Advanced Fuzzy Possibilistic C-means Clustering Based on Granular Computing

Hung Quoc Truong, Long Thanh Ngo
Department of Information Systems
Le Quy Don Technical University, Hanoi, Vietnam
Email:truongqhung@gmail.com, ngotlong@gmail.com

Witold Pedrycz
Department of Electrical & Computer Engineering
University of Alberta, Edmonton T6R 2V4 AB Canada
Email:wpedrycz@ualberta.ca

***Abstract*—With the rapid development of uncertain and large-scale datasets, Fuzzy Possibilistic C-means Clustering (FPCM) and Granular Computing (GrC) were introduced together with the aim to solve the feature selection and outlier detection problems. Utilizing the advantages of the FPCM and GrC, an Advanced Fuzzy Possibilistic C-means Clustering based on Granular Computing (GrFPCM) was proposed to select features as a preprocessing step for clustering problems and granular space is used to handle the uncertainty. Experimental results reported for various datasets in comparison with other approaches exhibit the advantages of the proposed method.**

***Index Terms*—fuzzy clustering, fuzzy possibilistic c-means clustering, granular computing, feature selection, outlier detection.**

I. Introduction

Clustering is a technique widely used in data mining. It is used to detect any structures or patterns in datasets, in which objects in one cluster exhibit substantial similarity. Clustering algorithms involve various methods, e.g hard clustering like k-means and its various [1] or fuzzy clustering like Fuzzy C-mean clustering (FCM) [5].

Fuzzy clustering algorithms were designed to deal with the uncertainty or vague information. A variant of fuzzy clustering is based on possibilistic approach which was first proposed by Krishnapuram et al. [4]. The algorithm determines a possibilistic partition in which a possibilistic membership is used to define the absolute degree of typicality of a point in any particular clusters. The larger the distance between an object to a centroid, the lower the possibilistic membership grade and the lower the object affects on clustering of the centroid. Therefore, methods of outlier detection or noise removal may be applied. However, the possibilistic approach still exists some drawbacks such as identical clusters and choosing its parameters. Therefore, Zhang et al. [2] proposed a combination method between Fuzzy C-means and Possibilistic C-means, namely Fuzzy Possibilistic C-means (FPCM), to resolve the identical issue of the possibilistic approach.

Nowadays, clustering problems often are used to deal with large and high dimensional datasets which also raises some issues to be resolved to retrieval useful information from these datasets [11]. The one is how to remove noises and redundant features, also called dimensional reduction. Most of the clustering algorithms in general and FPCM algorithm in particular are sensitive to large or high dimensional data or both. In deed, an object has many features and each feature has a different role or some features evenly are noises. An efficient way to handle this issue is selection of a subset of relevant features. Dimensional reduction will be efficient method to find clusters by mining better datasets through the relevant features and reducing data size for efficient storage, collection and processing [16].

Feature selection is one of the commonly used techniques to reduce the number of dimensions. It aims to select a subset of the relevant features according to an evaluation criterion so that the selected features fully represent the dataset to solve the problem [3]. The feature selection is capable of improving the learning performance by removing irrelevant features and decreasing the required storage through reducing the number of features. Thus, feature selection has commonly been used to remove irrelevant features and improve the performance of classification.

Many heuristic algorithms of feature selection have been proposed to reduce the number of dimensions. While J. Qian proposed an attribute reduction algorithms for big data using a map reduce [13], L.Sun et al. designed a feature selection method based on rough entropy [8], [10] or granular computing [15]. In addition, Q.H.Hu et al. introduced a feature selection method which was formed by combining granular computing and approximation theory [11]. However, these feature selection methods need the labelled samples as training samples to select the necessary features, applications were only focused on the classification or decision system problems.

Meanwhile, clustering is completed for unlabeled data which poses a challenge in feature selection. In such cases, definition of relevant features is unclear and need to be resolved. However, the feature selection methods, which were proposed for clustering as filter, wrapper and hybrid models, are usually designed based on the greedy approach according to an evaluation criterion [16]. It results in time-consuming and low efficiency for very large-scale dimensional data.

Recently, granular computing is a powerful tool to study granulation for handling complex problem, massive data, uncertain information and high dimensional data [6], [7]. Granular computing has become a new method which can simulate human thinking and solve problems in computational intelligence which concern the idea of granularity and the logic of granularity [14] and it is also used as a base for feature selection methods [11], [15].

There are many hybrid models between granular computing and other methods which are used to form a new type of machine learning algorithms. These methods are based on granule structure for various types of dataset or learning methods [9], [12].

In light of this brief review, GrC can be combined with clustering method to utilize feature selection for clustering to alleviate the negative impact of the high dimensional problem. In addition, selecting subset(s) of features may help improving the clustering results similar to improving the supervised methods as classification and decision systems [14].

Therefore, an advanced Fuzzy Possibilistic C-means Clustering is proposed on basis of a combination of FPCM algorithm with granular computing to handle noise removal or outlier detection and feature selection for dealing with high dimensional data. The proposed method not only takes advantage of FPCM ability in handling noises but also uses the granular computing theory to assess the significance of the features, which is a basic for eliminating the effects of irrelevant features and noise objects. Thus, the algorithm potentially enhance the clustering results. Experiments are implemented on several high dimensional datasets to illustrate the proposed method.

The paper is organized as follows: Section II briefly introduces some backgrounds about fuzzy clustering, fuzzy possibilistic c-means clustering and granular computing; Section III proposes the advanced fuzzy possibilistic c-means clustering based on granular clustering; Section IV offers some experimental results and section V covers a conclusion and proposes future research directions.

II. Preliminaries

A. Fuzzy Possibilistic C-Means Clustering Algorithm

Fuzzy Possibilistic C-Means Clustering Algorithm (FPCM) was proposed by Zhang et al. [2]. FPCM algorithm is built based on a combination of two algorithms FCM [5] and PCM [4], which has two types of memberships: 1) A possibilistic membership that measures the absolute degree of typicality of a point in any particular cluster, and 2) a fuzzy membership that measures the relative degree of sharing of a point among the clusters.

The objective function for FPCM was built as follows:

$$J_{FPCM}(T,U,V;X,\gamma)=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^{m}t_{ik}^{p}d_{ik}^{2},\ 1\le m,p\le\infty \tag{1}$$

$$+\sum_{i=1}^{c}\gamma_i\sum_{k=1}^{n}u_{ik}^{m}\left(1-t_{ik}\right) \tag{2}$$

in which $d_{ik}=\parallel x_k-v_i\parallel$ is Euclidean distance between the object x_k and the centroid v_i, c is the number of clusters, n is the number of objects, p is a weighting exponent of possibilistic membership ($p>1$) and fuzzifier m ($m>1$).

The scale parameter γ_i standing in (2) is to incorporate the possibilistic membership degrees and fuzzy membership ones:

$$\gamma_i=K\frac{\sum_{k=1}^{n}t_{ik}^{p}u_{ik}^{m}d_{ik}^{2}}{\sum_{k=1}^{n}t_{ik}^{p}u_{ik}^{m}},K>0 \tag{3}$$

where K is a constant.

t_{ik} denoted the possibilistic membership degree of x_k belonging to the i^{th} cluster and u_{ik} denoted the degree of fuzzy membership. They are determined as follows:

$$t_{ik}=\frac{1}{1+\left(\frac{d_{ik}^{2}}{\gamma_i}\right)^{\frac{1}{p-1}}},\forall i,k \tag{4}$$

$$u_{ik}=\frac{1}{\sum_{j=1}^{c}\left(\frac{t_{ik}^{(p-1)/2}\,d_{ik}}{t_{jk}^{(p-1)/2}\,d_{jk}}\right)^{\frac{2}{m-1}}} \tag{5}$$

in which $i=1,2,...,c$; $k=1,2,...,n$.

The centroids are computed as follows:

$$v_i=\frac{\sum_{k=1}^{n}t_{ik}^{p}u_{ik}^{m}x_k}{\sum_{k=1}^{n}t_{ik}^{p}u_{ik}^{m}},\forall i \tag{6}$$

in which $i=1,2,...,c$.

Defuzzification in the FPCM is made as if $u_{ik}>u_{jk}$ for $j=1,2,...,c$ and $j\neq i$ then x_k is assigned to the i^{th} cluster.

This algorithm can be briefly described as follows:

Algorithm 1 Fuzzy Possibilistic C-Means Clustering algorithm

1 Input: A dataset $X=\{x_i,x_i\in R^d\},i=1,2,...,n$, the number of clusters $c\,(1<c<n)$, weighting exponents $p,m(1<p,m<+\infty)$ and error ε.

2 Output: The possibilistic membership matrix T, fuzzy membership matrix U and the centroid matrix V.

3 Step 1:
3.1 The number of iterations is set to $l=0$.
3.2 Execute a fuzzy c-means clustering algorithm [5] to find an initial $U^{(l)}$ and $V^{(l)}$.
3.3 Compute $\gamma_1,\gamma_2,...,\gamma_c$ based on the $U^{(l)}$ and $V^{(l)}$ as follows: $\gamma_i=\frac{\sum_{k=1}^{n}u_{ik}^{m}d_{ik}^{2}}{\sum_{k=1}^{n}u_{ik}^{m}}$

4 Step 2:
repeat :
4.1 $l=l+1$.
4.2 Update the possibilistic membership matrix $T^{(l)}$ by using (4).
4.3 Update the fuzzy membership matrix $U^{(l)}$ by using (5).
4.4 Update the centroid matrix $V^{(l)}=\left[v_1^{(l)},v_2^{(l)},...,v_c^{(l)}\right]$ by using (6).
4.5 Apply (3) to compute $\gamma_1,\gamma_2,...,\gamma_c$ based on the $T^{(l)}$,$U^{(l)}$ and $V^{(l)}$
until :
$$Max\left(||U^{(l+1)}-U^{(l)}||\right)\le\varepsilon$$

5 Assign data x_k to i^{th} cluster if $u_{ik}>u_{jk}$, $j=1,2,...,c$ and $j\neq i$.

B. Granular Computing

Granular computing is a new concept and computing paradigm of processing information. When using granular computing in clustering, a granule is formed by a set of elements which are drawn together by indistinguishability, similarity, proximity or functionality.

Considering a clustering system $S = (X, A, V, f)$ denoted as $S(X, A)$ with $X = \{x_1, x_2, ..., x_n\}$ being a non-empty finite set of objects; $A = \{a_1, a_2, ..., a_d\}$ is a non-empty finite set of features; $V = \bigcup_{a \in A} V_a$ with V_a is called the value domain of the feature a, f is the information function of the system, $f : X \times A \rightarrow V$

Some definitions which were proposed in [17], are introduced to granulate the clustering system as follows.

While Def.2.1 is used to determine an indiscernibility relation between objects of X of a clustering system $S = (X, A)$ on a subset of features, Def.2.2 is used to construct an information granule of the clustering system S based on an relation which is defined in Def.2.1. On the basis of the granule, a definition of granularity of the clustering system on a subset of features is built in Def.2.3, which is used to assess the impact of the features on the clustering system S.

Definition 2.1: For each subset of features $B \subseteq A$, the non-empty set determines an indiscernibility relation on X as follows: $R_B = \{(x_i, x_j) \in X \times X | f_a(x_i) = f_a(x_j), \forall a \in B\}$ R_B is an equivalence relation on X, and it forms a partition of X, denoted by $X/R_B = \{[x_i]_B | x_i \in X\}$ where $[x_i]_B = \{x_j \in X | (x_i, x_j) \in R_B\}$ is called an equivalence class of x_i with respect to B.

A granule for clustering system is defined as follows:

Definition 2.2: Let $S = (X, A)$ be a clustering system. An information granule is defined as $gr_k = (\varphi_k, m(\varphi_k))$, where φ_k refers to the intention of information granule, and $m(\varphi_k)$ represents the extension of information granule.

Suppose that $B = \{a_1, a_2, ..., a_{d'}\} \in A$ then there must exist $\varphi_k = \{I_1, I_2, ..., I_{d'}\}$ such that $I_j \in V_{a_j}$ is a set of feature values corresponding to B. Then, the intention of an information granule can be denoted by $\varphi_k = \{I_1, I_2, ..., I_{d'}\}$, and the extension can be denoted by $m(\varphi_k) = \{x \in X | f(x, a_1) = I_1 \wedge f(x, a_2) = I_2 \wedge ... \wedge f(x, a_{d'}) = I_{d'}, a_j \in B\}$, $j \in \{1, 2, ..., d'\}$. Here, $m(\varphi_k)$ describes the internal structure of the information granule.

A granularity of system of features set B, denoted $GK(B)$, which is defined for examining the maintenance of clustering system.

Definition 2.3: Let $S = (X, A)$ be a clustering system, the concept Granularity of System of features set B based on the Granules set $Gr = \{gr_k\}$ denoted $GK(B)$, $B \subseteq A$, is determined as follows:

$$GK(B) = \sum_{k=1}^{|Gr/B|} |m(\varphi_k)|^2 / |X|^2,\ m(\varphi_k) \in gr_k.$$

For example, the dataset $X = \{x_1, x_2, x_3, x_4\}$, $x_i \in R^3$, the set of features $A = \{a_1, a_2, a_3\}$ and $B = \{a_1, a_2\}$, where $x_1 = (1, 2, 3), x_2 = (1, 2, 1), x_3 = (2, 3, 1)$ and $x_4 = (1, 2, 2)$. Suppose $I_j = f(x_i, a_j) = x_i^{(j)}$ then we obtain the set of granules $Gr/B = \{gr_1, gr_2\}$, in which $gr_1 = (\varphi_1, m(\varphi_1))$, $\varphi_1 = (1, 2)$, $m(\varphi_1) = \{x_1, x_2, x_4\}$, and $gr_2 = (\varphi_2, m(\varphi_2))$, $\varphi_2 = (2, 3)$, $m(\varphi_2) = \{x_3\}$. Resulting in $GK(B) = (3^2/4^2) + (1^2/4^2) = 10/16$.

III. Advanced Fuzzy Possibilistic C-means Clustering based on Granular computing

A. Feature reduction base on Granular Computing

According to concepts of granular computing, the significance of a set of features in clustering system was proposed [17]. Given a clustering system $S = (X, A)$, there is a feature in A, denoted $a \in A$, so that we can measure the degree of importance through the quantity of the granularity of A if the feature a is removed. Thus, Def.3.1,Def.3.2 and Def.3.3 are constructed to determine a reduction set of features $C : C \subseteq A$ based on the concept of the granularity of A which is defined in Def.2.3.

Definition 3.1: The significance degree of feature $a \in A$, denoted $Sig_{A-\{a\}}(a)$, is defined as follows:

$$Sig_{A-\{a\}}(a) = GK(A - a) - GK(A)$$

Note that the larger degree $Sig_{A-\{a\}}(a)$ takes, the more important the feature a is.

Definition 3.2: Given an information system $S = (X, A)$ and feature $a \in A$, the feature a is called redundant feature to A if the value of $GK(A - a)$ is equal $GK(A)$. Otherwise, the feature a is called necessary feature to A. The set of all the necessary features is the core of A, denoted $Core(A)$.

Definition 3.3: Given an information system $S = (X, A)$ and a set of features $C : C \subseteq A$. Set C is called a reduction of A if C is independent. All the reduction of A is denoted by Red(A).

The reduction algorithm can be briefly described as follows:

Algorithm 2 Feature reduction based on Granular Computing

1 Input: A granular information system S=(X,A) where $X \neq \varnothing$ is the universe and $A \neq \varnothing$ is the set of features. The granularity of A is denoted as $GK(A)$.
2 Output: C is as the minimum reduction of A.
3 Step 1. Determine the core of features $Core(A)$ as follow: Calculate the significance degree of each feature $a \in A$, denoted $Sig_{A-\{a\}}(a)$, if $Sig_{A-\{a\}}(a) \neq 0$ then select feature a into $Core(A)$.
4 Step 2.
  4.1 Assign $C := Core(A)$.
  4.2 If $GK(C) = GK(A)$ then terminal criteria is meet.
  4.3 **repeat** :
    4.3.1 For each feature $a \in A - C$ to C, calculate its significance degree to $C \cup \{a\}$: $Sig_C(a)$.
    4.3.2 Find the feature a so that its significance degree to C reach the maximal value, i.e. $Sig_C(a) = \max_{a' \in A-C} \left(Sig_C \left(a' \right) \right)$
    4.3.3 Add feature a to the core, i.e. $C := C \cup \{a\}$
  until : $GK(C) = GK(A)$

B. Granular space construction and feature selection

Let consider a clustering system $S = (X, A)$ where $X = \{x_1, x_2, ..., x_n\}$ and $A = \{a_1, a_2, ..., a_d\}$. We construct a granular space as follows:

First, the objects $X = \{x_1, x_2, ..., x_n\}$ are clustered into c clusters on each j^{th} feature by FPCM algorithm, $j \in A$. On each j^{th} feature, the clusters are labeled by numbering in ascending order, i.e. 1, 2, 3.

Second, a cluster label matrix, denoted F, is formed from $f(i, j)$ which is the label of the i^{th} object on the j^{th} feature, $1 \leq f(i, j) \leq c$, i.e. $F = [f(i, j)]_{(n \times d)}$.

Finally, from the values $\{f_1, f_2, ..., f_d\}$ of a row in the cluster label matrix F, we can construct a granule $gr_k = \{\varphi_k, m(\varphi_k)\}$ where $\varphi_k = \{f_1, f_2, ..., f_d\}$, $m(\varphi_k) = \{x_i \in X : f(i, 1) = f_1 \wedge f(i, 2) = f_2 \wedge ... \wedge f(i, d) = f_d\}$. So a granular space, denoted G, is formed from the set of granules, i.e. $G = \{gr_k\}, k = 1, 2, ..., n_g$ with n_g is the number of the granules, $1 \leq n_g \leq n$, denoted $n_g = |G|$.

However the set of granules can be divided into two types of granule: non-conflict and conflict granules which are defined as follows:

Definition 3.4: Consider a granular clustering system $S = (G, A)$, granular space $G = \{gr_k\}, k = 1, 2, ..., n_g$ and $n_g = |G|$. A non-conflict granular space with respect to A, denoted $GrSP$, is formed by $GrSP = \{gr_{k_1}\}$, in which $gr_{k_1} = \{\varphi_{k_1}, m(\varphi_{k_1})\}$ where $\varphi_{k_1} = \{f_1, f_2, ..., f_d\}$ and $f_1 = f_2 = ...f_d$. Otherwise, a conflict granular space with respect to A, denoted $GrSN$, is formed by $GrSN = \{gr_{k_2}\}$, in which $gr_{k_2} = \{\varphi_{k_2}, m(\varphi_{k_2})\}$, $\varphi_{k_2} = \{f_1, f_2, ..., f_d\}$ and $\exists f_p \neq f_q$

Remark: The significance of a feature only affect on the $GrSN$, thus the feature selection method be only applied on the $GrSN$.

In the FPCM algorithm, the outlier or noisy object x_k can be removed, $X := X - \{x_k\}$ if x_k satisfies the following conditions:

$$t_{ik}^{(j)} < \theta \ with \ \forall i = 1, 2, ..., c \ and \ j = 1, 2, ..., d \quad (7)$$

Where $t_{ik}^{(j)}$ is the possibilistic membership degree of x_k on the j^{th} feature in cluster i and θ is a noisy parameter.

Besides, the noisy feature a_j, $a_j \in A$ can be also removed, if $f(1, j) = f(2, j) = ... = f(n', j)$, where n' is the number of object in X after removing the outliers.

$$A := A - \{a_j\} \quad (8)$$

The granular space construction and feature selection method can be briefly described as follows:

Algorithm 3 The granular construction and feature selection

1 Input: A dataset $X = \{x_i\}, i = 1..n$, $A = a_1, a_2, ..., a_d$, c is the number of cluster and θ is a noise filter parameter.

2 Output: The feature set C is the minimum reduction of A and the granular space G=$GrSN \cup GrSP$

3 Step 1:
3.1 Execute Algorithm 1 on each feature $a_j \in A$ to form a cluster label matrix $F = [f(i, j)]_{(n \times d)}$ where $f(i, j)$ is the cluster label of the i^{th} object on the j^{th} feature.
3.2 Remove outlier objects and features by using (7) and (8), respectively.

4 Step 2: Construct granular space
4.1 Initialize $GrSP = \varnothing, GrSN = \varnothing, r = 0, ID = \{1, 2, ..., n\}, k = 0$, where r is the index of row of the matrix F, ID is the index set and k is the number of granules.
4.2 **repeat**
4.2.1 $k = k + 1$
4.2.2 **repeat**
$r = r + 1$
until $r \in ID$
4.2.3 Set φ_k to set of values of r^{th} row in the matrix F: $\varphi_k = f(r, 1), f(r, 2), ..., f(r, d')$, where d' is the number of features in A after removing the outliers.
4.2.4 Find $m(\varphi_k) = \{x_i \in X : f(i, 1) = f(r, 1) \wedge f(i, 2) = f(r, 2) \wedge ... \wedge f(i, d') = f(r, d')\}$.
if $|m(\varphi_k)| > 0$ **then**
4.2.4.1 **for** each $x_i \in m(\varphi_k)$:
$X = X - \{x_i\}, ID = ID - \{i\}$
4.2.4.2 $gr_k = (\varphi_k, m(\varphi_k))$
4.2.4.3 **if** $f(r, 1) = f(r, 2) = ... = f(r, d')$ **then**
$GrSP = GrSP \cup \{gr_k\}$
else
$GrSN = GrSN \cup \{gr_k\}$
until $ID = \varnothing$

5 Step 3: Apply Algorithm 2 on the the granular set GrSN to reach the minimum reduction C of A.

C. Advanced FPCM based on Granular computing

Consider a granular clustering system $S = (G, A)$, granular space $G = \{gr_k\}$, $k = 1, 2, ..., n$ and $n = |G|$.

The valued interval of the j^{th} feature of a input granule $gr_k = (\varphi_k, m_k(\varphi_k))$ is denoted $I_j^{(k)} = [a_j \ b_j]$ where a_j and b_j is defined as follows:

$$a_j = min(x_i^{(j)}), \forall x_i \in m_k(\varphi_k) \quad (9)$$

$$b_j = max(x_i^{(j)}), \forall x_i \in m_k(\varphi_k) \quad (10)$$

in which $x_i^{(j)}$ is the value of the object x_i on the j^{th} feature.

The new distance between a granule gr_k and the centroid $v_i = \{v_{i1}, v_{i2}..., v_{id}\}$, $d = |A|$, $i = 1, 2, ..., c$ is defined as follows:

$$\|gr_k - v_i\| = \sqrt{\sum_{j=1}^{d} \left(\left\|I_j^{(k)} - v_{ij}\right\|\right)^2} \quad (11)$$

where

$$||I_j^{(k)} - v_{ij}|| \overset{def}{=} \begin{cases} 0, \text{ if } v_{ij} \in [a_j, b_j] \\ \min\left(|a_j - v_{ij}|, |b_j - v_{ij}|\right) \end{cases} \quad (12)$$

The distance (11) is used to compute the possibilistic membership function and fuzzy membership function as follows:

t_{ik} is the possibilistic membership degree of the granule gr_k in the i^{th} cluster and u_{ik} is the fuzzy membership degree. They are determined in a similar was as in the FPCM algorithm:

$$t_{ik} = \frac{1}{1 + \left(\frac{d_{ik}^2}{\gamma_i}\right)^{\frac{1}{p-1}}}, \forall i, k \quad (13)$$

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{t_{ik}^{(p-1)/2} \, d_{ik}}{t_{jk}^{(p-1)/2} \, d_{jk}}\right)^{\frac{2}{m-1}}} \quad (14)$$

in which $i = 1, 2, ..., c$, $k = 1, 2, ..., n$.

d_{ik} is calculated by using (11), if the distance between granule gr_k and v_i equals to 0 then the fuzzy membership u_{ik} is assigned to 0.

Cluster centroids are computed in the same way of FPCM as follows:

$$v_i = \frac{\sum_{k=1}^{n} t_{ik}^{p} u_{ik}^{m} \sum_{t=1}^{|m_k(\varphi_k)|} x_t | x_t \in m_k(\varphi_k)}{\sum_{k=1}^{n} t_{ik}^{p} u_{ik}^{m}}, \forall i \qquad (15)$$

in which $i = 1, 2, ..., c$.

The GrFPCM algorithm can be briefly described as follows:

Algorithm 4 Advanced FPCM based on Granular Computing

1. Input: A clustering system $S(X, A)$ where a dataset $X = \{x_1, x_2, ..., x_n\}$, a set of features $A = a_1, a_2, ..., a_d$, the number of cluster c, error ε and noisy parameter θ.
2. Output: The possibilistic membership matrix T, fuzzy membership matrix U and the centroid matrix V
3. Step 1: Apply Algorithm 3 on the clustering system $S(X, A)$ to obtain the feature set C which is the minimum reduction of A and the granular space G.
4. Step 2:
   Apply Algorithm 1 on the clustering system $S = (G, C)$ as follows:
   4.1 The number of iterations is set to $l = 0$
   4.2 **repeat** :
   4.2.1 $l = l + 1$.
   4.2.2 Update the possibilistic membership matrix $T^{(l)}$ by using (13).
   4.2.3 Remove the outlier or noisy granular $gr_{t_{ik} \geq \theta} = \{gr_k \in G : \max(t_{ik}) \geq \theta, \forall i = 1, 2, ..., c\}$
   4.2.4 Update the fuzzy membership matrix $U^{(l)}$ by using (14).
   4.2.5 Update the centroids $V^{(l)} = \left[v_1^{(l)}, v_2^{(l)}, ..., v_c^{(l)}\right]$ by using (15).
   4.2.6 Apply (3) to compute $\gamma_1, \gamma_2, ..., \gamma_c$ based on the $T^{(l)}$, $U^{(l)}$ and $V^{(l)}$
   until :
   $$Max\left(||U^{(l+1)} - U^{(l)}||\right) \leq \varepsilon$$
5. Assign data gr_k to the i^{th} cluster if $u_{ik} > u_{jk}$, $j = 1, 2, ..., c$ and $j \neq i$.

IV. Experiments

In this section, some well-known available datasets with the pre-defined number of clusters are used in the experiments. We also offer a comparative analysis of the clutering results between various clustering methods involved: FCM, PCM, FPCM and GrFPCM (the proposed methods in this study).

Through the adjustments in the experiments, the clustering results are stable with parameters which are set as follows:

Exponential parameters m and p are set to 2, the noise parameter $\theta = 0.1$, error $\varepsilon = 0.00001$, the adjustment γ in FPCM and GrFPCM methods is calculated with $K = 1$

The resulting classification performance of the classification is evaluated by determining True Positive Rate (TPR) and False Positive Rate (FPR) defined as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}; \text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \qquad (16)$$

where TP is the number of correctly classified data, FN is the number of incorrectly misclassified data, FP is the number of incorrectly classified data and TN is the number of correctly misclassified data.

A. Experiment 1

In the first experiment, the well-known datasets are Wisconsin Diagnostic Breast Cancer (WDBC), E. coli promoter gene sequences (DNA) and Madelon [1] are considered. Detailed characteristic datasets and the set of minimum reduction of features are shown in Table I. In which the features of datasets are indexed from 1 (not 0).

TABLE I
CHARACTERISTIC DATASETS AND FEATURE SELECTION

Dataset	No of Instance	No of Features	Class	Feature Selection
WDBC	569	30	2	a_7, a_{11}, a_{22}, a_{27}
DNA	106	57	2	a_{14}, a_{16}
Madelon	4400	500	2	a_{48}, a_{64}, a_{119},a_{201}, a_{241}, a_{277}, a_{310}, a_{321}, a_{362}, a_{417}, a_{472}, a_{475}

The datasets in Table I are clustered by running FCM, FPCM and GrFPCM with the number of clusters is the number of classes. While FCM and FPCM perform the clustering on the datasets with all features, GrFPCM performs clustering on the granular space G with the reduced features which is the output of Algorithm 3. The clustering results or the quality of classification are reported in terms of indices TPR and FPR, which are shown in Table II.

Table II shows the clustering results in which the higher TPR value and lower FTR value, the better algorithm is. The GrFPCM obtained the highest TPR and the smallest FPR on three datasets with the feature selection is shown in Table II.

B. Experiment 2

The five public cancer datasets are Lymphoma, Leukaemia, Global Cancer Map, Embryonal Tumours and Colon [2], are used to illustrate the proposed method in the case of high-dimensional datasets. The datasets are shown in Table IV.

TABLE IV
CHARACTERISTIC DATASETS AND FEATURE SELECTION (FS)

Dataset	No of Instance	No of Original Features	Class	FS
Lymphoma	45	4026	2	15
Leukaemia	38	7129	2	6
Global Cancer Map	190	16063	14	16
Embryonal Tumours	60	7129	2	8
Colon	62	2000	2	9

The experiment has been carried out in the following scenario:

First, every dataset with the original features in Table IV is clustered by FCM and FPCM. Second, the features of datasets in Table IV is reduced by Algorithm 3. Then, every dataset with the reduced features is clustered by running

[1] http://www.ics.uci.edu/ mlearn/mlrepository.html
[2] http://www.upo.es/eps/bigs/datasets.html

TABLE II
CLUSTERING RESULTS FOR EXPERIMENT 1

Dataset	FCM			FPCM			GrFPCM		
	FS	TPR	FPR	FS	TPR	FPR	FS	TPR	FPR
WDBC	30	89.5%	4.5%	30	92.7%	2.8%	4	95.4%	1.9%
DNA	57	85.6%	6.7%	57	91.4%	3.1%	2	96.1%	1.7%
Madelon	500	86.1%	5.9%	500	90.8%	3.3%	12	94.8%	2.1%

TABLE III
CLUSTERING RESULTS FOR EXPERIMENT 2

Dataset	FCM		FCM (FS)		FPCM		FPCM(FS)		GrFPCM	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
Lymphoma	89.2%	4.6%	89.9%	4.2%	89.8%	3.1%	93.2%	1.8%	96.1%	1.7%
Leukaemia	72.1%	9.5%	82.1%	7.2%	81.4%	7.3%	89.4%	4.2%	93.6%	1.4%
Global Cancer Map	89.6%	4.8%	90.4%	3.2%	90.2%	5.5%	93.2%	2.5%	96.8%	1.2%
Embryonal Tumours	80.1%	9.1%	87.6%	6.3%	88.1%	7.6%	91.1%	4.6%	95.3%	1.9%
Colon	79.1%	7.9%	81.7%	6.9%	80.9%	9.5%	86.8%	4.9%	92.8%	3.4%

sequentially algorithms FCM and FPCM. Finally, GrFPCM performs clustering on the granular spaces with the reduced features which are the output of Algorithm 3. In which the number of clusters is assigned to the number of classes.

The clustering results or the quality of classification are reported in terms of indices TPR and FPR in Table III. In which FCM and FPCM columns are results by running the algorithms datasets with all features, FCM(FS) and FPCM(FS) columns are results obtained by running the ones with the reduced features.

From Table III, the TPR values obtained by running GrF-PCM on five datasets are greater 92% and obviously higher than the ones obtained from other methods. In addition, the FPR values are also smaller than the ones reached from other algorithms. In addition, the TPR and FPR values obtained by FCM and FPCM after reducing features are better than ones obtained by FCM and FPCM without reducing features, respectively.

Therefore, we can conclude that as forming the granular space for experimental datasets for handling the uncertainties, noises and the irrelevant features , the quality of the clustering results has been improved.

V. CONCLUSION

This paper presented an advanced fuzzy possibilistic c-means clustering method based on granular computing, which can reduce the features of datasets to obtain a set of key features, while eliminating the facial features. In addition, the proposed method being endowed with granular computing becomes beneficial when it comes to handle the uncertainties.

The experiments completed for several well-known datasets show that the proposed method generates better results than those produced by some other existing clustering methods.

Some next studies may be focused on the use of evolutionary methods (such as Genetic algorithms) to optimize parameters of the clustering method.

REFERENCES

[1] K.R. Zalik, An efcient k-means clustering algorithm, Pattern Recognition Letters, vol. 29, pp. 1385 - 1391, 2008.

[2] J.-S Zhang and Y.-W. Leung, Improved Possibilistic C-Means Clustering Algorithms, IEEE Trans. on Fuzzy Systems, vol. 12(2),pp.209-217, 2004.

[3] K. Kira and L. A. Rendell, A Practical Approach to Feature Selection, the Ninth Int' Workshop on Machine Learning, pp.249-256, 1992.

[4] R. Krishnapuram and J. Keller, A possibilistic approach to clustering, IEEE Trans. Fuzzy Syst., vol. 1, pp. 98110, 1993.

[5] J. C. Bezdek, Pattern Recognition With Fuzzy Objective Function Algorithms, New York: Academic, 1981.

[6] L.-y. Gao, L. Sang, Y.-c. Hu, L.-l. Zhou, Research on Granular Computing Cased on Rough Set Theory and Its Application, Control and Automation Publication Group, vol.24(12-3),pp.189191, 2008.

[7] H. Li,Research on Knowledge Reduction based on Knowledge Granularity vol.25(2), pp.1619, 2010.

[8] L. Sun, J. C. Xu, S. Q. Li, X. Z. Cao, and Y. P. Gao, New Approach for Feature Selection by Using Information Entropy, Journal of Information and Computational Science, vol.8, pp.2259-2268, 2011.

[9] S. Ding, H. Huang, J. Yu, Artificial Intelligence Review, Research on the hybrid models of granular computing and support vector machine, vol.43(4), pp.565-577, 2015.

[10] L. Sun, J. C. Xu, and Y. Tian, Feature Selection Using Rough Entropy-Based Uncertainty Measures in Incomplete Decision Systems, Knowledge Based Systems, vol.36, pp.206-216, 2012.

[11] Q. H. Hu, J. F. Liu, and D. R. Yu, Mixed Feature Selection Based on Granulation and Approximation, Knowledge-Based System, vol.21, pp.294-304, 2008.

[12] Y. Qian, Y. Li, J. Liang, Fuzzy Granular Structure Distance, IEEE Trans. on Fuzzy Systems , vol.23(6), pp. 2245-2259,2015.

[13] J. Qian, L. Ping, X. Yue, C. Liu, Hierarchical attribute reduction algorithms for big data using Map Reduce, Knowledge-based Systems, vol.73, pp.18-31, 2015

[14] W. Pedrycz, From fuzzy data analysis and fuzzy regression to granular fuzzy data analysis, Fuzzy Sets and Systems, vol.274, pp.1217,2015.

[15] L. Sun, J. Xu, Y. Hu, and L. Du Granular Space-Based Feature Selection and Its Applications, Journal of Software, vol. 8(4),pp.817-826,2013.

[16] B. M. Joshi, G.B. Jethava, H. B. Bhavsar, High Dimensional Unsupervised Clustering Based Feature Selection Algorithm, International Journal of Engineering Science and Technology (IJEST), vol.4(5), pp.2022-2029,2012.

[17] H. Runxin and H. Nian, The Reduction of Facial Feature Based on Granular Computing, Electronics and Signal Processing, LNEE 97, pp. 10151021, 2011.