Rough Set Based FCM Algorithm for Image Segmentation

Zhenghao Shi\textsuperscript{1*}, Yuyan Chao\textsuperscript{2}, Lifeng He\textsuperscript{3}, Tsuyoshi Nakamura\textsuperscript{1} and Hidenori Itoh\textsuperscript{1}

\textsuperscript{1} School of Computer Science and Engineering, Nagoya Institute of Technology, 464-8555, Japan
\{zhshi, tnaka\}@juno.ics.nitech.ac.jp
\textsuperscript{2} Faculty of Environment and Information Management, Nagoya Sangyo University,
Owariasahi-Shi, Aichi 488-8711, Japan
chao@nagoya-su.ac.jp
\textsuperscript{3} Department of Radiology, The University of Chicago, 5841 South Maryland Avenue,
MC 2026, Chicago, IL 60637, USA
helifeng@uchicago.edu

Abstract. In this paper, a modified FCM (Fuzzy C-Means) algorithm based on Rough Set for image segment is proposed. The processes of the approach include two stages. In first stage, a cluster center set is built by reduction theory (the core of Rough Sets). In this stage, a decision table is designed firstly, where an initial cluster center set which contains granules with ill-defined boundaries is treated as an object, and each granule is treated as the object’s attribute. Discernibility of the initial cluster center sets in terms of attributes is then computed in the form of a discernibility matrix. Then using rough set theory, a number of decision rules are generated from the discernibility matrix. The rules represent rough clusters of points in the original feature space. In second stage, the eliminated cluster center set is input to FCM as the initial cluster center for the soft evaluation of the segmentation, where the fuzzy membership functions is modeled by a rule and utilized to compute the similarity of each image pixel. Experimental results demonstrate the efficiency and the effectiveness of the proposed method.

Keywords: Rough Set, FCM (Fuzzy C-Means), Image Segmentation.
1 Introduction

Image segmentation is an important step in image analysis. It has seen an explosion of interest over the last few years, and many methods have been proposed [1], [2]. Among them, the fuzzy c-means (FCM) clustering [3] method has been paid more attention [4], and such a success chiefly attributes to the introduction of fuzziness for the belongingness of each image pixel [5]. This allows for the ability to make the clustering methods able to retain more information from the original image than the crisp or hard segmentation methods and function well on segmenting most noise-free images. However, if the initial clustering of the FCM data set is unsuitable chosen, the FCM algorithm will get trapped in local minima, and fails to segment images [6]. In addition, its clustering number need determined in advance, and its calculated amount is also huge. In order to overcome drawbacks mentioned above, many modified FCM methods have been proposed [3]. However, for there exists many vagueness and uncertainty inherent in building clustering set, most methods are lacking of instructiveness.

Rough sets theory [7] proposed by Pawlak is a mathematical tool to analyze vagueness and uncertainty inherent in making decision. It doesn’t rely on accessional information out of data set, and it analyzes and discovers reliant relation among data just from the point of view of data’s discreditable attribute, just based on the concept of an upper and a lower approximation of a set, as well as the approximation space and models of sets. For the sake of this reason, it can provide more convenience for knowledge reduction than other methods and also superiority to other methods in analyzing vagueness and uncertainty inherent in making decision.

Based on analysis mentioned above, this paper introduces Rough Set theory into FCM image segmentation, by reduction theory (the core of Rough Sets), redundant initial cluster centers in the initial cluster set is eliminated, this is very useful for improving the convergence of the FCM algorithm. Experimental results demonstrate the superiority of the proposed method in image FCM segmentation.

The rest of this paper is organized as follows. In Section II, Rough Set theory is described. In Section III, a novel image FCM clustering segmentation algorithm based on Rough Sets is presented in detail. Then experimental results are discussed. In the last section we give some conclusions.

2 Rough Sets [7]

Let us present here some preliminaries of rough set theory which are relevant to this article. For details, one may refer to [7].

The basic idea of rough set theory is exporting decision or classified rule of some concept by knowledge reduction. A basic tool to represent and dispose knowledge is decision table, in which rows are labeled by objects of the universe and columns by the attributes. Let \( U = \{ x_1, x_2, \ldots, x_N \} \) denotes a finite set of \( N \) objects (named as universe), \( Q = C \cup D \) (where \( C \) is a set of condition at-
tributes, D is a set of decision attributes) denotes a finite set of attributes, f: U × Q → V denotes the information function (which designates the attribute value of each object x in U). Then a decision table knowledge representation system can be represented as

$$S=<U, Q, V, f>$$  \hfill (1)

This table may be unnecessarily large because it could be redundant at least in two ways. The same or indiscernible objects may be represented several times, or some attributes may be superfluous. The notion of equivalence relation is used to tackle this problem.

For each a ∈ A, if there exits f(x, a)=f(y, a), then IND(A)={(x, y) ∈ U is called as an equivalence relation IND(A) on U, and IND(A) is called A-indiscernibility relation. The partition induced by the equivalence relation IND(A) can be used to build new subsets of the universe.

It also determines the approximation space AS = (U, IND(A)) in set S. Assuming A ⊆ Q and X ⊆ U, lower/upper approximate set of X can be defined as

$$A^-(X) = \bigcup \{ Y | (Y \in U \land IND(A) \land Y \cap X \neq \emptyset) \}$$  \hfill (2)

$$A_-(X) = \bigcup \{ Y | (Y \in U \land IND(A) \land Y \subseteq X) \}$$  \hfill (3)

Here $A^-(X)$ and $A_-(X)$ constitute a dualistic group $(A^-(X), A_-(X))$, which is called rough set based on A. A positive domain of X is defined as POS$_A$(X) = A(X).

Based on what mentioned above, knowledge reduction can be implemented by identifying equivalence classes. This is core of Rough sets.

3 FCM Segmentation Based on Rough Set

3.1 The Conventional FCM Algorithm \cite{8}

Let $S = \{s(x)\}$, where $s(x)$ presents a given set of feature vectors associated with an image defined in domain I. The objective of FCM-algorithm is to minimize the Fuzzy C-Means cost function $J_{FCM}$ formulated as

$$J_{FCM}(U, V) = \sum_{x \in I} \sum_{k=1}^{c} u_{k,x}^m d_{k,x}^2 \quad \text{Subject to} \quad \sum_{k=1}^{c} u_{k,x} = 1 \quad \forall x \in I$$  \hfill (4)

Where $V = \{v_1, v_2, \ldots, v_c\}$ is a cluster centers set, $U = \{u_{k,x}\}$ is a fuzzy partition matrix, in which each $u_{k,x}$ indicates the degree of membership between the image pixel vector $s(x)$ and the $k$th cluster center $v_k$ $m \in (1, \infty)$ is the fuzzy exponent and usually is set to 2, $c$ is the total number of clusters. $d_{k,x}$ is distance norm, which indexes the vector distance between a feature vector and a cluster center in the feature space. Usually the Euclidean distance is used i.e.
Minimization of the cost function $J_{FCM}$ is a nonlinear optimization problem, which can be implemented with the following iterative process:

Step 1: Select appropriate value for $\alpha$, and a small positive number $\varepsilon$. Initialize the cluster centers $V$ randomly. Set step variable $t=0$.

Step 2: Calculate (at $t=0$) or update (at $t>0$) the membership matrix $U = \{u_{k,x}\}$ by

$$
u_{k,x}^{(t+1)} = \frac{1}{\sum_{h=1}^{c} \frac{d_{h,x}}{d_{k,x}}} \quad \text{for} \quad k = 1,\ldots,c \quad \text{and} \quad x \in I$$

Step 3: Update the cluster centers $V$ by

$$v_{k}^{(t+1)} = \frac{\sum_{x=1}^{m} \nu_{k,x}^{(t+1)} s(x)}{\sum_{x=1}^{m} \nu_{k,x}^{(t+1)}} \quad \text{for} \quad k = 1,\ldots,c$$

Step 4: Repeat step 2-3 until $\|V^{(t+1)} - V^t\| < \varepsilon$.

After FCM clustering, each data sample will be associated with a membership value for each class. By assigning the data sample to the class with the highest membership value, a segmentation of the data can be obtained.

### 3.2 Reducing FCM Initial Cluster Centers Set by Rough Sets

In the following section, the method proposed in this paper is described in detail.

#### 3.2.1 Obtain Each Feature’s membership value

First, initial cluster centers $\{P_1, P_2, \ldots, P_c\}$ were generated by randomly choosing $c$ points from an image point set. Where $c \in [c_{\min}, c_{\max}]$, $c_{\min} = 2$, $c_{\max} = \sqrt{n}$ (n is the image pixels number). Each cluster centers $P_i$ is represented by $n$ numeric image features $\{F_i, l = 1, 2, \ldots, n\}$. Then each feature $F_i$ is described in terms of its fuzzy membership values corresponding to three linguistic fuzzy sets, namely, low (L), medium (M), and high (H), which characterized respectively by a $\pi$-membership function[9].
\[
\mu(F_i) = \begin{cases} 
2 \left(1 - \frac{|F_i - c|}{\lambda}\right)^2 & \text{for } \frac{\lambda}{2} \leq |F_i - c| \leq \lambda \\
1 - 2 \left(\frac{|F_i - c|}{\lambda}\right)^2 & \text{for } 0 \leq |F_i - c| \leq \frac{\lambda}{2} \\
0 & \text{otherwise}
\end{cases}
\tag{8}
\]

Where \(\lambda\) is the radius of the \(\pi\)-membership function with \(c\) as the central point. To select the center \(c\) and radius \(\lambda\), the method described in [9] is used in this paper. Thus, we obtain an initial clustering centers set where each cluster center is represented by a collection of fuzzy set.

3.2.2 Constitute a Decision Table for the Initial Cluster Centers Set

Definition 1 Degree of similarity between two different cluster centers is defined as

\[\alpha = \frac{\sum_{i=1}^{n} \mu(F_i)}{n}\tag{9}\]

higher the value of the similarity, the closer the two clustering center is.

Definition 2 In a same cluster centers set, if a cluster center has a same similarity value to another one, then they are called redundant cluster center each other.

Proposition 1 If \(A\) and \(B\) are redundant cluster center each other, \(B\) and \(C\) are redundant cluster center each other, then \(A\), \(B\) and \(C\) belong to a same redundant cluster center, Viz.

\[A \leftrightarrow B, B \leftrightarrow C \Rightarrow A \leftrightarrow B \leftrightarrow C\tag{10}\]

Based on what mentioned above, taking initial cluster centers as objects, taking cluster centers features \(F_i\), the central point \(c\) and the radius \(\lambda\) as conditional attributes, taking degree of similarity between two different cluster centers as decision attribute by computing the \(\pi\)-membership function value, then a decision table for the initial cluster centers set can be constituted as follows:

\[T = <U, P \cup R, C, D>\tag{11}\]

Where \(U = \{x_i, i=1, 2...m\}\), it denotes a initial cluster centers set; \(P \cup R\) is a finite set of the initial cluster center category attributes (where \(P\) is a set of condition attributes, \(R\) is a set of decision attributes); \(C = \{p_i, i=1, 2...n\}\) (where \(p_i\) is a domain of the initial cluster center category attribute);
$D: U \times P \cup R \rightarrow C$ is the redundant information mapping function, which defines an indiscernibility relation on $U$.

3.2.3 Eliminating redundant cluster centers from the initial cluster centers set

Assuming $D(x)$ denotes a decision rule, $D(x)\mid P(\text{condition})$ and $D(x)\mid R(\text{decision})$ denote the restriction that $D(x)$ to $P$ and $R$ respectively, $i$ and $j$ denotes two different cluster centers respectively, and other assumptions are as the same as what mentioned above. Based on what described above, the initial cluster centers set can be optimized by reduction theory according to the following steps:

1. deducing the compatibility of each rule of an initial cluster center set decision table
   - If $D(l)\mid P(\text{condition}) = D(j)\mid P(\text{condition})$ and $D(i)\mid R(\text{decision}) = D(j)\mid R(\text{decision})$, then rules of an initial cluster center set decision table are compatible;
   - If $D(i)\mid P(\text{condition}) = D(j)\mid P(\text{condition})$, but $D(i)\mid R(\text{decision}) \neq D(j)\mid R(\text{decision})$, then rules of an initial cluster center set decision table are not compatible.

2. Ascertaining redundant conditional attributes of an initial cluster center set decision table
   If an initial cluster center set decision table are compatible, then when $p \in P$ and $\text{Ind}(P) = \text{Ind}(P-p)$, $p$ is a redundant attribute and it can be leaved out, otherwise $p$ can’t be leaved out.
   If an initial cluster center set decision table are not compatible, then computing its positive region $\text{POS}(P, R)$. If $p \in P$, when $\text{POS}(P, R) = \text{POS}(P-p, D)$, then $p$ can be leaved out, otherwise $p$ can’t be leaved out.

3. Eliminating redundant decision items from an initial cluster center decision table. For each condition attribute $p$ carries out the process mentioned above until condition attribute set does not change.

As soon as redundant initial cluster centers in the initial cluster set is eliminated, a reduced cluster center set is used as the FCM initial input variance for image segmentation.

To evaluate the quality of clusters, the Xie-Beni index [10] was used as the cluster validity index in this paper. The Xie-Beni index is expressed as:

$$XB = \frac{\sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^2 \| x_i - v_j \|^2}{\min_{j} \| v_i - v_j \|}$$

(12)

A smaller XB reflects that the clusters have greater separation from each other and are more compact.

Based on what descript above, now the procedure for Rough Sets based FCM image segmentation method can be summered as follows:

Step1 Randomly initialize the number of clusters to $c$, where $2 \leq c \leq \sqrt{n}$ and $n$ is number of image points.
Step2 Randomly chooses c pixels from the image data set to be cluster centers.
Step3 Optimize the initial cluster centers set by Rough Sets.
Step4 Set step variable t=0, and a small positive number $\varepsilon$.
Step5 Calculate (at t=0) or update (at t>0) the membership matrix $U = \{u_{k,i}\}$ using equation (6).
Step6 Update the cluster centers by equation (7).
Step7 Compute the corresponding Xie-Beni index using equation (12).
Step8 Repeat step 5-8 until $\|XB^{(t+1)} - XB^t\| < \varepsilon$.
Step9 Return the best XB and best center positions.

4 Experiment Results

In this section, experimental results on real images are described in detail. There are totally six algorithms used in these experiments, i.e., standard FCM, FCM_S[11], KFCM_S[12], IFCM[13], PFCM[14] and RFCM (algorithm propose in this paper). In these experiments, the number of different types of object elements in each image from manual analysis was considered as the number of clusters to be referenced. They were also used as the parameter for FCM. The Xie-Beni index value has been utilized throughout to evaluate the quality of the classification for all algorithms. All experiments were implemented on PC with 1.8GHz Pentium IV processor using MATLAB (version 6.5).

The first experiment is to test the influence of initial clustering centers on the convergence performance of FCM on the standard Lena image (Figs.1). Four algorithms were referenced here: FCM, IFCM, PFCM and RFCM. The number of classes in these segmentations was set to three, which corresponds to the woman Lena, adorning in her head and background pixels.

Figs.1. (a) shows the initial clustering centers produced by four different methods, i.e., random initialize, immunity initialize, polynomial initialize and rough sets initialize. In the figure, the initial clustering centers produced by the four methods are tagged respectively, i.e. “□” for random initialize, “○” for immunity initialize, “△” for polynomial initialize and “φ” for rough sets initialize. “×” denotes the output clustering centers of FCM. Figs.1. (b) shows the evolutions of the object function for FCM when different initial clustering centers methods are adopted.

From Figs.1 (a), it can be seen that the initial clustering center produced by rough sets initialize method is the most closest one to the last output of the FCM. Output of the immunity initialize method is followed. Initial clustering center produced by random method is the most furthest one to the last output of the FCM.

From Figs.1 (b), it can be seen that RFCM has the fastest convergence speed; it can converged almost in the fourth iteration. IFCM also has good convergence performance, but it converged about in the seventh iteration. As a compared, the conventional FCM has the most poor convergence performance. It converged almost in the twenty-eight iteration.
Fig. 1. The influence of initial clustering centers on the convergence performance of FCM
(a) The initial clustering centers produced by four different methods, i.e., random initialize, immunity initialize, polynomial initialize and rough sets initialize. (b) The evolutions of the object function for FCM when different initial clustering centers methods are adopted.

Figs. 2 shows the image segmentation results by the above four algorithms respectively. The margins of Lena’s hat in images segmented by FCM and PFCM are both a little blurred, while the ones obtained by IFCM and RFCM both have clear margin. Especially the one segmented by RFCM is more clearly compared to others.

Fig. 2. Comparison of segmentation results on standard lean image
(a) Original standard lean image. (b) FCM. (c) PFCM. (d) IFCM. (e) RFCM.

The second experiment is tested on a MR image (Figs. 3) obtained from the BrainWeb Simulated Brain Database [15]. The number of tissue classes in the segmentation was set to three, which corresponds to gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Background pixels are ignored in the computation. Here four algorithms are referenced, i.e., FCM, FCM_S, KFCM_S, and RFCM.

From Figs. 3, it can be seen that the margins of the segmented image about GM and WM by FCM is blurred, while the one obtained by FCM_S, KFCM_S, and RFCM achieve much better segmentation compared to it and the results of the last two ones are as almost the same, and also a little superior to the former two algorithms.
Fig. 3. Comparison of segmentation results on a simulated brain MR image  
(a) Original MR brain image. (b) FCM. (c) FCM_S. (d) KFCM_S. (e) RFCM.

The third experiment is done on the image of “cameraman” with 3% “Gaussian” noise provided by image tools of MATLAB (figs.4). The number of object classes in the segmentation was set to two, which corresponds to cameraman, and Background pixels. Here all the six algorithms mentioned above are referenced.

Fig. 4. Comparison of segmentation results on an image with 3% “Gaussian” noise  
(a) Original noise image. (b) FCM. (c) PFCM. (d) IFCM. (e) FCM_S. (f) KFCM_S. (g) RFCM.

From Figs.4, it can be seen that KFCM_S and RFCM achieved better segmentation results under noises, while the other algorithms fail to remove the effect of added noises. Especially the first three ones almost do anything to noises.

Table 1 shows the average XB index values and its corresponding clustering results obtained when using FCM, FCM_S, KFCM_S, IFCM, PFCM and RFCM for image segmentation respectively.

From Table 1, it can be seen that in all experiments, the average XB values of the solutions found by KFCM_S and RFCM both are smaller than other FCM methods. This means that the clusters obtained by KFCM_S and RFCM both have better XB index values than the other approaches. These also indicate that KFCM_S and RFCM both are able to escape sub-optimal solutions better than the other methods.

Finally, Table 2 gives the comparison of the running time on the last two experiments. From Table 2, it can be seen that RFCM, IFCM and PFCM are much faster than the standard FCM.
(typically 2 times faster) and FCM_S (typically 2.6 times faster) and KFCM_S (typically 1.7 times faster), and in both experiments, RFCM is the fastest.

Table 1. Average of the XB index value and it corresponding clustering results obtained when using FCM, FCM_S, KFCM_S, IFCM, PFCM and RFCM for image segmentation respectively

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Average of the XB index values</th>
<th>Clustering Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCM</td>
<td>0.034024</td>
<td>3</td>
</tr>
<tr>
<td>PFCM</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IFCM</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FCM_S</td>
<td>0.031689</td>
<td>3</td>
</tr>
<tr>
<td>KFCM_S</td>
<td>0.031574</td>
<td>3</td>
</tr>
<tr>
<td>RFCM</td>
<td>0.031578</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2. Comparison of Clustering time (sec)

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCM</td>
<td>13.64</td>
<td>12.52</td>
</tr>
<tr>
<td>PFCM</td>
<td>-</td>
<td>6.65</td>
</tr>
<tr>
<td>IFCM</td>
<td>-</td>
<td>6.54</td>
</tr>
<tr>
<td>FCM_S</td>
<td>16.92</td>
<td>15.86</td>
</tr>
<tr>
<td>KFCM_S</td>
<td>11.20</td>
<td>9.78</td>
</tr>
<tr>
<td>RFCM</td>
<td>6.48</td>
<td>6.44</td>
</tr>
</tbody>
</table>

5 Conclusions

We employed Rough Sets to FCM image segmentation. By reduction theory (the core of Rough Sets), the vagueness and uncertainty information inherent in a given initial cluster center set is analyzed, and those redundant initial cluster centers in the initial cluster set is then eliminated, the reduced initial cluster center set as input to FCM for the soft evaluation of the segments, this is very useful for overcoming the drawbacks of conventional FCM segmentation over-dependence on initial value. To evaluate the quality of clusters, the Xie-Beni index was used as the cluster validity index. Experimental results indicate the superiority of the proposed method in image segmentation.
References