Gradient Based Fuzzy c_means (GBFCM) Algorithm

Dong C. Park and Issam Dagher
Intelligent Computing Research Lab
Department of Electrical and Computer Engineering
Florida International University
Miami,Florida 33199
e-mail: parkd@cherry.fiu.edu

Abstract

In this paper, a clustering algorithm based on the Fuzzy c_means algorithm (FCM) and the gradient descent method is presented. In the FCM, The minimization process of the objective function is proceeded by solving two equations alternatively in an iterative fashion. Each iteration requires the use of all the data at once. In our proposed approach one datum at a time is presented to the network, and the minimization is proceeded using the gradient descent method. Compared to FCM, the experimental results show that our algorithm is very competitive in terms of speed and stability of convergence for large number of data.

1 Introduction

The objective of clustering algorithms is the grouping of similar objects and separating of dissimilar ones [1]. Kohonen [2] introduced a network where the learning is based on the Winner_Take_All rule. It is based on the premise that one of the neurons has the maximum response due to an input. Only one winner is declared (membership of 1) and all others are loosers (memberships of 0).

In 1973, Dunn [3] introduced a generalization of the Winner_Take_All rule. He combines Zadeh's set concept with the criterion approach to clustering. The membership grade is spread over $\{0, \cdots, 1\}$ which will be able to "signal the presence or absence of Well-Separated Clusters". He derived the necessary conditions for minimizing an objective function. He called this method *fuzzy ISODATA*. A generalization of the *fuzzy ISODATA* is done by Bezdek [4]. He defined a family of objective functions $\{J_m,\ 1 < m < \infty\}$, and established a convergence theorem for that family. Windham presented a cluster validity for the FCM algorithm[5]. He obtained a measure by computing the ratio of the smallest membership to the largest one and transforming this ratio into a probability function. In [6], an unsupervised fuzzy clustering algorithm had been presented. The algorithm located the regions where the data have low degree of memberships, and tried to augment the number of clusters according to these regions.

In this paper, we present a new algorithm (GBFCM) which combines the characteristics of Kohonen network (presenting one data at a time and applying the gradient descent method) and the FCM algorithm (continuous values of the membership grades in the range $\{0, \cdots, 1\}$). In section 2, the minimization procedures for both FCM algorithm and Kohonen network are presented. A derivation of the proposed algorithm is presented in section 3 establishing a comparison between FCM and GBFCM and showing the pseudocode of the algorithm. Section 4 contains the results of applying FCM and GBFCM on the same problems. We conclude in section 5 with a short discussion about the algorithm and how it could be improved.

0-7803-1901-X/94 $4.00 ©1994 IEEE

2 Minimization procedures

The minimization procedure for Kohonen network is based on the gradient descent method which tries to minimize the distance $\|x_j - w_{ij}\|$ of an input x_j to the associated weight w_{ij}. This can be achieved by moving the weights along the negative gradient direction such that:

$$-\frac{\partial(x_j - w_{ij})^2}{\partial w_{ij}} = 2(x_j - w_{ij}) \tag{1}$$

The weights which are connected to the winning neuron will be updated using the following formula [1]:

$$\Delta w_{ij} = c(n)(x_j - w_{ij}) \tag{2}$$

where $c(n)$ is a small learning constant, and is typically decreasing as learning proceeds.

For FCM, the objective function to be minimized is defined as [7]:

$$J_m(U, v) = \sum_{k=1}^{n} \sum_{i=1}^{c} \{(\mu_i(x_k))^m (d_i(x_k))^2\} \tag{3}$$

where $d_i(x_k)^2$ is the distance from the input data x_k to v_i, v_i is the center of cluster i, $\mu_i(xk)$ is the membership value of xk in the cluster i, m is a weighting exponent. $m \in \{1, \cdots, \infty\}$, c is the number of clusters,and n is the number of data.
Bezdek found the following conditions for minimizing $J_m(U, v)$[2-5]:

$$\mu_i(x_k) = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_i(x_k)}{d_j(x_k)}\right)^{\frac{2}{m-1}}} \tag{4}$$

$$v_i = \frac{\sum_{k=1}^{n} (\mu_i(x_k))^m x_k}{\sum_{k=1}^{n} (\mu_i(x_k))^m} \tag{5}$$

The fuzzy clustering is achieved by applying equations 4 and 5 alternatively, in an iterative fashion until the error reaches a small value.

3 Gradient-Based FCM (GBFCM)

One attempt to improve the FCM algorithm in this paper is made by trying to minimize the objective function using one data at a time instead of the whole data.

3.1 Derivation of the algorithm

Given one datum x_i and c clusters with centers at v_j $(j = 1, 2, \cdots, c)$, the objective function to be minimized is :

$$J_i = \mu_{1i}^2 (v_1 - x_i)^2 + \mu_{2i}^2 (v_2 - x_i)^2 + \cdots + \mu_{ci}^2 (v_c - x_i)^2 \tag{6}$$

with the constraint:

$$\mu_{1i} + \mu_{2i} + .. + \mu_{ci} = 1 \tag{7}$$

The basic procedure of the gradient descent method is that starting from an initial center vector v, the gradient ΔJ_i of the current objective function is computed [8]. The next value of v is obtained by moving in the direction of the negative gradient along the multidimensional error surface such that:

$$v_{k+1} = v_k - \eta \frac{\partial J_i}{\partial v_k} \tag{8}$$

where :

$$\frac{\partial J_i}{\partial v_k} = 2\mu_{ki}^2 (v_k - x_i) \tag{9}$$

Equivalently,

$$v_{k+1} = v_k - \eta \mu_{ki}^2 (v_k - x_i) \tag{10}$$

where η is a small learning constant.

For the membership grades, we set :

$$\frac{\partial J_i}{\partial \mu} = 0 \tag{11}$$

and obtain :

$$\mu_i(x_k) = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_i(x_k)}{d_j(x_k)}\right)^2} \tag{12}$$

3.2 Comparison between FCM and GBFCM

FCM and GBFCM both have an objective function which tries to minimize the distance between each center and the data with a membership grade reflecting the degree of their similarities with respect to other centers. On the other hand, they differ in the way they try to minimize it:

- In the FCM algorithm, all the data are present in the objective function, and the gradients are set to zero in order to obtain the equations necessary for minimization[2-5] .
- In the GBFCM, only one datum is present at a time. And only the gradients of the objective function with respect to the membership grades are set to zero. The gradients with respect to the centers are not set to zero. their negative values are used to minimize the objective function.

3.3 GBFCM algorithm

The algorithm for implementing this method is as follows :

Procedure main()
  Read c, ϵ, m
    While ($error > \epsilon$)
      $e \leftarrow 0$
      While (input file is not empty)
        Read one datum x_i
        $v_{k+1} \leftarrow v_k - \eta \mu_{ki}^2 (v_k - x_i)$
        $\mu_i(x_k) \leftarrow \frac{1}{\sum_{j=1}^{c} \left(\frac{d_i(xk)}{d_j(xk)}\right)^2}$

$e \leftarrow e + v_i(k+1) - v_i(k)$
end while
error $\leftarrow e$
end while
Output μ_i and v_i.
end main()

4 Experiments and Results

Two experiments have been implemented using a Guassian function generator which creates an input file contained 1000 data.
we set $m = 2$ and $\epsilon = .001$, and to improve the speed of convergence of the algorithm, the following two steps are added:

- after one epoch calculate the difference of errors $dif = e(k+1) - e(k)$
- if that difference is very small decrease the step η.

In our algorithm, we used $dif = .0001$ and the decreasing step is $\frac{\eta}{16}$.
These two steps guaranteed a speed of convergence comparable to both Kohonen network and FCM algorithm. And for a large number of input data, GBFCM converges faster.

4.1 Experiment 1

The input data in figure 1(a)has been used for testing:

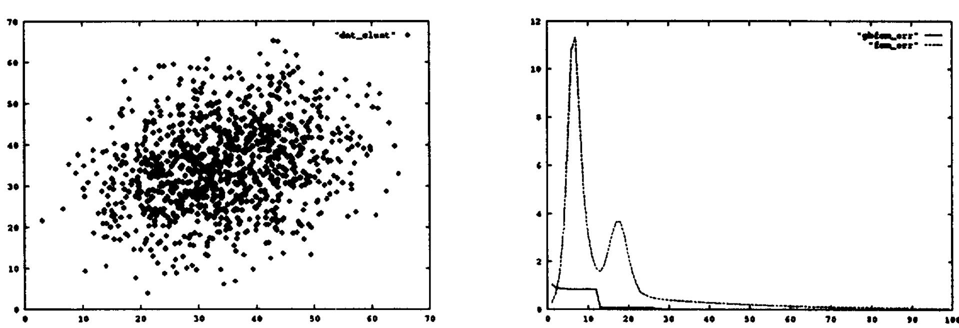

Figure 1: (a)input data (b)Error function vs. iteration for 4 clusters (solid:GBFCM,dotted:FCM)

GBFCM converged much faster than FCM for $\eta = .1$. The behavior of the error with respect to each iteration for both algorithms is shown in figure 1(b). In the following table, we show the number of iterations needed for both algorithms in achieving $\epsilon = .001$ for the cases of $c = 2, 3, 4$ clusters.

n.of clusters	2	3	4
n. of it. for FCM	14	60	60
n. of it. for GBFCM	12	26	26

4.2 Experiment 2

We used another type of input data ,figure 2(a). And we had the following table:

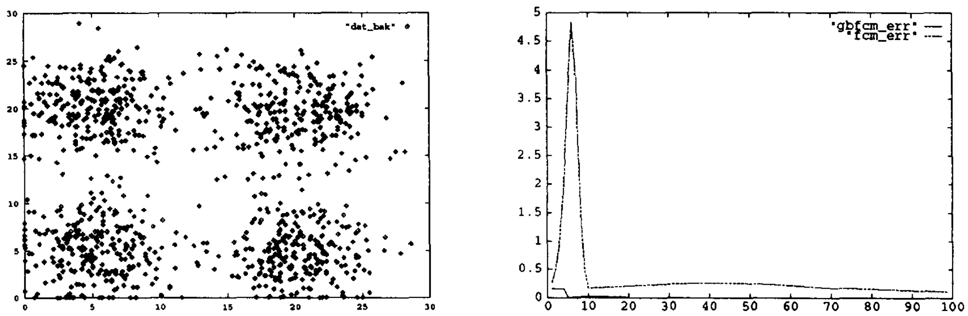

Figure 2: (a)input data, (b)Error function vs. iteration for 2 clusters (solid:GBFCM,dotted:FCM)

n.of clusters	2	3	4
n. of it. for FCM	40	40	12
n. of it. for GBFCM	20	20	8

4.3 Discussions

FCM and GBFCM approximately required the same number of iterations when the number of clusters was 2 for experiment 1 and 4 for experiment 2. This is due to the way of defining the number of groups in the Gaussian function generator. We defined 2 groups for experiment 1 and 4 groups for experiment 2. Then we used different means and different standard deviations. We can conclude that if the number of the specified clusters is the same as the natural number of clusters, both algorithms converged quickly with similar number of iterations. However, if the specified number of clusters is different from the natural number of clusters,as in the case of practical situations, GBFCM converges faster than FCM.

5 Conclusions

A Gradient-Based FCM (GBFCM) algorithm is presented. The proposed GBFCM requires only one datum at a time unlike the FCM algorithm which requires the presence of the whole data at each iteration. The comparative study shows that GBFCM converges faster than FCM in practical applications.

Acknowledgements

The research described in this paper was supported in part by National Science Foundation grant CDA-9313624. We thank ICRL members at FIU for fruitful discussions.

References

[1] J.M.Zurada, *Artificial neural systems*, West 1992

[2] T. Kohonen," Learning Vector Quantization," Helsinki University of Technology, Laboratory of Computer and Information Science, Report TKK-F-A-601, 1986

[3] J.C.Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters," *J. Cybern.*, 1973,(3):32-75.

[4] J.C.Bezdek, " A convergence theorem fo the fuzzy ISODATA clustering algorithms,"*IEEE trans. pattern Anal. Mach. int.*, 1980,(2):1-8. 1975,(24):835-838.

[5] Windham M.P.,"Cluster Validity for the Fuzzy c_means clustering algorithm,"*IEEE trans. pattern Anal. Mach. int.*, 1982,(4):357-363.

[6] I.Gath,A.B.Geva, "Unsupervised optimal fuzzy clustering," *IEEE trans. pattern Anal. Mach. int.*, 1989,(11):773-781.

[7] J.C.Bezdek, *Pattern recognition with fuzzy objective function algorithms,* New York: Plenum,1981.

[8] B.S.Gottfried,J.weisman, *Introduction to optimization theory*, Prentice Hall 1973

[9] J.C.Bezdek,J.C.Dunn," Optimal fuzzy partition:A heuristic for estimating the parameters in a mixture of normal distributions," *IEEE Trans. Comput.*, 1975,(24):835-838.

[10] Xuanli l.x.,BEN G., " A validity measure for fuzzy clustering," *IEEE trans. pattern Anal. Mach. int.*, 1991,(13):841-847.

[11] J. Hertz et al., *Introduction to the Theory of Neural Computation*, Reading, MA: Addison-Wesley, 1990

[12] G.J.Klir,T.A.Folger, *Fuzzy sets, uncertainty, and information*, Prentice Hall 1988

[13] V.Tiel, *Convex Analysis*, Wiley 1984