GeCiM: A Novel Generalized Approach to C-Means Clustering

László Szilágyi[1,2], David Iclănzan[1,3],
Sándor M. Szilágyi[1], and Dan Dumitrescu[3]

[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
[3] Faculty of Mathematics and Computer Science,
Babeş-Bolyai University of Cluj-Napoca, Romania

Abstract. All three conventional c-means clustering algorithms have their advantages and disadvantages. This paper presents a novel generalized approach to c-means clustering: the objective function is considered to be a mixture of the FCM, PCM, and HCM objective functions. The optimal solution is obtained via evolutionary computation. Our main goal is to reveal the properties of such mixtures and to formulate some rules that yield accurate partitions.

Keywords: fuzzy c-means clustering, possibilistic c-means clustering, hard c-means clustering, evolutionary computation.

1   Introduction

Zadeh's pioneering work [19] opened new dimensions in several sciences and research fields. The influence of fuzzy logic also reached the theory of clustering, causing a division between crisp and fuzzy partitioning. The relaxation of the probabilistic constraint in fuzzy clustering techniques gave birth to the absolute or possibilistic fuzzy clustering [2]. The most popular (in the sense of most frequently used) examples for these c-means clustering categories are the k-means or hard c-means (in the followings: HCM), fuzzy c-means (FCM) [6], and possibilistic c-means (PCM) [15] algorithms, probably because of their easily understandable and implementable alternating optimization (AO) solution.

All these three algorithms have their merits and drawbacks. HCM converges quickly, but it is very sensitive to initialization. FCM is better in partition quality than HCM, but it converges much slower. PCM emerged by the relaxation of the constraint that normalizes the degrees of membership, which caused several differences. Possibilistic memberships are context-insensitive, that is, they are only influenced by the distance of the given input vector from the given cluster prototype, and is not influenced by other clusters [2]. PCM reportedly tends

to produce coincident clusters [3], as its objective function treats each cluster independently.

It is obvious, that none of these three algorithms are generally suitable for every kind of clustering problem [5]. That is why, several attempts have been made to improve these clustering techniques by combining them with each other or with other algorithms (e.g. the improved version [14] of the generalized learning vector quantization (GLVQ) [17], the generalized c-means clustering scheme [18], the fuzzy J-means heuristic fuzzy clustering [4], or the suppressed FCM [10]).

In this paper we define a generalized c-means clustering model (GeCiM) , which minimizes a linear combination of the objective functions of HCM, FCM, and PCM. Based on the reformulated objective functions given in [12], the optimal solution is reached using a genetic algorithm, and validated on the IRIS data [1].

The remainder of the paper is organized as follows. Section 2 is a brief review of the methods upon which GeCiM is built. Section 3 defines the GeCiM clustering model and presents its optimization details. Section 4 relates on the clustering results and validation, while Section 5 presents the conclusions.

2   Background

2.1   Hard, Fuzzy, and Possibilistic C-Means Clustering

By definition, clustering means the separation of a set of objects into a given number of groups based on similarities. Objects are generally described with feature vectors, which contain numerical and/or logical values of the object. Let us denote the feature vectors by x_k, $k \in \{1, 2, \ldots, n\}$.

Clusters are generally represented by a prototype, centroid, or representative element, which is either an averaged or a carefully selected vector from the class. We will denote these prototypes by v_i, $i = 1 \ldots c$, where c represents the number of clusters.

The degree to which feature vector x_k belongs to cluster i is described by a real variable u_{ik}, $i = 1 \ldots c$, $k = 1 \ldots n$. Clustering algorithms are distinguished by the membership logic and the constraints that control these membership values.

HCM and FCM clustering minimize the same objective function, with different parametrization. Their objective function is:

$$J_{\text{F\&H}} = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m ||x_k - v_i||^2 = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m d_{ik}^2 \ , \tag{1}$$

where m is the so-called fuzzyfication parameter, and d_{ik} represents the distance between vector x_k and cluster prototype v_i. In case of HCM we have $m = 1$, while FCM uses $m > 1$. HCM allows only degrees of membership equaling to 0 or 1, while in FCM, $u_{ik} \in [0, 1]$, but in case of both algorithms we have the constraint $\sum_{i=1}^{c} u_{ik} = 1$. HCM resolves ties by chance.

The objective function of PCM clustering contains an extra term, which helps us find non-trivial solutions at minimization:

$$J_P = \sum_{i=1}^{c} \sum_{k=1}^{n} \left[u_{ik}^m ||x_k - v_i||^2 + \eta_i (1 - u_{ik})^m \right] \quad , \tag{2}$$

where η_i, $i = 1 \ldots n$ represents the penalty term of cluster i, fuzzyfication rate is $m > 1$, and degrees of membership are constrained by: $u_{ik} \in [0, 1]$; $\forall k = 1 \ldots n \, \exists i \in \{1, 2, \ldots, c\}$ such that $u_{ik} > 0$; and $\forall i = 1 \ldots c$ we have $0 < \sum_{k=1}^{n} u_{ik} < n$.

All these three approaches can be optimized via AO techniques: by alternately optimizing u_{ik} with v_i fixed, and v_i with u_{ik} fixed, until cluster prototypes stabilize. The membership update rule differs from algorithm to algorithm, while in all cases the cluster prototypes are updated using the same weighted averaging formula.

2.2 Reformulated Objective Functions

Hathaway and Bezdek [12] introduced reformulated objective functions to all three approaches, by eliminating the degrees of membership u_{ik}. They found the following objective functions:

$$R_{\text{HCM}} = \sum_{k=1}^{n} \min\{d_{1k}^2, d_{2k}^2, \ldots, d_{ck}^2\} \quad , \tag{3}$$

$$R_{\text{FCM}} = \sum_{k=1}^{n} \left(\sum_{i=1}^{c} d_{ik}^{2/(1-m)} \right)^{1-m} \quad , \tag{4}$$

$$R_{\text{PCM}} = \sum_{k=1}^{n} \sum_{i=1}^{c} \left[d_{ik}^{2/(1-m)} + \eta_i^{1/(1-m)} \right]^{1-m} \quad . \tag{5}$$

The authors proved the equivalence of these objective functions with their original, and proposed using these new optimization criteria for reducing the dimensions of the search space.

In this paper, we will apply these formulae to define mixtures of the HCM, FCM, and PCM clustering algorithms, and will optimize them via evolutionary computation.

3 GeCiM: The Generalized Clustering Model

3.1 The Proposed Model

Based on the reformulated objective functions of HCM, FCM, and PCM, let us formulate the following mixture criteria:

$$J_{\text{GeCiM}} = \beta \alpha R_{\text{HCM}} + \beta(1 - \alpha) R_{\text{FCM}} + (1 - \beta) R_{\text{PCM}} \quad , \tag{6}$$

where α, $\beta \in [0,1]$ represent balancing parameters that control the tradeoff among the three terms. Obviously, setting $\beta = 0$ returns us to PCM, $\beta = \alpha = 1$ means HCM, while $\beta = 1$ and $\alpha = 0$ yield FCM clustering. Further parameters are m (applied to the FCM and PCM terms) and η_i (for PCM only).

In any non-trivial case, when $0 < \alpha, \beta < 1$, the function J_{GeCiM} is not likely to have an AO scheme for minimization. That is why, in the followings, in order to study the properties of the proposed clustering criteria, we will turn to evolutionary computation techniques.

3.2   Optimization Via Evolutionary Computation

Genetic Algorithms (GAs) [11,13] are inspired and closely linked to Darwinian evolution. They mimic many mechanisms of natural evolution like mating, mutation, and survival of the fittest, in the hope that the recursive process of recombining lower order partial results in higher order ones will ultimately yield an optimal solution.

The complex schemes arising from the computational models of evolutionary processes make these methods applicable to a wide spectrum of problems. Due to their capabilities to handle several real world problems, they gained much popularity and are applied to more and more challenging optimizations tasks [7,8]. The behavior and performance of GAs and other heuristic based global optimizers is determined by the diversification/intensification balance. Diversification generally refers to the exploration of the search space, whereas the term intensification refers to the exploitation of the accumulated search experience.

As Eq. (6) can be a highly multimodal function due to its three main components, we use a multipopulation GA approach, which creates subpopulations within the niches defined by multiple potential optima. The employment of multiple groups of individuals facilitate a good diversity of the population as subpopulations that are occupying different niches do not need to outperform each other in order to propagate, resulting in a proper exploration of the search space. Subpopulations only interact by the means of a migration process, where the best individuals from one subpopulation are copied in another subpopulation, replacing the worst individuals from the destination search niche.

Intensification is achieved by allocating a separate portion of the search space to each subpopulation. After convergence of the global search, a Nelder-Mead simplex algorithm [9,16] is further employed starting from the best solution found so far.

The genome of the individuals is formed by a 12 element floating-point vector, with values between 0 and 1, as we use the IRIS data set in normalized form. These vectors can be logically regarded as a concatenation of three four-tuples, representing potential cluster centers for the three iris specimen classes from the Fisher's normalized IRIS data [1].

To ensure a proper exploration of the solution space and to foster population diversity, in our approach we use three subpopulations, each one initially slightly biased toward a certain iris specimen class. We achieve this effect, by setting the ith; $i \in \{1, 2, 3\}$ four-tuple in the ith subpopulation to vectors randomly picked

from the ith class of the normalized iris data, in 20 % of cases. The rest of the genome partitions and the rest of 80 % individuals from the subpopulations are uniformly randomly initialized with values from 0 to 1. While artificially introducing good genes into the subpopulations, biasing them towards certain classes, this process also ensures enough initial raw genetic material in order to sustain genetic variation and the opportunity for change over successive generations or "evolution".
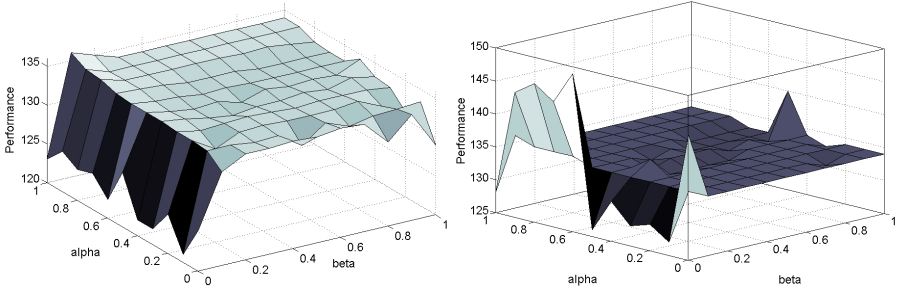

Fig. 1. Graphical representation of average (left) and maximum (right) number of correct decisions out of 150, vs. α and β tradeoff parameters

Recombination of best solutions from different subpopulations and increased intensification is promoted by the migration process, which copies the top 5 % individuals from each subpopulation to the other ones and discards the top 10 % worst individuals from every subpopulations, in every 20th generation.

Each subpopulation uses scaled fitness-proportionate selection, a crossover process that mixes up the genes with respect to original parental chromosomes in a uniform fashion and a mutation operator that mutates selected genes to a random number chosen uniformly from 0 to 1.

The subpopulation sizes were set to $SubPopulation_{size} = 120$, the mutation rate to $P_{mut} = 0.1$ and the crossover rate used was $R_c = 0.8$.

The algorithm was run for a maximum of 100 generations or until convergence was detected. The convergence criterion checks, whether the cumulative fitness improvement over the last 10 generations of the best individual is not exceeding the 10^{-6} threshold.

4   Results

The properties of the proposed generalized clustering method were tested using the IRIS data, which consists of 150 labeled vectors of four dimensions. Obviously, the clustering is performed in an unsupervised manner, that is, the labels are only used at evaluation.

We tested GeCiM with various parameter settings: α and β independently varied from 0 to 1, with steps of 0.1. The algorithm was performed 10 times
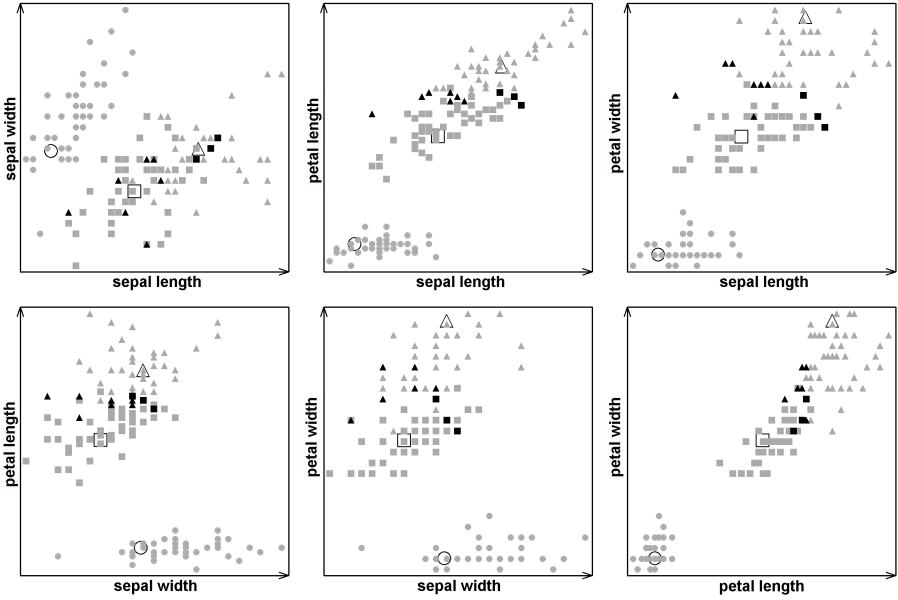
Fig. 2. Partition with 140/150 success rate signifying 6.67 % misclassifications. Shapes indicate the ground truth labeling; misclassified vectors are drawn in black. Cluster prototypes are indicated by the large circle, triangle and square.

with each parameter setting. Figure 1 shows the average and maximum number of correctly detected vectors for all parametrized cases.

Looking at the graphs in Fig. 1 we can formulate the following. Two interesting cases need to be remarked. Most of the accurate clustering techniques report 133-135 correct decisions out of 150. Our graph in Fig. 1 shows two independent spots with better decision rate. We can have 140 correct decision at $\beta = 1$, that is, when GeCiM is reduced to a mixture of HCM and FCM. This suggests that under certain circumstances, mixing FCM and HCM can lead to a superior clustering algorithm. One set of prototype, which was found to give this partition, is:

$$v_1 = \begin{pmatrix} 4.65310 \\ 3.07784 \\ 1.50021 \\ 0.20540 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 5.82625 \\ 2.70022 \\ 3.98397 \\ 1.31061 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 6.72685 \\ 3.09917 \\ 5.59478 \\ 2.43399 \end{pmatrix}. \tag{7}$$

The whole IRIS data set and its partition obtained with the vectors given above, are represented in two-dimensional projections in Fig. 2.

The other key thing that deserves a remark is the fact, that pure PCM ($\beta = 0$) is quite unstable, which means in most cases it leads to poor quality partitions (as indicated in [2]), but sometimes it finds a set of prototypes offering 149 correct decisions. One such case is depicted in eq. (8), and Fig. 3:
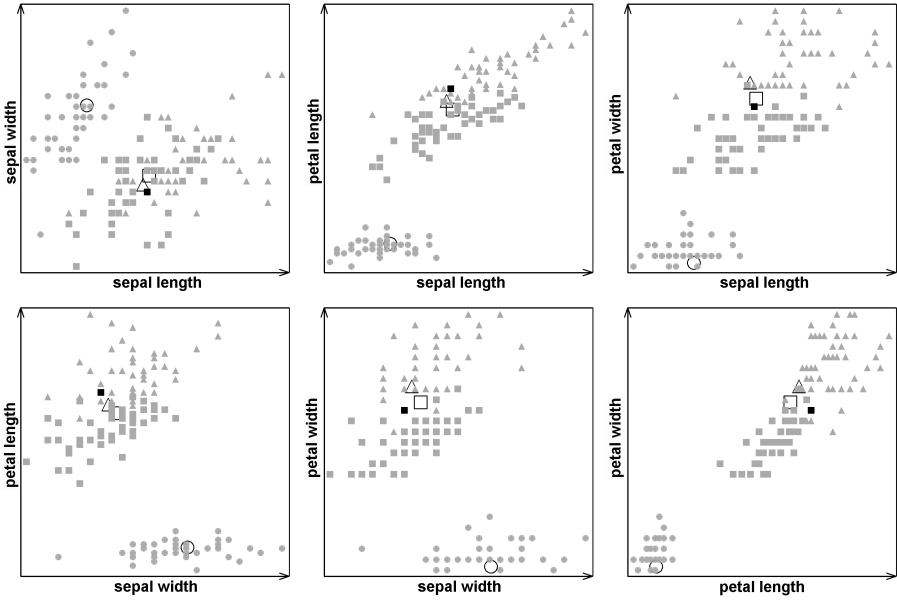
Fig. 3. Partition with 149/150 success rate signifying 0.67 % misclassifications

$$v_1 = \begin{pmatrix} 5.14843 \\ 3.51076 \\ 1.51996 \\ 0.13272 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 6.02770 \\ 2.85190 \\ 4.62071 \\ 1.67480 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 5.94317 \\ 2.76658 \\ 4.81742 \\ 1.82569 \end{pmatrix}. \tag{8}$$

These poor quality results provided by pure PCM are easily corrected by giving β a small, but non-zero value, for example 0.1. Figure 1 suggests this choice to be the best, whatever α might be. In other words, for any $\alpha \in [0, 1]$, the best choice is to set $\beta = 0.1$, and we get a stable and accurate solution with 136-137 correct decisions.

5   Conclusions

In this paper we proposed a mixture c-means clustering model, and involved a genetic algorithm for its optimization. Using the IRIS data to test and evaluate the proposed approach, we have concluded the followings:

– Mixtures of HCM and FCM, under certain circumstances, can outperform pure HCM or pure FCM.
– Pure PCM is very unstable, but mixing 90 % of PCM with some HCM and FCM show excellent in both partition quality and stability.
– The more FCM and less HCM we have in a mixture, the better the accuracy will be. However, this variation is slow.

Further works will aim at testing with lots of other data, and with smaller steps of α and β.

References

1. Anderson, E.: The IRISes of the Gaspe peninsula. Bull. Amer. IRIS Soc. 59, 2–5 (1935)
2. Baraldi, A., Blonda, P.: A survey of fuzzy clustering algorithms for pattern recognition - Part I. IEEE Trans. Syst. Man Cybern. Part B. 29, 778–785 (1999)
3. Barni, M., Capellini, V., Mecocci, A.: Comments on a possibilistic approach to clustering. IEEE Trans. Fuzzy Syst. 4, 393–396 (1996)
4. Belacel, N., Hansen, P., Mladenovic, N.: Fuzzy J-means: a new heuristic for fuzzy clustering. Patt. Recogn. 35, 2193–2200 (2002)
5. Benyó, B.: Analysis of temporal patterns of physiological parameters. In: Kamruzzaman, J., Begg, R.K., Sarker, R.A. (eds.) Artificial neural networks in finance, health and manufacturing: potential and challenges, pp. 285–317. Idea Group Publishing, Hershey (2006)
6. Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. Plenum, New York (1981)
7. Cartwright, H.: An introduction to evolutionary computation and evolutionary algorithms. In: Johnston, R., Mingos, D.M.P. (eds.) Applications of evolutionary computation in chemistry. Structure and Bonding, vol. 110, pp. 1–32. Springer, Heidelberg (2004)
8. Dasgupta, D., Michalewicz, Z. (eds.): Evolutionary algorithms in engineering applications. Springer, New York (2001)
9. Dennis Jr., J.E., Woods, D.J.: Optimization on Microcomputers: The Nelder-Mead Simplex Algorithm. In: Wouk, A. (ed.) New Computing Environments: Microcomputers in Large-Scale Computing, Soc. Ind. Appl. Math., Philadelphia, PA, pp. 116–122 (1987)
10. Fan, J.L., Zhen, W.Z., Xie, W.X.: Suppressed fuzzy c-means clustering algorithm. Patt. Recogn. Lett. 24, 1607–1612 (2003)
11. Goldberg, D.E.: Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Reading (1989)
12. Hathaway, R.J., Bezdek, J.C.: Optimization of clustering by reformulation. IEEE Trans. Fuzzy Syst. 3, 241–245 (1995)
13. Holland, J.H.: Adaptation in natural artificial systems. University of Michigan Press, Ann Arbor (1975)
14. Karayiannis, N.B., Bezdek, J.C., Pal, N.R., Hathaway, R.J., Pai, P.I.: Repairs to GLVQ: a new family of competitive learning scheme. IEEE Trans. Neural Networks 7, 1062–1071 (1996)
15. Krishnapuram, R., Keller, J.M.: A possibilistic approach to clustering. IEEE Trans. Fuzzy Syst. 1, 98–110 (1993)
16. Lagarias, J., Reeds, J., Wright, M., Wright, P.: Convergence properties of the Nelder-Mead simplex algorithm in low dimensions. SIAM J. Optim. 9, 112–147 (1998)
17. Pal, N.R., Bezdek, J.C., Tsao, E.C.K.: Generalized clustering networks and Kohonen's self-organizing scheme. IEEE Trans. Neural Networks 4, 549–557 (1993)
18. Yu, J.: General c-means clustering model. IEEE Trans. Patt. Recogn. Mach. Intell. 27, 1197–1211 (2005)
19. Zadeh, L.A.: Fuzzy sets. Inform. Contr. 8, 338–353 (1965)