

Sample-Weighted Fuzzy Clustering with Regularizations

Miin-Shen Yang and Yee-Shan Pan

Abstract—Although there have been many researches in cluster analysis to consider on feature weights, little effort is made on sample weights. Recently, Yu et al. (2011) considered a probability distribution over a data set to represent its sample weights and then proposed sample-weighted clustering algorithms. In this paper, we give a sample-weighted version of generalized fuzzy clustering regularization (GFCR), called the sample-weighted GFCR (SW-GFCR). Some experiments are considered. These experimental results and comparisons demonstrate that the proposed SW-GFCR is more effective than the most clustering algorithms.

Keywords—Clustering; fuzzy c-means; fuzzy clustering; sample weights; regularization.

I. INTRODUCTION

CLUSTERING methods are used to partition a data set into several subsets so that the sample points in the same subsets are the most similar to each other and the sample points in the different subsets are the most dissimilar. Nowadays, clustering algorithms have been widely and successfully applied in a variety of substantive areas, such as image processing, data mining, pattern recognition, machine learning, etc. In general, clustering methods are based on an objective function of similarity or dissimilarity measures in which partitional methods are popularly used [1]. The most popular partitional methods with cluster prototypes are k-means [2]-[3], fuzzy c-means (FCM) [4]-[5] and possibilistic c-means (PCM) [6]-[7].

In general, clustering algorithms treat feature components as equal weights. To improve their clustering strengths, there are many researches in considering feature weighting extensions of clustering methods, such as Modha and Spangler [8], Huang et al. [9], Wang et al. [10] and Hung et al. [11]. However, most clustering methods, even with those feature weighting extensions, are always considering all sample points as equal weights during clustering processes. In practice, it is not good to suppose that every sample in a data set has the same weight in cluster analysis. Recently, Yu et al. [12] had considered a probability distribution over a data set to represent its sample weights and then proposed sample-weighted clustering algorithms.

In this paper, we first consider fuzzy clustering with regularizations. We then give a sample-weighted version of these fuzzy clustering regularization methods, called the

sample-weighted generalized fuzzy clustering with regularizations (SW-GFCR). Some examples are considered for comparisons. These experimental results and comparisons actually demonstrate that the proposed SW-GFCR is more effective than the clustering algorithms. The rest of this paper is organized as follows. In Section II, we first considered a probability distribution over a data set to represent its sample weights and also review the generalized fuzzy clustering with regularizations (GFCR). We then propose the sample-weighted generalized fuzzy clustering with regularizations (SW-GFCR). In Section III, the example is made with real data sets to demonstrate the effectiveness and usefulness of the proposed algorithms. Finally, conclusions are stated in Section IV.

II. SAMPLE-WEIGHTED FUZZY CLUSTERING WITH REGULARIZATIONS

Let $X = \{x_1, x_2, \dots, x_n\}$ be an s -dimensional data set. Suppose that $u = [u_{ik}]_{c \times n} \in M_{fcm}$

$= \left\{ u = [u_{ik}]_{c \times n} \mid \sum_{i=1}^c u_{ik} = 1, u_{ik} \geq 0, 0 < \sum_{i=1}^c u_{ik} < n \right\}$ is a partition

matrix and $v = \{v_1, \dots, v_c\}$ is the set of cluster centers. The fuzzy c-means (FCM) objective function [4]-[5] is defined as

$J_m(u, v) = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m \|x_k - v_i\|^2$. By Lagrange multiplier, the

necessary conditions for the minimum of $J_m(u, v)$ with respect to a partition u_{ik} and a cluster prototype v_i are the following update equations:

$$v_i = \frac{\sum_{k=1}^n u_{ik}^m x_k}{\sum_{k=1}^n u_{ik}^m} \text{ and } u_{ik} = \left(\frac{\|x_k - v_i\|^2}{\sum_{j=1}^c (\|x_k - v_j\|^2)^{\frac{1}{m-1}}} \right)^{-1}.$$

The FCM algorithm is iterated with the above update equations where it becomes the most used clustering algorithm in the literature.

There are many generalizations of FCM. In Yu and Yang [13], they proposed a generalized model, called a generalized fuzzy clustering regularization (GFCR) method with the constraint on membership functions u_{ik} . The GFCR objective function was defined as follows [13]:

Miin-Shen Yang is with the Department of Applied Mathematics, Chung Yuan Christian University, Chung-Li 32023, Taiwan. (e-mail: msyang@math.cycu.edu.tw)

$$J_m^\lambda(u, v) = \sum_{k=1}^n \sum_{i=1}^c [u_{ik}^m d_{ik} + \lambda p(u_{ik})] \quad (1)$$

where $p(u_{ik})$ is a function of u_{ik} . By the Lagrange multiplier δ_k , the necessary conditions of minimization of equation (1) are derived as follows:

$$m u_{ik}^{m-1} d_{ik} + \lambda p'(u_{ik}) + \delta_k = 0 \quad (2)$$

$$\sum_{k=1}^c u_{ik}^m (x_k - v_i) = 0 \quad (3)$$

We see that the update (3) of the cluster centers $\{v_1, \dots, v_c\}$ for GFCR is the same as that for the FCM. The iterations with update (2) and (3) are called the GFCR algorithm. We mention that Wei and Fahn [14] considered the function $p(u_{ik})$ with $p(u_{ik}) = \sum_{k=1}^n \sum_{i=1}^c u_{ik} \log u_{ik}$. Yang [15] and Özdemir and Akarun [16] considered the function $p(u_{ik})$ with $p(u_{ik}) = \sum_{k=1}^n \sum_{i=1}^c u_{ik} \log u_{ik}^m$. Yasuda et al. [17] considered the function $p(u_{ik})$ with $p(u_{ik}) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik} \log u_{ik} + (1 - u_{ik}) \log(1 - u_{ik}))$.

On the other hand, Yu et al. [12] considered a probability distribution over a data set to represent its sample weights and then proposed sample-weighted clustering algorithms. We review it as follows. If the distortion measure between a sample point x_k and the cluster centers $\{v_1, \dots, v_c\}$ is denoted as d_k and a distribution over X is denoted as $p(x)$, then the expected distortion can be defined as

$$D = \sum_{k=1}^n p(x_k) d_k, \quad (4)$$

where for each k , $p(x_k) \geq 0$ and $\sum_{k=1}^n p(x_k) = 1$. In cluster analysis, the best partitioning results should minimize the objective function D . However, the direct minimization of equation (4) with respect to a distribution $p(x)$ and the cluster centers $\{v_1, \dots, v_c\}$ may produce unreasonable clustering results. This is because, for $d_k^* = \min_{1 \leq k \leq n} d_k$, we have $D = \sum_{k=1}^n p(x_k) d_k \geq d_k^*$. Therefore, we obtain a unique minimum for equation (4) with the sample weight $p^0(x_j) = 1$

if $j = \arg \min_{1 \leq i \leq n} d_i$ and 0 elsewhere. However, this weighting function for a data set only produces a cluster with a data point, which obviously makes no sense for clustering. This kind of result is produced because there is no prior knowledge about the distribution $p(x)$. In this sense, we may apply the maximum entropy principle so that Yu et al. [12] proposed the following objective function:

$$D_E = \sum_{k=1}^n p(x_k) d_k + \zeta^{-1} \sum_{k=1}^n p(x_k) \ln p(x_k) \quad (5)$$

where $\forall k, p(x_k) \geq 0$ and $\sum_{k=1}^n p(x_k) = 1$ and $\zeta > 0$. Thus, by the Lagrange multiplier, the following update equation for $p(x)$ can be obtained:

$$p(x_k) = \frac{\exp(-\zeta \times d_k)}{\sum_{k=1}^n \exp(-\zeta \times d_k)} \quad (6)$$

Next, we propose our sample-weighted generalized fuzzy clustering with regularizations (SW-GFCR). To have the GFCR with sample weights, we consider the SW-GFCR objective function as follows:

$$F(w, u, v) = \sum_{k=1}^n \sum_{i=1}^c w_k (u_{ik}^m d_{ik} + \lambda p(u_{ik})) + \zeta^{-1} \sum_{k=1}^n w_k \log w_k$$

where $\forall i, \forall k, u_{ik} \geq 0, \sum_{i=1}^c u_{ik} = 1, w_k \geq 0, \sum_{k=1}^n w_k = 1$ and $\zeta > 0, 2 \leq c < n, m > 0$. Thus, we have the Lagrangian as follows:

$$F(w, u, v, \lambda) = \sum_{k=1}^n \sum_{i=1}^c w_k (u_{ik}^m d_{ik} + \lambda p(u_{ik})) + \zeta^{-1} \sum_{k=1}^n w_k \log w_k + \delta_1 \left(\sum_{k=1}^n w_k - 1 \right) + \delta_2 \left(\sum_{i=1}^c u_{ik} - 1 \right)$$

Let $\frac{\partial F}{\partial v_i} = 0$. We obtain $\sum_{k=1}^n w_k u_{ik}^m (x_k - v_i) = 0$. Thus, we have that

$$v_i = \frac{\sum_{k=1}^n w_k u_{ik}^m x_k}{\sum_{k=1}^n w_k u_{ik}^m}. \quad (7)$$

Let $\frac{\partial F}{\partial w_k} = 0$. We obtain $w_k = e^{-1} e^{-\zeta \delta_1} e^{-\zeta \sum_{i=1}^c (u_{ik}^m d_{ik} + \lambda p(u_{ik}))}$. Since

$$\sum_{k=1}^n w_k = 1, \text{ we have that } e^{-\zeta \delta_1} = \frac{e}{\sum_{k=1}^n e^{-\zeta \sum_{i=1}^c (u_{ik}^m d_{ik} + \lambda p(u_{ik}))}}. \text{ Thus, we}$$

can obtain

$$w_k = \frac{e^{-\zeta \sum_{i=1}^c (u_{ik}^m d_{ik} + \lambda p(u_{ik}))}}{\sum_{q=1}^n e^{-\zeta \sum_{i=1}^c (u_{iq}^m d_{iq} + \lambda p(u_{iq}))}} \quad (8)$$

Furthermore, we consider $\frac{\partial F}{\partial u_{ik}} = 0$. We have that

$$m u_{ik}^{m-1} w_k d_{ik} + \lambda w_k p'(u_{ik}) + \delta_2 = 0 \quad (9)$$

In this paper, we consider the following three cases:

(I) $p(u_{ik}) = u_{ik} \log u_{ik}$; We have that $p'(u_{ik}) = \log u_{ik} + 1$ and

$$u_{ik} = \exp\left(\frac{-m(u_{ik})^{m-1} w_k d_{ik} - \lambda w_k - \delta_2}{\lambda w_k}\right). \text{ Thus, we obtain}$$

$$u_{ik} = \frac{\exp(-m\lambda^{-1} u_{ik}^{m-1} d_{ik} - 1)}{\sum_{j=1}^c \exp(-m\lambda^{-1} u_{jk}^{m-1} d_{jk} - 1)} \quad (10)$$

(II) $p(u_{ik}) = u_{ik}^m$; We have that $p'(u_{ik}) = m(u_{ik})^{m-1}$ and

$$u_{ik} = \left(\frac{-\delta_2}{(w_k d_{ik} + \lambda w_k) m}\right)^{\frac{1}{m-1}}. \text{ Thus, we obtain}$$

$$u_{ik} = \frac{(m(w_k d_{ik} + \lambda w_k))^{\frac{-1}{m-1}}}{\sum_{j=1}^c (m(w_k d_{jk} + \lambda w_k))^{\frac{-1}{m-1}}} \quad (11)$$

(III) $p(u_{ik}) = u_{ik} \log u_{ik} + (1-u_{ik}) \log(1-u_{ik})$; We have that $p'(u_{ik}) = \log u_{ik} - \log(1-u_{ik})$ and

$$u_{ik} = \exp(-m\lambda^{-1} u_{ik}^{m-1} d_{ik} + \log(1-u_{ik}) - (\delta_2/\lambda w_k)).$$

Thus, we obtain

$$u_{ik} = \frac{\exp(-m\lambda^{-1} u_{ik}^{m-1} d_{ik} + \log(1-u_{ik}))}{\sum_{j=1}^c \exp(-m\lambda^{-1} u_{jk}^{m-1} d_{jk} + \log(1-u_{jk}))} \quad (12)$$

For learning the parameters ζ and λ , we consider to use $\zeta^{(l+1)} = 0.999^{l+1}$ and $\lambda^{(l+1)} = 0.999^{l+1}$. Thus, we can construct a new sample-weighted generalized fuzzy clustering with regularizations (SW-GFCR) as follows:

SW-GFCR algorithm

Step 1: Fix $2 \leq c \leq n$ and fix any $\varepsilon > 0$. Give initials

$$u^{(0)} = [u_{ik}^{(0)}]_{c \times n}$$

and $w^{(0)} = \{w_1^{(0)}, \dots, w_k^{(0)}\}$ and let $t = 0$.

Step 2: Learning the parameters ζ and λ by

$$\zeta^{(t+1)} = 0.999^{t+1} \text{ and } \lambda^{(t+1)} = 0.999^{t+1}.$$

Step 3: Compute the cluster center $v^{(t+1)}$ with $u^{(t)}$ using equation (7).

Step 4: Compute the probability weight $w_i^{(t+1)}$ by equation (8);

Step 5: Update $u^{(t+1)}$ with $v^{(t+1)}$ using equation (10) or (11) or (12).

Step 4: Compare $u^{(t+1)}$ to $u^{(t)}$ in a convenient matrix norm $\|\cdot\|$.

$$\text{IF } \|u^{(t+1)} - u^{(t)}\| < \varepsilon, \text{ STOP}$$

ELSE $t = t + 1$ and return to step 2.

III. EXAMPLE AND COMPARISONS

In this section, we compare FCM, sample-weighted FCM (SW-FCM), sample-weighted k-means (SW-KM) with our SW-GFCR(I), SW-GFCR(II) and SW-GFCR(III) in the cases (I), (II) and (III), respectively, for the following two real data sets.

Data 1: The IRIS data set [18] that has 150 data points with each four attributes. It is divided into three clusters of Iris Setosa, Iris Versicolor and Iris Virginica, and two of them are overlapping where each cluster has 50 data points.

Data 2: The liver disorders data set has 290 data points with each seven attributes [19]. The first six components are continuous-type attributes as follows: mean corpuscular volume, alkaline phosphatase, alamine aminotransferase, aspartate aminotransferase, gamma-glutamyl transpeptidase and number of half-pint equivalents of alcoholic beverages drunk per day. The last component is a categorical attribute of selector that is a field used to split data into two sets.

We first implement SW-KM, SW-FCM, SW-GFCR(I), SW-GFCR(II) and SW-GFCR(III) for the IRIS data set with different values of the parameter ζ under the learning way $\lambda^{(t+1)} = 0.999^{t+1}$. We consider average error counts of different algorithms with each 100 runs. The results are shown in Table I. We find that, for all different algorithms, if the value of ζ is larger than 1.0, then the average error counts will become larger. We also find that, the parameter λ has the similar situation as the parameter ζ . In this sense, we consider the learning way for the parameters ζ and λ with $\zeta^{(t+1)} = 0.999^{t+1}$ and $\lambda^{(t+1)} = 0.999^{t+1}$ during implementing SW-FCM, GFCR and SW-GFCR. Furthermore, from Table I, we also find that SW-GFCR(I) presents the best for the IRIS data set among SW-GFCR(I), SW-GFCR(II) and SW-GFCR(III). In this sense, we pick the case (I) of $p(u_{ik}) = u_{ik} \log u_{ik}$ for regularization. In Table II, we compare the average error counts of SW-GFCR(I) with FCM, SW-FCM and GFCR(I) under

$\zeta^{(t+1)} = 0.999^{t+1}$ and $\lambda^{(t+1)} = 0.999^{t+1}$ for the IRIS data set with 100 runs. We can see that SW-GFCR(I) is the best one among them. Moreover, we consider FCM, GFCR(I) and SW-GFCR(I) under $\zeta^{(t+1)} = 0.999^{t+1}$ and $\lambda^{(t+1)} = 0.999^{t+1}$ for the liver disorders data set. These average error counts are shown in Table III. We find that SW-GFCR(I) also gives the best result for the liver disorders data set.

TABLE I
AVERAGE ERROR COUNTS OF ALGORITHMS WITH $\lambda^{(t+1)} = 0.999^{t+1}$ FOR THE IRIS DATA SET (100 RUNS)

	$\zeta = 0.004$	$\zeta = 0.26$	$\zeta = 0.48$	$\zeta = 0.89$	$\zeta = 1.62$	$\zeta = 5.50$	$\zeta = 10.00$
SW-K M	36.71	37.58	41.69	41.91	52.88	54	34.88
SW-FC M	16	15	15	41	42	44	56.27
SW-GF CR(I)	13.83	13.83	13.83	13.93	14.15	15.13	22.41
SW-GF CR(II)	24.68	24.68	24.68	24.68	27.15	32.08	33.32
SW-GF CR(III)	27.37	27.37	29.37	29.30	29.45	35.45	40.33

TABLE II
AVERAGE ERROR COUNTS FOR THE IRIS DATA SET

FCM	SW-FCM	GFCR(I)	SW-GFCR(I)
16	15.2	15.93	13.83

TABLE III
AVERAGE ERROR COUNTS FOR THE LIVER DISORDERS DATA SET

FCM	GFCR(I)	SW-GFCR(I)
27.7%	26.7%	25.5%

IV. CONCLUSIONS

In this paper, we propose the sample-weighted generalized fuzzy clustering with regularizations (SW-GFCR). The proposed method can obtain probability weights over the data points with membership regularizations. It is not only to consider the sample weights, but also to adjust the bias with memberships. The experiments and comparisons demonstrate the superiority and effectiveness of the proposed SW-GFCR.

REFERENCES

[1] L. Kaufman, P.J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York, 1990.
 [2] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proc. 5th Berkeley Symp. Math. Statist. Prob. 1, 1967, pp. 281-297.
 [3] D. Pollard, "Quantization and the method of k-means," *IEEE Trans.*

Information Theory, vol. 28, pp. 199-205, 1982.
 [4] J.C. Bezdek, J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.
 [5] M.S. Yang, "A survey of fuzzy clustering," *Mathematical and Computer Modeling*, vol. 18, pp. 1-16, 1993.
 [6] R. Krishnapuram, J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Systems*, vol. 1, pp. 98-110, 1993.
 [7] M.S. Yang and C.Y. Lai, "A robust automatic merging possibilistic clustering method," *IEEE Trans. on Fuzzy Systems*, vol. 19, pp. 26-41, 2011.
 [8] D.S. Modha and W.S. Spangler, "Feature weighting in k-means clustering," *Machine Learning*, vol. 52, pp. 217-237, 2003.
 [9] J.Z. Huang, M.K. Ng, H. Rong, and Z. Li, "Automated variable weighting in k-means type clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 657-668, 2005.
 [10] X.Z. Wang, Y.D. Wang, L.J. Wang, "Improving fuzzy c-means clustering based on feature-weight learning," *Pattern Recognition Letter*, vol. 25, pp. 1123-1132, 2004.
 [11] W.L. Hung, M.S. Yang and D.H. Chen, "Bootstrapping approach to feature-weight selection in fuzzy c-means algorithms with an application in color image segmentation," *Pattern Recognition Letters*, vol. 29, pp. 1317-1325, 2008.
 [12] J. Yu, M. S. Yang, E. S. Lee, "Sample-weighted clustering methods," *Computers and Mathematics with Applications*, vol. 62 (2011) 2200-2208.
 [13] J. Yu, M. S. Yang, "A generalized fuzzy clustering regularization model with optimality tests and model complexity analysis," *IEEE Transactions on Fuzzy Systems*, vol. 15, pp. 904-915, 2007.
 [14] C. Wei, C. Fahn, "The multisynapse neural network and its application to fuzzy clustering," *IEEE Transactions on Neural Networks*, vol. 13, pp. 600-618, 2002.
 [15] M. S. Yang, "On a class of fuzzy classification maximum likelihood procedures," *Fuzzy Sets Systems*, vol. 57, pp. 365-375, 1993.
 [16] D. Özdemir, L. Akarun, "A fuzzy algorithm for color quantization of images," *Pattern Recognition*, vol. 35, pp. 1785-1791, 2002.
 [17] M. Yasuda, T. Furuhashi, M. Matsuzaki, S. Okuma, "Fuzzy clustering using deterministic annealing method and its statistical mechanical characteristics," in Proc. 10th IEEE International Conference on Fuzzy Systems, 2001, pp. 797-800.
 [18] E. Anderson, "The IRISes of the Gaspé Peninsula," *Bull. Am. IRIS Soc.*, vol. 59, pp. 2-5, 1935.
 [19] UCI machine learning repository, Center for Machine Learning and Intelligent Systems, Liver Disorders Data Set. Available from: <http://archive.ics.uci.edu/ml/datasets/Liver+Disorders>