Remote Sensing Image Classification based on Clustering Algorithms

B.Sai Chandana*, M.L.Phanendra*, Samuel Wesley Babu*, Dr.K.Srinivas*, Dr.R.Kiran Kumar@

*Assistant Professor, Department of CSE, GIT GITAM University, Visakhapatnam.
*PG Student, Department of IT, GIT, GITAM University, Visakhapatnam
*Professor, VR Siddhartha Engineering College, Vijayawada.
*Assistant professor, Department of CSE, Krishna University, Machilipatnam.
bschandana@gmail.com

Abstract: Clustering is an unsupervised classification method widely used for classification of remote sensing images. As the spatial resolution of remote sensing images getting higher and higher, the complex structure is the simple objects becomes obvious, which makes the classification algorithm based on pixels being losing their advantages. In this paper, four different clustering algorithms such as K-means, Moving K-means, Fuzzy K-means and Fuzzy Moving K-means are used for classification of remote sensing images. The Fuzzy Moving K-means clustering algorithm avoids the problems such as, the occurrence of dead centers, center redundancy and trapped center at local minima. The final center values obtained are located with their respective group of data. The experimental results show that Fuzzy Moving K-means has classified the remote sensing image more accurately than other three algorithms.

Keywords: Image Clustering, Remote Sensing, Image Processing.

1. INTRODUCTION

Remote sensing can be defined as any process whereby information is gathered about an object, area or phenomenon without physical contact with the object [1]. The remote sensing technology (aerial sensor technology) is used to classify objects on the Earth (both on the surface, and in the atmosphere and oceans) by means of propagated signals. New opportunities to use remote sensing data have arisen, with the increase of spatial and spectral resolution of recently launched satellites. Remote sensing image classification is a key technology in remote sensing applications [2]. Rapid and high accuracy remote sensing image classification algorithm is the precondition of kinds of practical applications. In remote sensing, sensors are available that can generate multispectral data, involving five to more than hundred bands.

At present, there is different image classification methods used for different purposes by various researches. These techniques are distinguished in two main ways as supervised and unsupervised classifications [3]. Supervised classification has different sub-classification methods which are named as parallelepiped, maximum likelihood, minimum distances and Fisher classifier methods. Unsupervised classification has evolved in two basic strategies [4], Iterative and Sequential. In an iterative procedure such as K-Means or ISODATA, an initial number of desired clusters are selected, and the centroid locations are then moved around until a statistically optimal fit is obtained. In a sequential algorithm such as Classification by Progressive Generalization, the large number of spectral combinations is gradually reduced through a series of steps using various proximity measures [5][10].

In this paper, we used three versions of conventional k-means clustering algorithm for classification of remote sensing image. The qualitative and quantitative results show that Fuzzy Moving K-means clustering algorithm has classified the image better than other clustering algorithms. The paper is organized as follows: Section 2 presents the K-means clustering algorithm, Section 3 presents Moving K-means clustering algorithm, Section 4 presents Fuzzy K-means clustering algorithm, Section 5 presents Fuzzy Moving K-means clustering algorithm.
Clustering algorithm, Section 6 presents Experimental results and finally Section 7 report conclusions.

2. K-MEANS CLUSTERING ALGORITHM

K-means is one of the basic clustering methods introduced by Hartigan [6]. This method is applied to segment the remote sensing image in recent years. The K-means clustering algorithm for classification of remote sensing image is summarized as follows:

Algorithm K-means(x,n,c)

Input:
N: number of pixels to be clustered;
x={x₁,x₂,x₃,……,x₈} : pixels of remote sensing image
c ={c₁,c₂,c₃,……,cₖ} : clusters respectively.

Output:
cl: cluster of pixels

Begin

Step 1: cluster centroids are initialized.

Step 2: compute the closest cluster for each pixel and classify it to that cluster, ie: the objective is to minimize the sum of squares of the distances given by the following:

\[ \Delta_0 = \| x_i - c_j \| \quad \text{arg min} \sum_{i=1}^{N} \sum_{j=1}^{C} \Delta_j^2 \]  \hspace{1cm} (1)

Step 3: Compute new centroids after all the pixels are clustered. The new centroids of a cluster is calculated by the following

\[ c_j = \frac{1}{N_j} \sum_{x_i \text{ belongs to } c_j} x_i \]  \hspace{1cm} (2)

Step 4: Repeat steps 2-3 till the sum of squares given in equation is minimized.

End.

The K-means clustering algorithm has many weaknesses which are as follows:

1. The number of clusters must be determined before the algorithm is executed.
2. The algorithm is sensitive to initial conditions. It produces different results for different initial conditions.
3. The K-means algorithm may be trapped in the local optimum. As a result, the trapped clusters would represent wrong group of data.
4. Data which are far away from the centers may pull the centers away from the optimum location, leading to poor representation of data.

3. MOVING K-MEANS CLUSTERING ALGORITHM

The Moving K-means clustering algorithm is the modified version of K-means proposed in [7]. It introduces the concept of fitness to ensure that each cluster should have a significant number of members and final fitness values before the new position of cluster is calculated. The Moving K-means clustering algorithm for classification of remote sensing image is summarized as follows:

Algorithm Moving K-means(x,n,c)

Input:
N: number of pixels to be clustered;
x={x₁,x₂,x₃,……,x₈} : pixels of remote sensing image.
c ={c₁,c₂,c₃,……,cₖ} : clusters respectively.

Output:
cl: cluster of pixels

Begin

Step 1: cluster centroids are initialized

Step 2: compute the closest cluster for each pixel and classify it to that cluster, ie: the objective is to minimize the sum of squares of the distances given by the following:

\[ \Delta_0 = \| x_i - c_j \| \quad \text{arg min} \sum_{i=1}^{N} \sum_{j=1}^{C} \Delta_j^2 \]  \hspace{1cm} (3)

Step 3: The fitness for each cluster is calculated using

\[ f(c_j) = \sum_{x_i \in c_j} (\| x_i - c_j \|)^2 \]  \hspace{1cm} (4)

All centers must satisfy the following condition:

\[ f(c_j) \geq \alpha s f(c_i) \]  \hspace{1cm} (5)

where \( \alpha_s \) is small constant value initially with value in range 0< \( \alpha_s \)<1/3, \( c_i \) and \( c_j \) are the centers that have the smallest and the largest fitness values. If (5) is not fulfilled, the members of \( c_i \) are assigned as members.
of \( c_{n} \), while the rest are maintained as the members of \( c_{1} \). The positions of \( c_{n} \) and \( c_{1} \) are recalculated according to:

\[
C_{i} = \frac{1}{n_{ci}} \left( \sum_{j \in c_{i}} x_{j} \right) \quad (6)
\]

\[
C_{i} = \frac{1}{n_{ci}} \left( \sum_{j \in c_{i}} x_{j} \right) \quad (7)
\]

The value of \( a_{w} \) is then updated according to:

\[
a_{w} = a_{w} - a_{w} / n_{c} \quad (8)
\]

The above process are repeated until (5) is fulfilled. Next all data are reassigned to their nearest center and the new center positions are recalculated using (2).

Step 4: The iteration process is repeated until the following condition is satisfied:

\[
f(c_{i}) \geq a_{w} \cdot f(c_{i}) \quad (9)
\]

End

The Moving K-means algorithm has the following drawbacks:

1. The Moving K-means algorithm is sensitive to noise.
2. For some cases of Moving k-means, the clusters or centers are not located in the middle or centroid of a group of data, leading to imprecise results.
3. The fitness concept in the Moving k-means algorithm lead to a problem where some members of centers with the largest fitness are enforced to be assigned as a members of a center with the smallest fitness.

4. **FUZZY K-MEANS CLUSTERING ALGORITHM**

The Fuzzy K-means [8] is an unsupervised clustering algorithm. The main idea of introducing fuzzy concept in the Fuzzy K-means algorithm is that an object can belong simultaneously to more than one class and does so by varying degrees called memberships. It distributes the membership values in a normalized fashion. It does not require prior knowledge about the data to be segmented. It can be used with any number of features and number of classes. The fuzzy K-means is an iterative method which tries to separate the set of data into a number of compact clusters. The Fuzzy K-means algorithm is summarized as follows:

**Algorithm Fuzzy K-Means (x,n,c,m)**

**Input:**
- N=number of pixels to be clustered;
- \( x = \{x1, x2, ..., xN\} \): pixels of remote sensing image;
- \( c = \{c1, c2, c3, ..., cN\} \): clusters respectively.
- \( m=2 \): the fuzziness parameter;

**Output:**
- \( u \): membership values of pixels and clustered Image

**Begin**

Step_1: Initialize the membership matrix \( u_{ij} \) is a value in (0,1) and the fuzziness parameter \( m (m=2) \). The sum of all membership values of a pixel belonging to clusters should satisfy the constraint expressed in the following.

\[
\sum_{j=1}^{c} u_{ij} = 1 \quad (10)
\]

for all \( i = 1, 2, ..., N \), where \( c (\approx 2) \) is the number of clusters and \( N \) is the number of pixels in remote sensing image.

Step_2: Compute the centroid values for each cluster \( c_{j} \). Each pixel should have a degree of membership to those designated clusters. So the goal is to find the membership values of pixels belonging to each cluster. The algorithm is an iterative optimization that minimizes the cost function defined as follows:

\[
F = \sum_{j=1}^{c} \sum_{i=1}^{N} u_{ij}^m \| x_{i} - c_{j} \|^2 \quad (11)
\]

where \( u_{ij} \) represents the membership of pixel \( x_{i} \) in the \( i \)th cluster and \( m \) is the fuzziness parameter.

Step_3: Compute the updated membership values \( u_{ij} \) belonging to clusters for each pixel and cluster centroids according to the given formula.

\[
u_{ij} = \frac{1}{\sum_{k=1}^{c} \left( \| x_{i} - v_{k} \|^2 \right)^{2(m-1)}}
\]

and

\[
v_{j} = \frac{\sum_{i=1}^{N} u_{ij}^{m} x_{i}}{\sum_{j=1}^{N} u_{ij}^{m}} \quad (12)
\]

Step_4: Repeat steps 2-3 until the cost function is minimized.

**End.**
5. Fuzzy Moving K-Means Clustering Algorithm

In the Fuzzy Moving K-means clustering algorithm [9], the membership function is used in addition to the Euclidian distance to control the assignment of the members to the proper center. The algorithm minimizes the sensitivity to the noisy data by updating the moving member function. It is not obligatory for the members of the center with the largest fitness value to follow the center with the smallest fitness value. The Fuzzy Moving K-means clustering algorithm is summarized as follows:

Input:
N: number of pixels to be clustered;
\( x = \{x_1, x_2, x_3, \ldots, x_N\} \) : pixels of remote sensing image
\( c = \{c_1, c_2, c_3, \ldots, c_c\} \) : clusters respectively
\( m=2 \): the fuzziness parameter;

Output:
\( u \): membership values of pixels and clustered Image

Begin
Step_1: Initialize the membership matrix \( u_{ij} \) is a value in (0,1) and the fuzziness parameter \( m \) (\( m=2 \)).
The sum of all membership values of a pixel belonging to clusters should satisfy the constraint expressed in the following.
\[
\sum_{j=1}^{c} u_{ij} = 1 \tag{13}
\]
for all \( i=1,2,\ldots, N \), where \( c (=2) \) is the number of clusters and \( N \) is the number of pixels in remote sensing image.

Step_2: Compute the centroid values for each cluster \( c_i \). Each pixel should have a degree of membership to those designated clusters. So the goal is to find the membership values of pixels belonging to each cluster. The algorithm is an iterative optimization that minimizes the cost function defined as follows:
\[
F = \sum_{j=1}^{N} \sum_{i=1}^{c} u_{ij}^m \| x_j - c_i \|^2 \tag{14}
\]
where \( u_{ij} \) represents the membership of pixel \( x_j \) in the \( i^{th} \) cluster and \( m \) is the fuzziness parameter.

Step 3: The fitness for each cluster is calculated using
\[
f(c_i) = \sum_{j=1}^{N} (\| x_j - c_i \|^2) \tag{15}
\]
All centers must satisfy the following condition:
\[
f(c_i) \geq \alpha_u f(c_i) \text{ and } m(c_{ak}) > m(c_{bk}) \tag{16}
\]
where \( \alpha_u \) is small constant value initially with value in range \( 0 < \alpha_u < 1 / 3 \), \( c_1 \) and \( c_i \) are the centers that have the smallest and the largest fitness values, \( m(c_{ak}) \) is the membership value of point \( k \) according to the smallest centre and \( m(c_{bk}) \) is the membership value of point \( k \) according to the largest centre. If (5) is not fulfilled, the members of \( c_i \) are assigned as members of \( c_a \), while the rest are maintained as the members of \( c_i \). The positions of \( c_i \) and \( c_a \) are recalculated according to:
\[
C_i = \frac{1}{n_i} (\sum_{c_i} x_j ) \tag{17}
\]
\[
C_a = \frac{1}{n_a} (\sum_{c_a} x_j ) \tag{18}
\]
The value of \( \alpha_u \) is then updated according to:
\[
\alpha_u = \alpha_u \cdot \alpha / n_c \tag{19}
\]
The above process are repeated until (5) is fulfilled. Next all data are reassigned to their nearest center and the new center positions are recalculated using (2).
Compute the updated membership values \( u_{ij} \) belonging to clusters for each pixel according to given formula
\[
u_{ij} = \frac{1}{\sum_{k=1}^{c} (\| x_j - c_k \|^2)^{(m-1)}} \tag{20}
\]
and
\[
 v_i = \frac{\sum_{j=1}^{N} u_{ij}^m x_j}{\sum_{j=1}^{N} u_{ij}^m} \tag{20}
\]
Step 4: The iteration process is repeated until the following condition is satisfied.
\[
f(c_i) \geq \alpha_u f(c_i) \text{ and } m(c_{ak}) > m(c_{bk}) \tag{21}
\]
6. EXPERIMENTAL RESULTS

Qualitative Analysis:
The proposed four clustering algorithms are performed on a remote sensing image that consists of a total of 38808 pixels. The classification results by the proposed algorithms are shown in figure 1. From the results, we can visualize that Fuzzy Moving K-means have classified the remote sensing image more accurately than other three algorithms. It avoids the problems such as, the occurrence of dead centers, center redundancy and trapped center at local minima. It is also less sensitive to initialization process of clustering value. The final center values obtained are located with their respective group of data, enabling the size and shape of object to be maintained and preserved.

Quantitative Analysis:
Quantitative analysis is a numerically oriented procedure to figure out the performance of algorithms without any human error. The Mean Square Error (MSE) is significant metric to validate the quality of image. It measures the square error between pixels of the original and the resultant images. The MSE is mathematically defined as

\begin{equation}
MSE = \frac{1}{N} \sum_{j=1}^{k} \sum_{i \in C_j} ||x_i - c_j||^2
\end{equation}

Where N is the total number of pixels in an image and \(x_i\) is the pixel which belongs to the \(j^{th}\) cluster. The lower difference between the resultant and the original image reflects that all the data in the region are located near to its centre. Table 1 shows the quantitative evaluations of four clustering algorithms. The results confirm that Fuzzy Moving K-means algorithm produces the lowest MSE value for classifying the remote sensing image.

<table>
<thead>
<tr>
<th>Method</th>
<th>MSE Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-means</td>
<td>282.781</td>
</tr>
<tr>
<td>Moving K-means</td>
<td>216.392</td>
</tr>
<tr>
<td>Fuzzy K-means</td>
<td>138.327</td>
</tr>
<tr>
<td>Fuzzy Moving K-means</td>
<td>96.322</td>
</tr>
</tbody>
</table>

Fig 1: classification results (4 Clusters)

7. Conclusion
This paper has presented four clustering algorithms namely K-means, Moving K-means, Fuzzy K-means and Fuzzy Moving K-means for the classification of remote sensing image. The qualitative and quantitative analysis done proved that Fuzzy Moving K-means has higher classification quality than other clustering algorithms. The occurrence of dead centers, center redundancy and trapped center at local minima problems can be avoided. The algorithm is also less sensitive to initialization process of clustering value.

References

[2] Xiaofang Liu, Xiaowen Li, Ying Zhang, Cunjian Yang, Wenbo Xu, Min Li, Huanmin Luo”, Remote Sensing Image Classification Using Function Weighted FCM Clustering Algorithm”, 2007 IEEE.


