Performance Analysis of Kmeans and Fuzzy-Cmeans Classification Algorithms for Satellite Images

Aniruddha G
First year – Information Technology
Sri Ramakrishna Institute of Technology
Coimbatore, India
aniruddharao741@gmail.com

R. Nagendran
Department of Information Technology
Sri Ramakrishna Institute of Technology
Coimbatore, India
rnagu127@gmail.com

Abstract: Satellite image classification has become a tedious task as more and more satellites are launched and more data is being obtained. There are two types of classification namely, supervised and unsupervised. The present study deals with unsupervised classification focusing on its application for classification of satellite image. Two heavily used algorithms, kmeans clustering and fuzzy-cmeans clustering was taken into consideration, studied and their efficacies were tested on satellite images of various cities around the world. The results were conclusive enough to infer that fcm is a better classification algorithm than kmeans.

I. INTRODUCTION

Clustering algorithms are used to study large amounts of data and classify them as per requirement. Over the years, multiple algorithms for clustering, each with their own specialties and shortcomings have originated. Two such methods, Fuzzy cmeans (fcm) and kmeans have been selected and their performance studied. The results were compared and tabulated. Fuzzy Cmeans is a method of clustering where a single point/piece of data is allowed to be grouped into multiple clusters based on the percentage the data has in common with the concerned cluster. This is achieved by dividing the training vectors in accordance with the membership function. Probabilistic constraints are used to enable memberships of the training vector across clusters that sum up to one i.e., different grades of training vector are shared by distinct clusters but not degree of typicality. Kmeans is a clustering algorithm where n observations are sorted into k rigid clusters. The data values of one cluster cannot be present in another cluster.

The paper is organised as follows section II discusses History and logic behind classification, Kmeansand FCM, Section III explains the differences in output quality of Kmeans and FCM. Section IV gives conclusion based on the experimental results.

II. LITERATURE STUDY

Satellite images contain a lot of geographical information and are very useful for professionals in many fields[1].Satellite and remote sensing images provide qualitative and quantitative information that can be used to simplify tasks and reduce study time[2]. Satellites collect a huge amount of data and this keeps increasing as more and more satellites are launched. Due to this data volumes are growing exponentially[3]. To get useful data from these hundreds of images, powerful classification algorithms are needed.

Image classification is the task of obtaining information classes from a given raster image with multiple bands. Satellite image classification is a technique of grouping pixels of images obtained through remote sensing satellites into meaningful classes known as clusters[4].

a) Need for satellite image classification

The data that can be obtained from satellite images can be utilized in multiple ways. Some of them include:

- Thematic map creation
- Assisting planners and engineers
- Locate minerals using spectral analysis
- Disaster management
- Planning agriculture
- Field surveys

b) Types of image classification:

There exist two distinct types of image classification namely, supervised and unsupervised. Kmeans and fuzzy-cmeans are two algorithms that are extensively used in unsupervised classification.

c) Kmeans

This is one of the methods of unsupervised learning that is effective for unlabelled data[5][6]. Groups are formed from the data with number of groups being represented by the variable k. The algorithm uses iterations to assign each data point to one and only one of the 'k' number of groups. By kmeans clustering we can get the following data:

1. Centroids of each of the k clusters.
2. Labels for the training data

The kmeans of any set of data is obtained through the following equation:

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (\|x_i - v_j\|)^2$$

Where,

'$||x_i - v_j||$' is the Euclidean distance between x_i and v_j.

'c_i' is the number of data points in i^{th} cluster.

'c' is the number of cluster centres.

d) Fuzzy-Cmeans

It is a method developed by Dunn[7] in 1973 and improved by Bezdek[8] in 1981.

It is based on the minimization of the following function:

$$Jm(U,V) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^{m} \| x_k - v_i \|^2$$

Where,

$$u_{ik} = \left(\sum_{i=1}^{c} \frac{\left(\| x_k - c_i \|^2 \right)^{\frac{1}{(m-1)}}}{\left(\| x_k - c_l \|^2 \right)^{\frac{1}{(m-1)}}} \right)^{-1}$$

$$v_i = \frac{\sum_{x=1}^{n} u_{ik} x_k}{\sum_{x=1}^{n} u_{ik}}$$

When there are uncertainties related to data, fuzzy clustering technique allows us to provide a mathematical framework in order to capture them. Each training vector belongs to each cluster with some degree of membership depending on its similarities to the cluster.

III. EXPERIMENTAL RESULTS

The performance of kmeans technique and FCM technique were analysed for satellite images of various cities suchas Beijing, Zurich, Bayannaoer, Quebec-city, Sun-city and New York. The test images of size 128× 128, 200 × 200 and 400×400were taken as input.

Table 1 shows the output results obtained through providing the satellite images as input to Kmeans algorithm. From Table 1 it is clear that PSNR is undesirably low for output using Kmeans clustering. Sample result images obtained from kmeans clustering algorithm are presented after table 1.

Table 2 shows the output results obtained through providing the satellite images as input to Fuzzy-Cmeans algorithm. From Table 2 it is clear that PSNR is considerably higher for output obtained using clusteringcompared to the output using Kmeans clustering. Sample result images obtained from fuzzy-cmeans clustering algorithm are presented after table 2.

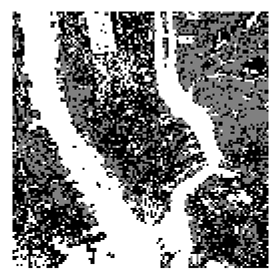
i) KMEANS ALGORITHM:

TABLE I

Index	Place	Size	PSNR	Max. error	MSE
1	Zurich, Switzerland	128×128	40.9999	3	5.1653
		200×200	41.5548	3	4.5457
		400×400	41.5644	3	4.5357
2	Bayannaoer, China	128×128	44.7147		2.1959
		200×200	44.8586	3	2.1243
		400×400	42.7634	3	3.4415
3	Beijing, China	128×128	43.2654	3	3.0658
		200×200	43.0694	3	3.2073
		400×400	42.7102	3	3.4839
4	Quebec city, Canada	128×128	40.6622	3	5.5829
		200×200	44.5825	3	2.2638
		400×400	41.6406	3	4.4567
5	Sun city, Arizona, USA	128×128	42.1062	3	4.0037
		200×200	43.8109	3	2.7039
		400×400	43.1227	3	3.1682
6	New York City, USA	128×128	42.5464	3	3.6178
		200×200	44.5069	3	2.3035
		400×400	43.6979	3	2.7752

Size Input Output

128×128



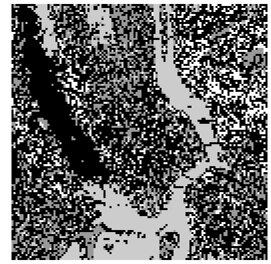
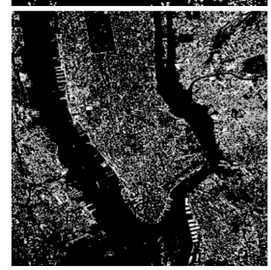
200×200

400×400		

ii) FUZZY-CMEANS ALGORITHM:

TABLE II

Index	Place	Size	PSNR	Max. error	MSE
1	Zurich, Switzerland	128×128	54.6205	1	0.2244
		200×200	54.9210	1	0.2094
		400×400	58.7574	1	0.0866
2	Bayannaoer, China	128×128	53.9966	1	0.2591
		200×200	53.6568	1	0.2802
		400×400	53.8801	1	0.2661
3	Beijing, China	128×128	56.1669	1	0.1572
		200×200	55.2266	1	0.1952
		400×400	55.1201	1	0.2000
4	Quebec city, Canada	128×128	54.8508	1	0.2128
		200×200	56.7710	1	0.1368
		400×400	55.8130	1	0.1705
5	Sun city, Arizona, USA	128×128	55.2696	1	0.1933
		200×200	55.8937	1	0.1674
		400×400	53.0486	1	0.3223
6	New York City, USA	128×128	55.1859	1	0.1970
		200×200	58.3182	1	0.0958
		400×400	59.9476	1	0.0658

Size	Input	Output
128×128		
200×200		
400×400		

IV. CONCLUSION

From the study it was found that Fuzzy-Cmeans gave far better results when compared to kmeans algorithm of classification. This is evident from the fact that Peak signal to noise ratio is higher, maximum error is lower and Mean Square error is quite reduced in the output images from Fuzzy-Cmeans clustering when compared to the results from Kmeans clustering algorithm.

REFERENCES

[1] S. Muhammed, G. Aziz, N. Aneela and S. Muhammed, Classification by Object Recognition in SatelliteImages by using Data Mining, vol. 1, London: World Congress on Engineering, 2012.

[2] V. Chaichoke, P. Supawee, V. Tansak and K. S. Andrew, A Normalized Difference Vegetation Index (NDVI) Time-Series of Idle Agriculture Lands: A Preliminary Study, vol. 15, Engineering Journal, 2011, pp. 9-16.

[3] X. Zheng, X. Sun, K. Fu and W. Hongqi, Automatic Annotation of Satellite Images via Multifeature Joint Sparse Coding With Spatial Relation Constraint, vol. 10, 2013, pp. 652-656.

[4] A. Karlsson, Classification of high resolution satellite images, 2003.

[5] A. Naik, "Google sites," . [Online]. Available: https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm.

[6] Tan, Steinbach, Kumar and Ghosh, "louisiana.edu," [Online]. Available: http://www.ucs.louisiana.edu/~xxw8007/kdd/PPT/Kmeans-ICDM06.pdf.

[7] J. C. Dunn, A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", vol. 3, Journal of Cybernetics, 1973, pp. 32-57.

[8] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algoritms, New York: Plenum Press, 1981