuzzy clustering algorithm

$$\begin{cases} U_j^{(\tau+1)}(k) = D^2(\tilde{x}_k, Cl_j^{(\tau)})^{\frac{1}{1-\beta}} * \left(\sum_{l=1}^m (D^2(\tilde{x}_k, Cl_l^{(\tau)}))^{\frac{1}{1-\beta}} \right)^{-1}, \\ Cl_j^{(\tau+1)} = \sum_{k=1}^N (U_j^{(\tau+1)})^\beta \tilde{x}_k * \left(\sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta \right)^{-1}, \end{cases} \quad (2)$$

in case when the $\beta = 2$ - parameter that is called fuzzyfier and defines "blurring" the boundaries between classes we come to a decision, that has the form

$$\begin{cases} U_j^{(\tau+1)}(k) = \|\tilde{x}_k - Cl_j^{(\tau)}\|^{-2} * \left(\sum_{l=1}^m \|\tilde{x}_k - Cl_l^{(\tau)}\|^{-2} \right)^{-1}, \\ Cl_j^{(\tau+1)} = \sum_{k=1}^N (U_j^{(\tau+1)}(k))^2 \tilde{x}_k * \left(\sum_{k=1}^N (U_j^{(\tau+1)}(k))^2 \right)^{-1}. \end{cases} \quad (3)$$

The process of fuzzy clustering can be organized in on-line adaptive mode

$$\begin{cases} U_j(k+1) = (D^2(\tilde{x}_{k+1}, Cl_j^{(k)}))^{\frac{1}{1-\beta}} * \left(\sum_{l=1}^m (D^2(\tilde{x}_{k+1}, Cl_l^{(k)}))^{\frac{1}{1-\beta}} \right)^{-1}, \\ Cl_j(k+1) = Cl_j(k) + \eta(k+1) U_j^\beta(k+1) (\tilde{x}_{k+1} - Cl_j(k)). \end{cases} \quad (4)$$

3 Adaptive algorithm for possibilistic fuzzy clustering (APosFC)

The goal function of possibilistic clustering has the form

$$E(U_j(k), Cl_j, \mu_j) = \sum_{k=1}^N \sum_{j=1}^m U_j^\beta(k) D^2(\tilde{x}_k, Cl_j) + \sum_{j=1}^m \mu_j \sum_{k=1}^N (1 - U_j(k))^\beta \quad (5)$$

where $\mu \geq 0$ - the scalar parameter that specifies the distance at which the level of membership equals 0.5.

Minimizing the equation (5) relatively, $U_j(k)$, Cl_j and μ_j we obtain the system of equations (6)

$$\begin{cases} U_j^{(\tau+1)}(k) = \left(1 + \left(\frac{D^2(\tilde{x}(k), Cl_j^{(\tau)})}{\mu_j^{(\tau)}} \right)^{\frac{1}{\beta-1}} \right)^{-1}, \\ Cl_j^{(\tau+1)} = \sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta \tilde{x}(k) * \left(\sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta \right)^{-1}, \\ \mu_j^{(\tau+1)} = \sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta D^2(\tilde{x}(k), Cl_j^{(\tau+1)}) * \left(\sum_{k=1}^N (U_j^{(\tau+1)}(k))^\beta \right)^{-1}, \end{cases} \quad (6)$$

in case when the $\beta = 2$ we come to a decision, that has the form (7):

$$\begin{cases} U_j^{(\tau+1)}(k) = \left(1 + \frac{\|\tilde{x}(k) - Cl_j^{(\tau)}\|^2}{\mu_j^{(\tau)}} \right)^{-1}, \\ Cl_j^{(\tau+1)} = \sum_{k=1}^N (U_j^{(\tau)}(k))^2 \tilde{x}(k) * \left(\sum_{k=1}^N (U_j^{(\tau)}(k))^2 \right)^{-1}, \\ \mu_j^{(\tau+1)} = \sum_{k=1}^N (U_j^{(\tau)}(k))^2 \|\tilde{x}(k) - Cl_j^{(\tau+1)}\|^2 * \left(\sum_{k=1}^N (U_j^{(\tau)}(k))^2 \right)^{-1}. \end{cases} \quad (7)$$

In the online mode algorithms (6), (7) can be written in recurrent form (8)

$$\begin{cases} U_j(k+1) = \left(1 + \left(\frac{D^2(\tilde{x}(k+1), Cl_j(k))}{\mu_j(k)} \right)^{\frac{1}{\beta-1}} \right)^{-1}, \\ Cl_j(k+1) = Cl_j(k) + \eta(k+1) U_j^\beta(k+1) (\tilde{x}(k+1) - Cl_j(k)), \\ \mu_j(k+1) = \sum_{p=1}^{k+1} U_j^\beta(p) D^2(\tilde{x}(p), Cl_j(k+1)) * \left(\sum_{p=1}^{k+1} U_j^\beta(p) \right)^{-1} \end{cases} \quad (8)$$

and

$$\begin{cases} U_j(k+1) = \left(1 + \frac{\|\tilde{x}(k) - c_j(k)\|^2}{\mu_j(k)} \right)^{-1}, \\ c_j(k+1) = c_j(k) + \eta(k+1)U_j^2(k+1)(\tilde{x}(k+1) - c_j(k)), \\ \mu_j(k+1) = \sum_{p=1}^{k+1} U_j^2(p) \|\tilde{x}(p) - c_j(k+1)\|^2 * \left(\sum_{p=1}^k U_j^2(p) \right)^{-1}, \end{cases} \quad (9)$$

that permits to solve the fuzzy clustering task in sequential mode.

4 Adaptive Credibilistic Fuzzy Clustering (ACrFC)

Credibilistic fuzzy clustering is associated with minimizing the goal function (10)

$$E(Cr_q(k), w_q) = \sum_{k=1}^N \sum_{q=1}^m Cr_q^\beta(k) D^2(x_k, w_q) \quad (10)$$

with constraints $0 \leq Cr_q(k) \leq 1 \forall q, k$; $\sup Cr_q(k) \geq 0,5 \forall k$; $Cr_q(k) + \sup Cr_r(k) = 1$ for any q and k for which $Cr_q(k) \geq 0,5$. Here $Cr_q(k)$ – credibility that observation x_k belongs to a cluster Cl_q . In this case, the membership level is calculated using on the membership function [5, 6]

$$U_q(k) = \varphi_q(D(x_k, Cl_q)) \quad (11)$$

where: $\varphi_q(\cdot)$ – monotonically decreases in the interval $[0, \infty]$, $\varphi_q(0) = 1$, $\varphi_q(\infty) \rightarrow 0$.

It is easy to see that function (11) is essentially measure of similarity based on distance [7]. As such a function, it was proposed in [8] to use the expression

$$U_q(k) = \left(1 + D^2(x_k, Cl_q) \right)^{-1}. \quad (12)$$

It is interesting to note that expression (12) can be rewritten in the form

$$\begin{aligned}
U_q(k) &= \left(D^2(x_k, Cl_q(k)) \right)^{\frac{1}{1-\beta}} \cdot \left(\sum_{l=1}^m \left(D^2(x_k, Cl_l(k)) \right)^{\frac{1}{1-\beta}} \right)^{-1} = \\
&= \left(D^2(x_k, Cl_q(k)) \right)^{\frac{1}{1-\beta}} \left(D^2(x_k, Cl_q(k)) \right)^{\frac{1}{1-\beta}} + \sum_{\substack{l=1 \\ l \neq q}}^m \left(D^2(x_k, Cl_l(k)) \right)^{\frac{1}{1-\beta}} \right)^{-1} = \quad (13) \\
&= \left(1 + \left(D^2(x_k, Cl_q(k)) \right)^{\frac{1}{1-\beta}} \sum_{\substack{l=1 \\ l \neq q}}^m \left(D^2(x_k, Cl_l(k)) \right)^{\frac{1}{1-\beta}} \right)^{-1}
\end{aligned}$$

that for the Euclidean metric and $\beta=2$ takes the form of a Cauchy distribution density function with a width parameter σ_q^2 [9]

$$U_q(k) = \left(1 + \frac{\|x_k - Cl_q(k)\|^2}{\sigma_q^2} \right)^{-1}, \quad (14)$$

$$\sigma_q^2 = \left(\sum_{\substack{l=1 \\ l \neq q}}^m \|x_k - Cl_l(k)\|^2 \right)^{-1}. \quad (15)$$

It is easy to see that the membership function (13) is a special case of (14) for $\sigma_q^2 = 1$.

Finally, a batch algorithm of credibilistic fuzzy clustering can be written in the form [5, 6]:

$$U_q^{(\tau+1)}(k) = \left(1 + D^2(x_k, Cl_q^{(\tau)}) \right)^{-1}, \quad (16)$$

$$U_q^{*(\tau+1)}(k) = U_q^{(\tau+1)}(k) \left(\sup_l U_l^{(\tau+1)}(k) \right)^{-1}, \quad (17)$$

$$Cr_q^{(\tau+1)}(k) = \frac{1}{2} \left(U_q^{*(\tau+1)}(k) + 1 - \sup_{l \neq q} U_l^*(k) \right), \quad (18)$$

$$Cl_q^{(\tau+1)} = \sum_{k=1}^N \left(Cr_q^{(\tau+1)}(k) \right)^\beta x_k \left(\sum_{k=1}^N \left(Cr_q^{(\tau+1)}(k) \right)^\beta \right)^{-1}. \quad (19)$$

Based on this formulas, we can introduce into consideration online version of the credibilistic fuzzy clustering method in the form

$$\left\{ \begin{array}{l}
\sigma_q^2(k+1) = \left(\sum_{\substack{l=1 \\ l \neq q}}^m \|x_{k+1} - Cl_l(k)\|^2 \right)^{-1}, \\
U_q(k+1) = \left(1 + \frac{\|x_{k+1} - Cl_q(k)\|^2}{\sigma_q^2(k+1)} \right)^{-1}, \\
U_{(k+1)}^* = U_q(k+1) \left(\sup_{l \neq q} U_l(k+1) \right)^{-1}, \\
Cr_q(k+1) = \frac{1}{2} \left(U_q^*(k+1) + 1 - \sup_{l \neq q} U_l^*(k+1) \right), \\
Cl_q(k+1) = Cl_q(k) + \eta(k+1) Cr_q^\beta(k+1) (x_{k+1} - Cl_q(k)).
\end{array} \right. \quad (20)$$

Therefore, from a computational point of view, the online algorithm for credibilistic fuzzy clustering is no more complicated than the recurrent versions of FCM and PCM, while retaining the advantages of a credibility approach.

5 Evolutionary Cat Swarm Optimization

To search global extremum of a function (1) (5) and (10) it is expedient to use the bio-inspired evolutionary particles swarm optimization algorithms [10]. Among the swarm algorithms, one of the fastest ones are the so-called algorithms of the cats swarm [11, 12], that proved to be effective in solving a wide range of Data Mining tasks.

Cat swarm-CS model of behavior, assumes that each cat cat_p of swarm consisting of Q individuals ($p=1,2,\dots,Q$), can be in one of two states: Seeking Mode (SM) and Tracing Mode (TM). In this case, the seeking mode is associated with slow movements with a slight amplitude near the initial position (space scanning in the vicinity of the current position), and the tracing mode that is determined by fast jumps with a large amplitude and allows the cat cat_p go put from local extremum, if she got there. The combination of local scanning and abrupt changes in the current state makes it more likely to find a global extremum compared to traditional multi-extrema optimization methods.

In the general case, both of these modes for each of the cats swarms can be described by the recurrent optimization procedure [13]

$$cat_p(\tau+1) = cat_p(\tau) - \alpha(cat_p(\tau) - cat_p(\tau-1)) - \eta \hat{\nabla} E_M(cat_p(\tau)) + \eta_\xi \Xi(\tau), \quad (21)$$

where $cat_p(\tau+1)$ - state of p -th cat of swarm on τ -th iteration of the search, α - parameter that determines the inertia properties of the tracing mode. In a case when $\alpha = 0$ process optimization approaches to the standard gradient search. η - seeking mode step, $\hat{\nabla}E(cat_p(\tau))$ - gradient estimate of the goal function (1), (5) and (10) in the neighborhood of the point $cat_p(\tau)$, $\Xi(\tau)$ - a random component that introduces additional stochastic motions into the tracing process, η_{ξ} - parameter that specifies the amplitude of these movements.

In this algorithm, each cat can have two parallel states: search mode and tracking mode. This approach provides a search for a global extremum in the case when the number of cats in the swarm is sufficient.

6 Experimental research

Fuzzy clustering based on evolutionary cat swarm optimization (CSO) was performed on four different data samples: Iris, Cancer, Wine and Glass. Each of the data sets has a number of parameters, presented in Table 1.

Table 1. Characteristic Parameters of the Samples

Data Set	Number of clusters	Number of attributes	Number of observations.
Iris	3	4	150
Cancer	2	9	683
Wine	3	13	178
Glass	6	8	214

Table 2. Parameters of cat swarm optimization algorithm (CSO)

Parameters	Value
SRD	Random [0,1]
Seeking memory Pool (SMP)	5
Population size	Number of clusters
r_1	Random in [0,1]
c_1	Const
SPC	Random in [0,1]
Number of iteration	Manually

Table 3. Comparative results of time processing of clustering algorithms such as Fuzzy c-means algorithm (FCM), Particle Swarm Optimization (PSO), Gauss – Seidel algorithm (GSA), CSO, APrFCCSO, APosFCCSO and ACrFCCSO

Data Set	FCM	PSO	GSA	CSO	APrFC CSO	APosFC CSO	ACrFC CSO
Cancer	0.009	0.138	0.204	0.026	0.007	0.010	0.012
Glass	0.010	0.431	0.431	0.021	0.020	0.025	0.024
Iris	0.008	0.020	0.022	0.043	0.012	0.017	0.015
Wine	0.009	0.282	0.098	0.076	0.013	0.015	0.013

Also conducted a comparative analysis of the quality of clustering data on the main characteristics quality ratings, such as: Partition Coefficient (PC), Partition Index (SC), Xie and Beni's Index (XB) of existing clustering methods and proposed method.

Table 4. Evaluation of the quality of fuzzy clustering methods data

Data clustering methods	PC	SC	XB
Fuzzy C-means	0.50	1.62	0.19
Adaptive probabilistic fuzzy clustering	0.25	1.44	0.01
Adaptive possibilistic fuzzy data clustering	0.26	1.22	0.01
Adaptive credibilistic fuzzy data clustering	0.24	1.25	0.01
Adaptive probabilistic fuzzy clustering CSO (APrFCCSO)	0.22	1.37	0.25
Adaptive possibilistic fuzzy data clustering CSO (APosFCCSO)	0.23	0.95	0.01
Adaptive credibilistic fuzzy data clustering CSO (ACrFCCSO)	0.22	0.70	0.15

Table 5. Results of clustering CSO, APrFCCSO, APosFCCSO and APosFCCSO with a different number of iterations (average error in%)

Data Set	Number of iterations CSO			Number of iterations APrFCCSO			Number of iterations APosFCCSO			Number of iterations APosFCCSO		
	50	100	150	50	100	150	50	100	150	50	100	150

Iris	23.34	20.84	21.67	17.55	14.78	16.46	17.65	14.88	16.55	17.45	14.78	16.45
Cancer	40,23	40,55	41,47	38.89	39.22	39.15	37.49	38.42	39.65	37.45	38.44	39.65
Wine	24,55	21,44	22,20	18.43	17.37	16.32	18.95	18.37	19.32	18.85	18.47	19.22
Glass	56.34	56.48	55.67	51.63	51.7	49.79	52.03	51.88	49.39	52.13	51.78	49.49

Table 6 – Comparative results of performance indicators of algorithms such as PSO, CSO, APrFCCSO, APosFCCSO and APosFCCSO for 1st database

Data Set	MSE	PSO	CSO	APrFCCSO	APosFCCSO	APosFCCSO
Iris	Best	4×10^{-7}	9×10^{-10}	8.4×10^{-11}	8.8×10^{-11}	8.56×10^{-11}
	Median	7×10^{-6}	7.3×10^{-9}	5.2×10^{-10}	5.4×10^{-10}	5.5×10^{-10}
	Worst	1.2×10^{-5}	9.3×10^{-9}	9.6×10^{-10}	9.4×10^{-10}	9.4×10^{-10}
Cance	Best	1.3×10^{-7}	8×10^{-10}	8.5×10^{-11}	8.4×10^{-11}	8.35×10^{-11}
	Median	7×10^{-6}	4.4×10^{-9}	7.6×10^{-11}	6.6×10^{-11}	6.55×10^{-11}
	Worst	2.02×10^{-5}	6.8×10^{-9}	7.8×10^{-10}	7.78×10^{-10}	7.8×10^{-10}
Wine	Best	1.4×10^{-6}	7×10^{-10}	9.5×10^{-11}	9.0×10^{-11}	8.95×10^{-11}
	Median	5×10^{-5}	4×10^{-9}	6.6×10^{-11}	6×10^{-11}	5.95×10^{-11}
	Worst	2×10^{-5}	6×10^{-9}	6.8×10^{-10}	6.7×10^{-10}	6.69×10^{-10}
Glass	Best	2.5×10^{-7}	7.9×10^{-10}	8.7×10^{-11}	8.4×10^{-11}	8.35×10^{-11}
	Median	6.9×10^{-6}	5×10^{-9}	7.7×10^{-11}	7.8×10^{-11}	7.77×10^{-11}
	Worst	3×10^{-5}	5.9×10^{-9}	7.7×10^{-10}	7.5×10^{-10}	7.49×10^{-10}

7 Conclusion

The problem of fuzzy clustering based on probabilistic, possibilistic and credibilistic online approaches was considered. The recurrent modifications of well-known batch procedures designed to solve problems of Data Stream Mining allow to process information in online mode as sequential additions to the system under consideration. Since the goal functions of fuzzy clustering in the general case are multi-extremal, was is proposed to refine the solutions using swarm evolutionary optimization

algorithms. The modification introduced on the basis of the optimization procedure cat swarms with improved properties through the use of a stochastic gradient estimate was proposed.

The experiments confirmed the effectiveness of the developed approach, which is characterized by simplicity of numerical implementation and a sufficiently high convergency rate.

References

1. Xu, R., Wunsch, D.C.: Clustering. Hoboken, N.J. John Wiley & Sons, Inc. (2009)
2. Aggarwal, C.C.: Data Mining: Text Book. Springer (2015)
3. Kohonen, T.: Self-Organizing Maps. Berlin: Springer-Verlag (1995)
4. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
5. Zhou, J., Wang, Q., Hung, C.-C., Yi, X.: Credibilistic clustering: the model and algorithms. *Int.J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, №4, pp.545-564 (2015).
6. Zhou, J., Wang, Q., Hung, C. C.: Credibilistic clustering algorithms via alternating cluster estimation. *J. Intell. Manuf.*, pp.727-738 (2017).
7. Young, F.W., Hamer, R.M.: Theory and Applications of Multidimensional Scaling-Hillsdale, N.J.: Erlbaum (1994).
8. Zhou, J., & Hung, C.-C.: A generalized approach to possibilistic clustering algorithms. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*. 15. pp.117–138. (2007).
9. Hu, Zh., Bodyanskiy, Ye, Tyshchenko, O., Shafronenko, A.: Fuzzy clustering of incomplete data by means of similarity measures. *Proc.2019 IEEE 2nd Ukr. Conf. on Electrical and Computer Engineering (UKRCON), Track 6. Lviv, Ukraine, pp.149-152 (2019).*
10. Grosan, C., Abraham, A., Chis, M.: Swarm intelligence in Data Mining - Studies in Computational Intelligence (2006)
11. Chu, S.-C., Tsai, P.-W., Pan, J.S.: Cat swarm optimization. In: *Lecture Notes in Artificial Intelligence*. Berlin Heidelberg: Springer-Verlag (2006).
12. Chu, S.-C., Tsai, P.-W.: Computational Intelligence based on the behavior of cats. In: “*Int. J. of Innovative Computing, Information, and Control*”, №1, pp.163 – 173 (2007)
13. Shafronenko, A., Bodyanskiy, Ye. Pliss, I., Patlan, K.: Fuzzy Clusterization of Distorted by Missing Observations Data Sets Using Evolutionary Optimization. In: *Proceedings “Advanced Computer Information Technologies (ACIT’2019)”*, Česke Budejovice, Czech Republic, June 5-7, 2019, pp. 217-220 (2019). doi: 10.1109/ACITT.2019.8779888
14. Hu, Zh., Bodyanskiy, Ye, Tyshchenko, O., Shafronenko, A.: Fuzzy clustering of incomplete data by means of similarity measures. In: *2019 IEEE 2nd Ukraine*

Conference on Electrical and Computer Engineering UKRCON - 2019, Conference Proceedings, July 2-6, 2019, Lviv, Ukraine, pp.149-152 (2019) doi: 10.1109/UKRCON.2019.8879844.

15. Shafronenko, A. Yu, Bodyanskiy, Ye. V., Pliss I.P.: The Fast Modification of Evolutionary Bioinspired Cat Swarm Optimization Method. In: 2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL), Sozopol, Bulgaria, 6-8 Sept. 2019, pp. 548-552. (2019) doi: 10.1109/CAOL46282.2019.9019583