Context-Based Gustafson-Kessel Clustering with Information Granules

Myung-Won Lee¹, Keun-Chang Kwak²*
¹Dept. of Control and Instrumentation Eng., Chosun University, 309 Pilmun-daero Dong-gu, Gwangju, Korea
²School of Information and Communication Engineering, Myongji University, Gyeonggi-do, Korea

Abstract—In this paper, we propose a Context-based Gustafson-Kessel (CGK) clustering algorithm that builds Information Granulation (IG) in the form of fuzzy set. The fundamental idea of this clustering is based on Conditional Fuzzy C-Means (CFCM) clustering introduced by Pedrycz. The proposed clustering develops clusters preserving homogeneity of the clustered patterns associated with the input and output space. Furthermore, this performs the local adaptation of the distance metric to the shape of the cluster based on fuzzy covariance matrix and linguistic contexts. The experimental results reveal that the proposed clustering algorithm yields better performance in comparison with Fuzzy C-Means (FCM), GK, and CFCM clustering introduced in the previous literature for synthesis data set.

Keywords—Gustafson-Kessel clustering, information granulation, conditional fuzzy c-means clustering, linguistic context

I. INTRODUCTION

Clustering techniques partition whole data set into several small clusters such that the dissimilarity measure within a cluster is smaller than that among clusters. Achieving such a partitioning requires a similarity measure that takes input vectors and returns a value reflecting their similarity. Clustering algorithms are frequently used in conjunction with Radial Basis Function Networks (RBFN) or Fuzzy Modeling (FM) primarily to determine initial locations for radial basis functions or fuzzy if-then rules, respectively [1]. The most representative clustering techniques used in conjunction with model construction are K-means clustering, Fuzzy C-Means (FCM) clustering [2], subtractive clustering [3], mountain clustering [4], Gustafson-Geva (GG) clustering [5], and Gustafson-Kessel (GK) clustering [6], etc. Although these clustering techniques showed a good classification and prediction performance in the design of intelligent model, these were performed by context-free clustering method without considering the homogeneity between input and output spaces [7]. In contrast to context-free clustering, context-based clustering is an interesting way to develop clusters by focusing on a certain portion of the original data. This gives rise to the modularization of clustering and helps carry out more focused data analysis.

From this viewpoint, Pedrycz [8] introduced FCM clustering method guided by a conditional variable, what so called Conditional Fuzzy C-Means (CFCM) clustering. This clustering estimates the clusters preserving homogeneity of the clustered patterns associated with the input and output space from an information of linguistic contexts. This clustering has been used successfully in conjunction with RBFN [9] and Linguistic Model (LM) [10]. However, the cluster shape of CFCM clustering is circle, because this clustering uses Euclidean distance between cluster centers and data points in each context. We can encounter difficulties when the cluster shape is similar to ellipse with correlation between input variables.

II. CONTEXT-BASED GK CLUSTERING ALGORITHM

Gustafson and Kessel (GK) extended the standard FCM clustering algorithm by using an adaptive distance measure to detect the clusters with different geometrical shapes. Each cluster has its own norm-inducing matrix A_i, which computes the following inner-product norm

\[ D_{ik}^2 = (z_k - v_i)^T A_i (z_k - v_i) \]  (1)

The matrices A_i are used as optimization variables, thus allowing each cluster to adapt the distance norm to the local topological structure of the data. The objective function of the GK clustering is defined as follows

\[ Q = \sum_{i=1}^{K} \sum_{k=1}^{N_i} \mu_{ik}^m D_{ik}^2 \]  (2)

This objective function cannot be directly minimized with respect to A_i, since it is linear in A_i. To obtain a feasible solution, A_i must be constrained in some way. In order to accomplish this, we constrain the determinant of A_i as follows

\[ |A_i| = \rho_i, \quad \rho_i > 0, \quad \forall i \]  (3)

The A_i is obtained by using the Lagrange multiplier method as the following equation

\[ A_i = [\rho_i \ det(F_i)]^{1/m} F_i^{-1} \]  (4)

Where F_i is the fuzzy covariance matrix of the i-th cluster.
2.2 Context-based GK (CGK) clustering

We consider the contexts to be described by triangular membership functions being distributed in the output space with the 1/2 overlap occurring between two successive fuzzy sets [10]. In the classification problem, the linguistic contexts are generated uniformly. We denote those fuzzy sets by \( T_1, T_2, \ldots, T_p \) as linguistic contexts.

\[
U(T_i) = \left\{ u_{ik} \in [0,1] \right\} \sum_{i=1}^{c} u_{ik} = f_{ik} \forall k
\]  

(5)

Where \( f_{ik} \) denotes a membership value of the \( k \)-th data point included by the \( t \)-th context. The minimization of objective function is realized by iteratively updating the values of the membership matrix and the prototypes. The update of the membership matrix is computed as follows

If \( D_{ikA} > 0 \) for all \( i = 1,2,\ldots,c \)

\[
u_{ik} = \frac{f_{ik}}{\sum_{j=1}^{c} (D_{ikA} / D_{jkA})^{(m-1)}}
\]  

(6)

Otherwise

\[
u_{ik} = 0
\]

Fig. 1 \( f_k \) value corresponding to output \( y_k \) in the \( t \)-th context

Where \( u_{ik} \) represents the element of the membership matrix induced by the \( i \)-th context and \( k \)-th data in the \( t \)-th context. The Mahalonobis distance can be expressed as follows

\[
D_{ikA}^2 = (z_k - v_i)^T [\mu_i \det(F_i)]^{1/m} F_i^{-1}(z_k - v_i)
\]  

(7)

\[
F_i = \frac{\sum_{k=1}^{N} (u_{ik})^m (z_k - v_i)(z_k - v_i)^T}{\sum_{k=1}^{N} (\mu_{ik})^m}
\]

(8)

The cluster centers \( v_i \) are calculated in the form

\[
v_i = \frac{\sum_{k=1}^{N} u_{ik} z_k}{\sum_{k=1}^{N} u_{ik}^m}
\]  

(9)

Fig. 2 visualizes the concept of CGK clustering in the development of a web of information granules when the number of context is 3 and the number of cluster per context is \([3 4 4]\), respectively. We can recognize from Fig. 2 that the clusters obtained from CGK clustering have the more homogeneity than those produced by context-free fuzzy clustering such as FCM and GK clustering.

![Contexts and input space](image)

**Fig. 2 Context-based GK clustering based on information granules**

III. CLUSTER VALIDATION AND RULE GENERATION

3.1 Cluster Validation

Cluster Validation (CV) is to find the optimal number of cluster from given data set. Several validity measures have been studied to determine the optimal number of cluster. Bezdek [11] proposed two CVs that are called the Partition Coefficient (\( V_{PC} \)) which minimizes an index value and Partition Entropy (\( V_{PE} \)) which maximizes an index using a partion matrix. Xie and Beni index [12] \( V_{XB} \) has the characteristic of intuitive meaning and relieves the problem of finding the optimal number of clusters. The validity measure \( V_{XB} \) is defined as the ratio of compactness and separation of fuzzy partition.
Kim [13] proposed a CV index (VKL) for GK clustering that also to find a validity index. VKL is defined as the average value of the relative degrees of sharing of c(c-1)/2 pairs of clusters, where the relative degree of sharing of each cluster pair is defined as the weighted sum of the relative degree of sharing at zj of two clusters in the pair.

\[ V_{PC} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^2}{n} \]  

(10)

\[ V_{PE} = -\frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{c} u_{ij} \log_2(\mu_{ij}) \]  

(11)

\[ V_{XB} = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^2 \left( \sum_{j=1}^{c} \sum_{i=1}^{n} \frac{\|v_i - z_k\|^2}{\|v_i - v_j\|^2} \right) \]  

(12)

\[ V_{KL} = \frac{2}{c(c-1)} \sum_{p \neq q} \sum_{j=1}^{c} \left[ \left( \mu_{p}^j(z_j) \cap \mu_{q}^j(z_j) \right) \mu(z_j) \right] \]  

(13)

3.2 Determination of initial locations for RBF or fuzzy rules

The CGK clustering algorithms can be used in conjunction with the design of RBFN or FM primarily to determine initial locations for Radial Basis Functions (RBF) of receptive field or fuzzy if-then rules. In the design of RBFN, the weights are characterized by the linguistic contexts between hidden layer and output layer, while the general RBFNs are connected by weights with real value. The cluster centers obtained by CGK clustering are used as those of Gaussian RBF in the hidden layer. The model output of RBFN can be represented by the uncertain output values with lower and upper bound. In the design of FM, fuzzy rules are generated by cluster centers obtained from each context in CGK clustering. The effective partitioning of the input space through CGK clustering can decrease the number of rules and thus increase the speed in both the learning and application phases.

IV. EXPERIMENTAL RESULTS

We start with two-dimensional set of 900 patterns (300 patterns each class) with three classes (red, blue, magenta) as shown in Fig. 3. For simplicity, we assume that the number of cluster per context is the same. We used the proposed CGK clustering for the classification of synthesis data set, while the number of cluster per context increases.

Fig. 4 and 5 show the cluster centers obtained by FCM and GK clustering when the number of cluster is 6, respectively. As shown in these figures, we can recognize that the clusters obtained by these clustering algorithms are not valid. Fig. 6 visualizes the clusters obtained by CFCM clustering when the number of cluster per context is 2 (p=3, c=2). As shown in Fig. 6, although the context-based clustering is used, the cluster shape is not appropriate to data distribution of each class. In contrast, Fig. 7 shows the cluster centers estimated by the CGK clustering when the number of cluster per context is the same as Fig. 6. As shown in Fig. 7, it represents a good shape of the cluster based on fuzzy covariance matrix and linguistic contexts.
Furthermore, these results revealed that the proposed CGK clustering method showed good classification performance in comparison to CFCM clustering.

<table>
<thead>
<tr>
<th>Method</th>
<th>#</th>
<th>classification rates(%)</th>
<th>$V_{PC}$</th>
<th>$V_{PE}$</th>
<th>$V_{XB}$</th>
<th>$V_{KL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFCM</td>
<td>6</td>
<td>84.11</td>
<td>0.57</td>
<td>0.89</td>
<td>10.43</td>
<td>472.19</td>
</tr>
<tr>
<td>[8]</td>
<td>9</td>
<td>89.22</td>
<td>0.49</td>
<td>1.15</td>
<td>20.56</td>
<td>620.72</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>92.33</td>
<td>0.47</td>
<td>1.26</td>
<td>21.07</td>
<td>653.39</td>
</tr>
<tr>
<td>CGK</td>
<td>6</td>
<td>93.00</td>
<td>0.66</td>
<td>0.67</td>
<td>13.16</td>
<td>238.75</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>94.33</td>
<td>0.56</td>
<td>0.93</td>
<td>5.49</td>
<td>365.70</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>94.89</td>
<td>0.52</td>
<td>1.08</td>
<td>4.38</td>
<td>413.00</td>
</tr>
</tbody>
</table>

**V. CONCLUSIONS**

We have developed a novel context-based GK clustering based on information granules. The experimental results lead us to the conclusion that the proposed clustering algorithm can develop clusters preserving homogeneity of the clustered patterns with the aid of linguistic contexts. Furthermore, this clustering can represent the local adaptation of the distance metric to the shape of the cluster based on fuzzy covariance matrix showing good classification performance and geometrical characteristics.

**ACKNOWLEDGMENTS**

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2013R1A1A2012127)

**REFERENCES**


Table 1 lists the comparison results of four cluster validations and classification rates. As listed in Table 1, it is concluded that the best number of cluster is 12 (p=3, c=4) in all of cluster validation methods.


