Generalised Fuzzy Clustering

Dat Tran and Dharmendra Sharma
School of Information Sciences and Engineering
University of Canberra, ACT 2601, Australia

Abstract-- This paper presents a general approach to fuzzy clustering methods. A generalised fuzzy objective function is used to combine fuzzy c-means clustering, fuzzy entropy clustering, and their extended versions into a generalised fuzzy clustering method. Some new extended versions of the above-mentioned clustering methods are proposed from this general approach. Several cluster data sets were analysed and experimental results showed that noise clustering-based fuzzy entropy clustering gives good classification rates on most of the cluster data sets.

1. INTRODUCTION

Cluster analysis is one of the main issues of pattern recognition to search for structure in data sets [1]. Clustering is the grouping similar feature vectors into clusters. Clusters can have sharp boundaries in a hard clustering method, where each vector is in only one cluster [3]. In contrast, boundaries generated by a fuzzy clustering method are vague. Each vector of a fuzzy partition belongs to different clusters with different fuzzy membership values [2]. The well-known fuzzy method is fuzzy c-means (FCM) clustering [4]. Its extended versions are Gustafson-Kessel clustering [6], Gath-Geva [7] clustering, and noise clustering [5]. Fuzzy entropy (FE) clustering and its extended versions were proposed in our previous work [8]. In general, the objective functions and algorithms of hard, FCM and FE clustering methods have some common characteristics, therefore we propose a general approach to combine these clustering methods. From this general approach, we can derive not only the above-mentioned clustering methods by setting particular values for fuzzy parameters, but also new extended versions of FE clustering, such as Gaussian-based FE clustering and noise clustering-based FE clustering. Several cluster data sets were analysed in this paper by those clustering methods. These data sets were designed by Wagner & Wagner [12] at University of Braunschweig. Experimental results showed that noise clustering-based fuzzy entropy clustering gives highest classification rates on most of the cluster data sets.

2. GENERAL CLUSTERING METHOD

Let \( X = \{x_1, x_2, \ldots, x_T \} \) be a set of feature vectors in a multi-dimensional space, the general objective function is written as follows

\[
G_{mn}(U, \Theta; X) = \sum_{i=1}^{C} \sum_{t=1}^{T} u_{it}^m d^2(x_i, \theta_t) + n \delta_{mn} \sum_{i=1}^{C} \sum_{t=1}^{T} u_{it} \log u_{it}
\]

(1)

where \( n \geq 1 \) is degree of fuzziness; \( m \geq 0 \) degree of fuzzy entropy; \( U = [u_{it}] \),
1 ≤ i ≤ C, 1 ≤ t ≤ T, a matrix whose elements \( u_{it} \) are memberships of \( x_t \) in class \( C_i \); 
\( \theta = \{ \theta_1, \theta_2, \ldots, \theta_C \} \), a set of \( C \) prototypes, each of which characterises one of the \( C \) clusters, each prototype is a parameter set; \( d(x_t, \theta_i) \) is the distance from \( x_t \) to prototype \( \theta_i \), known as a measure of dissimilarity; and \( \delta_{m1} \) is the Kronecker delta function \( (\delta_{m1} = 1 \text{ if } m = 1, \text{ and } \delta_{m1} = 0 \text{ otherwise}) \). Let \( \mu_i, \Sigma_i, \) and \( w_i, 1 \leq i \leq C \) are mean vectors, covariance matrices and mixture weights, denoting locations, shapes and densities of clusters. By setting particular values for \( m \) and \( n \), we can derive various clustering methods.

3. HARD CLUSTERING

Let \( m = 1, n = 0 \), the general objective function becomes the \( K \)-means objective function as follows [3]

\[
G_{10}(U, \theta; X) = \sum_{i=1}^{C} \sum_{t=1}^{T} u_{it} d^2(x_t, \theta_i) \quad (2)
\]

where \( \theta_i = [\mu_i] \) and 
\[
d^2(x_t, \theta_i) = \| x_t - \mu_i \|^2 = (x_t - \mu_i)'(x_t - \mu_i) \quad (3)
\]

The \( K \)-means parameter estimation equations are as follows

\[
u_{it} = \begin{cases} 
1 & \text{if } d(x_t, \theta_i) < d(x_t, \theta_k) \quad \forall k \neq i \\
0 & \text{otherwise}
\end{cases}, \quad \mu_i = \frac{\sum_{t=1}^{T} u_{it} x_t}{\sum_{t=1}^{T} u_{it}} \quad (4)
\]

4. FUZZY C-MEANS CLUSTERING

Let \( m > 1 \), note that \( \delta_{m1} = 0 \), the general objective function becomes the fuzzy c-means objective function as follows [4]

\[
G_m(U, \theta; X) = \sum_{i=1}^{C} \sum_{t=1}^{T} u_{it}^m d^2(x_t, \theta_i) \quad (5)
\]

where \( \theta_i \) and \( d^2(x_t, \theta_i) \) are defined as those in (3). The FCM parameter estimation equations are as follows

\[
u_{it} = \left\{ \sum_{i=1}^{C} \left[ \frac{d(x_t, \theta_i) / d(x_t, \theta_k)}{d(x_t, \theta_i) / d(x_t, \theta_k)} \right]^{\frac{2}{m-1}} \right\}^{-1} \quad \text{and} \quad \mu_i = \frac{\sum_{t=1}^{T} u_{it}^m x_t}{\sum_{t=1}^{T} u_{it}^m} \quad (6)
\]

5. GUSTAFSON-KESSEL CLUSTERING

Using the FCM objective function in (5) and redefine \( \theta_i \) and \( d^2(x_t, \theta_i) \) as follows

\[
\theta_i = [\mu_i, \Sigma_i], \quad d^2(x_t, \theta_i) = (x_t - \mu_i)' M_i^{-1} (x_t - \mu_i) \quad (7)
\]

where matrices \( M_i, 1 \leq i \leq C \), are symmetric and positive definite and subject to the following constraints \( |M_i| = \rho_i \), with \( \rho_i > 0 \) and fixed for each \( i \). We obtain 
\( M_i^{-1} = (|M_i| |\Sigma_i|)^{-\frac{1}{2}} \Sigma_i, 1 \leq i \leq C \), where \( |M_i| \) and \( |\Sigma_i| \) are the determinants of \( M_i \),
and $\Sigma$, respectively and $L$ is the vector space dimension. The Gustafson-Kessel parameter estimation equations are as follows [6]

$$u_{il} = \left\{ \sum_{k=1}^{C} \left[ d(x_i, \theta_l)/d(x_i, \theta_k) \right]^{\frac{2}{m-1}} \right\}^{-1}$$  \hspace{1cm} (8)

$$\mu_i = \frac{T}{T} \sum_{l=1}^{T} u_{il}^m x_i / \sum_{l=1}^{T} u_{il}^m, \quad \Sigma_i = \frac{T}{T} \sum_{l=1}^{T} u_{il}^m (x_i - \mu_i) (x_i - \mu_i) / \sum_{l=1}^{T} u_{il}^m$$  \hspace{1cm} (9)

6. **GAUSSIAN CLUSTERING**

Let $m = 1$ and $n = 1$, the general objective function in (1) is rewritten as follows

$$G(U, \theta; X) = \sum_{i=1}^{C} \sum_{l=1}^{T} u_{il} d^2(x_i, \theta_l) + \sum_{i=1}^{C} \sum_{l=1}^{T} u_{il} \log u_{il}$$  \hspace{1cm} (10)

Redefining $\theta_l$ and $d^2(x_i, \theta_l)$ as follows

$$\theta_l = (\mu_l, \Sigma_l, w_l), \quad d^2(x_i, \theta_l) = -\log P(x_i, \theta_l | \theta)$$  \hspace{1cm} (11)

where $\log P(x_i, \theta_l | \theta) = -\frac{1}{2}(x_i - \mu_l)^\top \Sigma_l^{-1} (x_i - \mu_l) + \log w_l - \log (2\pi)^{d/2} | \Sigma_l |$  \hspace{1cm} (12)

We obtain the Gaussian clustering method

$$u_{il} = P(x_i, \theta_l | \theta) / \sum_{k=1}^{C} P(x_i, \theta_k | \theta)$$  \hspace{1cm} (13)

$$w_l = \frac{1}{T} \sum_{i=1}^{T} u_{il}, \quad \mu_l = \frac{T}{T} \sum_{i=1}^{T} u_{il} x_i / \sum_{i=1}^{T} u_{il}, \quad \Sigma_l = \frac{T}{T} \sum_{i=1}^{T} u_{il} (x_i - \mu_i) (x_i - \mu_i) / \sum_{i=1}^{T} u_{il}$$  \hspace{1cm} (14)

7. **GATH-GEVA CLUSTERING**

The Gath-Geva (GG) parameter estimation equations are as follows [7]

$$u_{il} = \left\{ \sum_{k=1}^{C} [d(x_i, \theta_l)/d(x_i, \theta_k)]^{\frac{2}{m}} \right\}^{-1}$$  \hspace{1cm} (15)

$$w_l = \frac{1}{T} \sum_{i=1}^{T} u_{il}, \quad \mu_l = \frac{T}{T} \sum_{i=1}^{T} u_{il} x_i / \sum_{i=1}^{T} u_{il}, \quad \Sigma_l = \frac{T}{T} \sum_{i=1}^{T} u_{il} (x_i - \mu_l) (x_i - \mu_l) / \sum_{i=1}^{T} u_{il}$$  \hspace{1cm} (16)

where the distance is defined as $d^2(x_i, \theta_l) = 1/P(x_i, \theta_l | \theta)$ and $P(x_i, \theta_l | \theta)$ is shown in (12). As indicated in [11], in contrast to the FCM and GK methods, the GG method is not based on any objective functions, but is only a fuzzification of statistical estimators. If we were to apply for the GG clustering method the same technique as for the FCM models, i.e. minimising the function $G_m(U, \theta; X)$ in (5) using the GG’s distance, the resulting system of equations could not be solved analytically. In this
sense, the GG method is a good heuristic on the basis of an analogy with probability theory.

8. **Fuzzy Entropy Clustering**

Let $m = 1$ and $n > 0$, the general objective function in (1) becomes the FE objective function as follows

$$G_{m}(U, \theta; X) = \sum_{i=1}^{C} \sum_{r=1}^{T} u_{ir} d^{2}(x_{i}, \theta_{r}) + n \sum_{i=1}^{C} \sum_{r=1}^{T} u_{ir} \log u_{ir}$$

(17)

where $\theta_{r}$ and $d^{2}(x_{i}, \theta_{r})$ are defined as those in (3). The FE parameter estimation equations are as follows

$$u_{ir} = \left\{ \frac{C}{\sum_{k=1}^{C} e^{d^{2}(x_{i}, \theta_{k}) / e^{d^{2}(x_{i}, \theta_{r})}} \right\}^{-1} \quad \mu_{i} = \frac{\sum_{r=1}^{T} u_{ir} x_{i}}{\sum_{r=1}^{T} u_{ir}}$$

(18)

9. **Gustafson-Kessel-Based Fuzzy Entropy Clustering**

Applying (7) to (16), the parameter estimation equations for Gustafson-Kessel-based fuzzy entropy (GK-FE) clustering are as follows

$$u_{ir} = \left\{ \frac{C}{\sum_{k=1}^{C} e^{d^{2}(x_{i}, \theta_{k}) / e^{d^{2}(x_{i}, \theta_{r})}} \right\}^{-1} \quad \mu_{i} = \frac{\sum_{r=1}^{T} u_{ir} x_{i}}{\sum_{r=1}^{T} u_{ir}}$$

(19)

$$\Sigma_{i} = \frac{\sum_{r=1}^{T} u_{ir} (x_{i} - \mu_{i}) (x_{i} - \mu_{i})}{\sum_{r=1}^{T} u_{ir}}$$

(20)

10. **Gaussian-Based Fuzzy Entropy Clustering**

Applying (11) and (12) to (17), the parameter estimation equations for Gaussian-based fuzzy entropy (G-FE) clustering are as follows

$$u_{ir} = \left\{ \frac{C}{\sum_{k=1}^{C} e^{d^{2}(x_{i}, \theta_{k}) / e^{d^{2}(x_{i}, \theta_{r})}} \right\}^{-1} \quad \mu_{i} = \frac{\sum_{r=1}^{T} u_{ir} x_{i}}{\sum_{r=1}^{T} u_{ir}}$$

(21)

$$w_{j} = \frac{1}{T} \sum_{r=1}^{T} u_{ir} \quad \mu_{i} = \frac{\sum_{r=1}^{T} u_{ir} x_{i}}{\sum_{r=1}^{T} u_{ir}}$$

$$\Sigma_{i} = \frac{\sum_{r=1}^{T} u_{ir} (x_{i} - \mu_{i}) (x_{i} - \mu_{i})}{\sum_{r=1}^{T} u_{ir}}$$

(22)

It can be shown that as $n = 1$, the membership function $u_{ir}$ in (21) reduces to the function in (13) using the distance in (11). This means that FE-GMM clustering reduces to GMM clustering. So GMM clustering can be viewed as a special case of FE clustering.
11. **Noise Clustering-Based Fuzzy Entropy Clustering**

It was shown in our previous work [8], a noise clustering (NC) based version of FE clustering can be obtained by applying the following to compute the fuzzy membership function:

$$ u_{it} = \left[ \sum_{k=1}^{C} e^{d^2(s_i, \theta_k) / e^{d^2(s_i, \theta_k)}} \right]^{-1} $$

where the noise is considered to be a separate class and is represented by a prototype that has a constant distance $D$ from all feature vectors. Replacing (18), (19), and (21) with (23), we obtain the NC-FE, NC-GK-FE, and NC-G-FE clustering methods, respectively.

5. **Experimental Results**

Six data sets of two-dimensional vectors and three data sets of three-dimensional vectors [12] are analyzed by the fuzzy clustering methods. Each of clusters contains 100 vectors in the two-dimensional data sets and 200 vectors in the three-dimensional data sets. The two data sets Kreis1.inp in Figure 1 and Ellipsen.inp in Figure 2 consist of 2 clusters. The data set Ellipsen_gk.inp in Figure 3 consists of 3 clusters. The three data sets Ellipsen_gk_parallel.inp in Figure 4, Ellipsen_gg.inp in Figure 5 and Ellipsen_gg_parallel in Figure 6 consist of 4 clusters. The three-dimensional data set Ellipsoid_gk.inp contains 2 clusters and the other two data sets Ellipsoid_gg.inp, and Ellipsoid_gg_parallel.inp contains 3 clusters. We apply the maximum fuzzy membership rule to classify vectors into clusters. Classification rates (100 * number of misclassified vectors / number of total vectors) for the fuzzy clustering methods are shown in Table 1. It can be seen that the noise clustering-based fuzzy entropy (NC-FE) clustering method gives the best results on most of the cluster data sets.

<table>
<thead>
<tr>
<th></th>
<th>FE</th>
<th>GK-FE</th>
<th>G-FE</th>
<th>NC-FE</th>
<th>FCM</th>
<th>GK</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kreis1</td>
<td>0.0</td>
<td>44.5</td>
<td>38.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>47.5</td>
</tr>
<tr>
<td>Ellipsen</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>40.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Ellipsen_gk</td>
<td>36.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>36.0</td>
<td>66.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Ellipsen_gk_parallel</td>
<td>17.5</td>
<td>0.0</td>
<td>50.0</td>
<td>5.0</td>
<td>5.0</td>
<td>47.5</td>
<td>25.5</td>
</tr>
<tr>
<td>Ellipsen_gg</td>
<td>1.0</td>
<td>18.8</td>
<td>18.8</td>
<td>0.0</td>
<td>1.3</td>
<td>75.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Ellipsen_gg_parallel</td>
<td>22.0</td>
<td>24.0</td>
<td>50.0</td>
<td>0.0</td>
<td>11.5</td>
<td>75.0</td>
<td>5.3</td>
</tr>
<tr>
<td>Ellipsoid_gk</td>
<td>12.3</td>
<td>13.0</td>
<td>45.0</td>
<td>11.3</td>
<td>12.0</td>
<td>47.0</td>
<td>16.8</td>
</tr>
<tr>
<td>Ellipsoid_gg</td>
<td>4.2</td>
<td>34.0</td>
<td>33.3</td>
<td>4.0</td>
<td>5.0</td>
<td>22.3</td>
<td>33.3</td>
</tr>
<tr>
<td>Ellipsoid_gg_parallel</td>
<td>0.0</td>
<td>16.7</td>
<td>16.7</td>
<td>0.0</td>
<td>0.0</td>
<td>16.7</td>
<td>23.3</td>
</tr>
</tbody>
</table>

*Table 1: Classification rates (%) for the fuzzy clustering methods*
Figure 1: Kreis1.inp, 200 vectors, 2 clusters

Figure 2: Ellipsen.inp, 200 vectors, 2 clusters

Figure 3: Ellipsen_gk.inp, 300 vectors, 3 clusters

Figure 4: Ellipsen_gk_parallel.inp, 400 vectors, 4 clusters

Figure 5: Ellipsen_gg.inp, 400 vectors, 4 clusters

Figure 6: Ellipsen_gg_parallel.inp, 400 vectors, 4 clusters
7. CONCLUSION

A general fuzzy approach to fuzzy clustering methods has been proposed in this paper. Some new extended versions of fuzzy clustering methods have been proposed from this general approach. The noise clustering-based fuzzy entropy clustering method has shown the highest classification rates on most of the cluster data sets used in this paper.

8. REFERENCES