

A Novel Adaptive Clustering Algorithm Using Validity Index in Remote Sensing Data

Dalmiya C.P¹ and Santhi N²

¹Research Scholar, ²Associate Professor

^{1,2}Department of Electronics and Communication Engineering,
Noorul Islam Centre for Higher Education, Kumaracoil, Thuckalay
Kanyakumari dist., Tamil Nadu, 629180, India.

¹Orcid Id: 0000-0002-6619-5977, ²Orcid Id: 0000-0001-8034-242X

Abstract

Clustering refers to a non-predefined factor that is subject to concern more on unsupervised classification methods. Clustering technique in remote sensing images without human interference helps in various applications like change detection, land cover information, classification etc. In this paper a new adaptive clustering algorithm called Xie-Beni based Adaptive Fuzzy C Means Clustering (XB-AFCM) method is proposed and the validity indices are measured. The number of clusters is the major task in remote sensing applications that can be determined by cluster validity indices (CVIs). In recent years more number of CVIs is proposed in different papers and some common CVIs are evaluated in this paper. The proposed work is tested on 5 experimental datasets. Finally the calculations revealed that the proposed algorithm is best suited for accurate clustering of remote sensing images for unsupervised change detection methods.

Keywords: clustering, cluster validity index, xie-beni index, remote sensing, fuzzy C means.

INTRODUCTION

Clustering aims at partitioning the data set into significant groups. Clustering deals with data mining process and for identifying interesting distributions and patterns [1]. Clustering is based on a non-predefined criterion on the other hand classification is based on predefined criteria. Based on the literature there are different types of clustering methods such as K means, Fuzzy C means, Constrained k means [2], Hierarchical clustering [20,21,22] which produces accurate clustering results. As we know that clustering method is very sensitive to initializations, Fuzzy C Means Clustering [6] is one of the best methods for unsupervised change detection methods. The main concern in clustering is to decide accurate number of clusters. To evaluate the clustering process and for finding optimal number of clusters, Cluster Validity Indices (CVIs) are calculated [3]. A better clustering algorithm and validation measures are important for grouping a data [5]. For high-resolution remote sensing data it is difficult to visualize the data, so a better clustering analysis [17] with optimal

number of cluster is important. If the clustering parameters assigned an improper value, it may lead to wrong decisions [8]. Two criteria such as compactness and separation are proposed for a better clustering scheme [1,8]. Compactness indicates the variance of the cluster. The smaller the variance, the clustering technique is better. Separation indicates the distance between the cluster elements and if the distance is large the method is more isolated. In [4] 16 commonly used CVIs discussed are applicable to remote sensing data sets. In [4, 7, 10, 13, and 15] Xie-Beni index is proposed which is called compactness and separation validity function. It is used for evaluating Fuzzy C means clustering technique. Other validating indices used in Fuzzy clustering algorithms are I index [9], Fukuyama and Sugeno Index [7, 11], Partition Coefficient [7, 12, 13], Partition Entropy Coefficient [13], Fuzzy hyper volume [14], Partition Density [1, 14] etc. K means clustering is a partitioning algorithm, which decomposes the data set into group of disjoint clusters. Partitioning Around Medoids (PAM), CLARA (Clustering Large Applications) and CLARANS (Clustering Large Applications based on Randomized Search) [1]. The mostly used CVIs for K means clustering are Root Mean Square Standard Deviation (RMSSTD), Semi-partial R-squared (SPR) and R-Squared (RS) [1]. These indices can also be applied to Hierarchical clustering techniques. Moreover other algorithms such as BIRCH [16], CURE [18] and ROCK [19] use hierarchical structure for partitioning the data points.

PROPOSED CLUSTERING ALGORITHM

Xie-Beni based Adaptive Fuzzy C Means Clustering (XB-AFCM)

Fuzzy C means clustering is developed by Dunn in 1973 and the method is improved by Bezdek in 1981. This is based on minimization of the objective function J_f that can be calculated by the eqn. (2). From the analysis of validity indices it is noted that Xie-Beni index performs well in most of the images.

Let $S = \{s_1, s_2, s_3 \dots s_n\}$ be the set of data points and $T = \{t_1, t_2, t_3, \dots, t_c\}$ be the set of centers.

Step 1: Randomly select 'c' cluster centers.

Step 2: Calculate the fuzzy membership 'U_{ij}' using eqn. (1)

$$U_{ij} = \left[\sum_{j=1}^c \left(\frac{\|s_i - t_j\|^2}{\|s_i - t_c\|^2} \right)^{1/(m-1)} \right]^{-1} \quad (1)$$

Step 3: Compute the fuzzy centers 't_j'

Step 4: Repeat step 2) and 3) until the minimum 'J_f' value is achieved

$$J_f = \sum_{i=1}^N \sum_{j=1}^c (U_{ij})^f \|s_i - t_j\|^2 \quad (2)$$

Step 5: Calculate the Xie-Beni validity index, which is given in eqn. (5).

Step 6: Stop the iteration when the validity index reaches the minimum value.

CVIs for Fuzzy C Means Clustering

Partition Coefficient (PCoeff)

$$PCoeff = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{n_v} \|U_{ij}^2\| \quad (3)$$

PCoeff values are in the range [1/n_v, 1]. The value should be closer to unity for a better clustering technique.

Where n_v-number of clusters

U_{ij} – Membership function

Partition Entropy Coefficient (PECoeff)

$$PECoeff = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{n_v} U_{ij} \log_a(U_{ij}) \quad (4)$$

where a is the base of the logarithm. Partition entropy coefficient values ranges from 0 to log_a n_v. The value should be closer to 0 for a good clustering method.

Xie-Beni index (XBindex)

Xie-Beni index [10] is best-suited validity index in Fuzzy C means clustering. It works well with compact clusters.

$$XBindex = \frac{\sum_{i=1}^N \sum_{j=1}^N \mu_{ij}^2 \|a_i - b_j\|^2}{N \min_{j \neq N} \{\|b_j - b_n\|^2\}} \quad (5)$$

Fukuyama-Sugeno index (FSindex)

For well-separated clusters FS index value will be minimum. It measures the compactness and distance between the clusters.

$$FSindex = \sum_{i=1}^N \sum_{j=1}^{n_v} U_{ij}^f (\|a_i - b_j\|_X^2 - \|b_j - b\|_X^2) \quad (6)$$

In FS_{index} b is the mean vector of A and X is a symmetric matrix. For a well-separated cluster the FS_{index} should be as small as possible.

Total Fuzzy Hyper Volume (TFHV)

Total Fuzzy Hyper Volume is given by

$$TFHV = \sum_{i=1}^{n_v} V_i \quad (7)$$

$$V_i = |\Sigma_i|^{1/2} \quad (8)$$

$$\Sigma_i = \frac{\sum_{j=1}^N U_{ij}^f (a_i - b_j)(a_i - b_j)^T}{\sum_{j=1}^M U_{ij}^f} \quad (9)$$

Where b_j is the centre of cluster j.

RESULTS AND DISCUSSIONS

In order to evaluate the effectiveness of proposed method; experiments are carried out in 5 different types of satellite data. There are many validity indices available for evaluating the clustering techniques, but it is evident from the statistical assessment that all the validity indices does not work accurately in different remote sensing data. The proposed method categorizes the data into accurate number of clusters. For evaluating the correctness of result, a manual evaluation of validity indices is performed for the next cluster from which the clustering operation is stopped. For the analysis of validity indices, a dataset consists of five types of satellite data are taken such as GeoEye-1 (Image 1), Quick bird (Image 2), Worldview-2 (Image 3), IKONOS (Image 4) and Worldview-3 (Image 5). The details of dataset and specifications are given in Table1. The clustering result of each image is given in Figure 1,2,3,4 and 5.

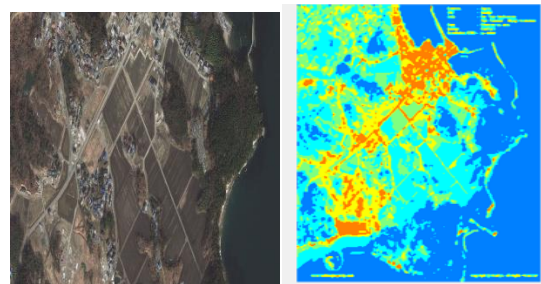


Figure 1.a GeoEye-1 b. Clustering results

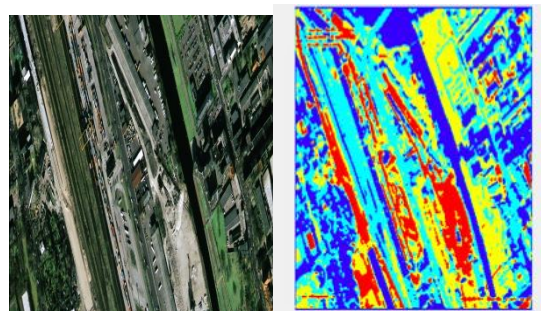


Figure 2.a Quick bird b. Clustering results

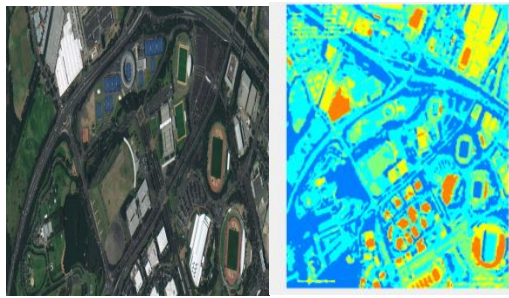


Figure 3.a Worldview 2 b. Clustering results

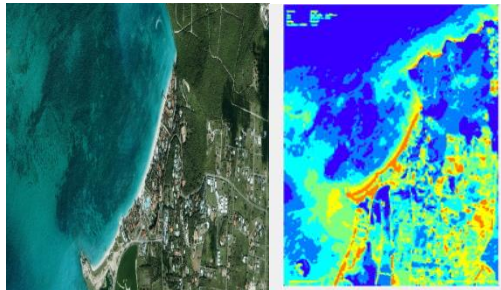


Figure 4.a IKONOS b. Clustering results

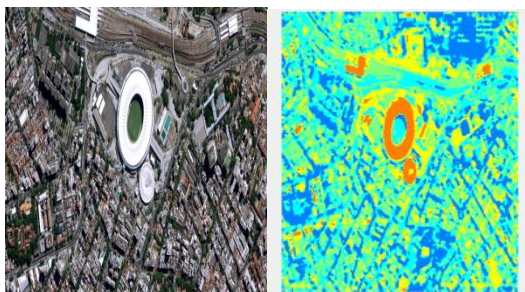


Figure 5 a.Worldview-3 b. Clustering results

Performance Evaluation

In fact cluster validity indices are certain measures to evaluate the number of clusters in an image. In the proposed XB-AFCM algorithm, the best cluster is formed which includes a correct classification of areas in each image. In Figure 6-10, the cluster evaluation of various datasets for the proposed algorithm is illustrated.

Table 2: CVI evaluation

Dataset	No. of clusters	PCoeff	PECoeff	FSindex(-) X 10 ⁻⁴	TFHV X 10 ⁻⁴	Proposed
1	2	0.87	0.21	0.55	0.66	0.14
	3	0.82	0.31	0.35	0.61	0.13
	4	0.80	0.37	0.72	0.43	0.124
	5	0.78	0.41	0.8	0.40	0.121
2	6	0.76	0.46	0.79	0.37	0.124
	2	0.83	0.27	0.51	1.07	0.16
	3	0.78	0.39	0.53	0.87	0.14
	4	0.77	0.44	0.68	0.72	0.12
3	5	0.74	0.51	0.67	0.65	0.13
	2	0.89	0.17	0.93	0.83	0.13
	3	0.79	0.36	0.94	0.75	0.18
	4	0.77	0.41	0.94	0.61	0.19
4	5	0.77	0.44	0.99	0.53	0.14
	6	0.74	0.50	0.89	0.47	0.15
	2	0.83	0.26	0.36	0.69	0.16
	3	0.78	0.38	0.42	0.57	0.146
5	4	0.76	0.44	0.58	0.51	0.137
	5	0.75	0.48	0.59	0.45	0.13
	6	0.73	0.54	0.63	0.43	0.12
	7	0.72	0.55	0.63	0.39	0.13
6	2	0.80	0.31	0.36	1.08	0.18
	3	0.77	0.39	0.43	0.8	0.14
	4	0.75	0.46	0.72	0.81	0.14
	5	0.74	0.50	0.76	0.81	0.13
7	6	0.72	0.55	0.81	0.64	0.14

STUDY AREA AND SPECIFICATIONS

Table 1: Dataset and Specifications

Sensor	Area	Resolution	Acquisition Date
Geo Eye-1	Sendai, Japan	0.5 m	February 22, 2012
Quick bird	Berlin, Germany	-	March 28, 2002
Worldview-2	2000 Summer Olympic Complex, Australia	0.5 m	October 20, 2009
IKONOS	Carribbean Sea, West Indies	1m	December 3, 2009
Worldview-3	Maracana Stadium, Brazil	30 cm	-



Figure 6: Cluster evaluation of GeoEye-1 Image

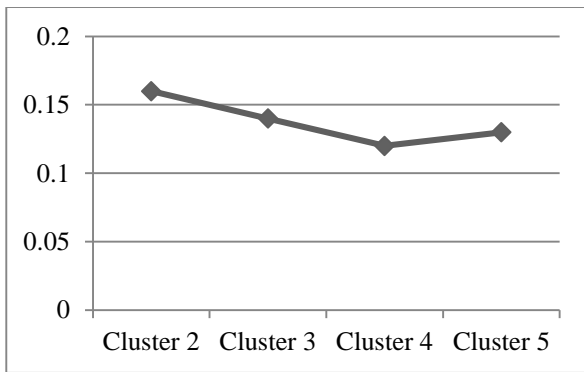


Figure 7: Cluster evaluation of Quick bird Image



Figure 8: Cluster evaluation of Worldview-2 Image

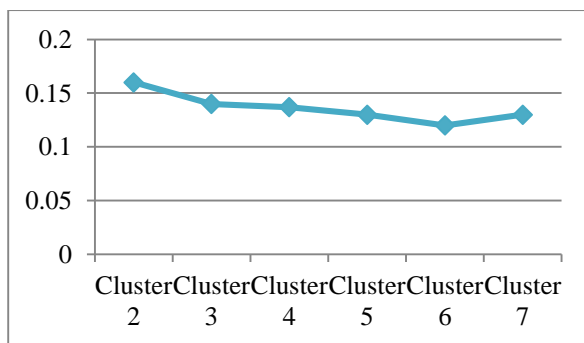


Figure 9: Cluster evaluation of IKONOS image



Figure 10: Cluster evaluation of Worldview-3 Image

As we can see the results, in GeoEye-1 image, clustering stops at 5th cluster in the proposed method. A manual calculation of 6th cluster is carried out to clarify the results. From the manual calculation it can be seen that value of XB-AFCM is low at 5th cluster and increasing in the next cluster. In Quick bird image, proposed method value of 4th cluster is 0.12 and for 5th cluster the value is 0.13. Thus clustering stops at 4th cluster. In Worldview-2 image also it shows the XB-AFCM is low at 5th cluster than 6th cluster, so clustering stops at 5th cluster. In IKONOS and Worldview-3 also it can be viewed as the clustering stops at the point where XB-AFCM value becomes minimum. In terms of results and discussions it is revealed that the proposed method shows accurate number of clusters.

CONCLUSION

In various remote sensing applications clustering evaluation plays a vital task for measuring the compactness of the method. In this work various CVIs are evaluated for different remote sensing images. For the study, five satellite images are taken and clustering performs in each image and CVIs are noted. From the validation it is vividly revealed that clustering stops at the minimum accurate value of XB-AFCM method. From the evaluation it is concluded that the proposed Xie-Beni based Adaptive Fuzzy C Means Clustering (XB-AFCM) performs well and produces accurate number of clusters.

ACKNOWLEDGEMENT

We express our sincere thanks to DST, New Delhi, India for the financial support under INSPIRE fellowship scheme for the above research work.

REFERENCES

- [1] M.Halkidi, Y.Batistakis and M.Vazirgiannis, "On clustering validation techniques," *Journal of Intell. Info. Sys.*, pp. 107-145, 2001.
- [2] A.M Lal and S.M Anoucia, "Semi-supervised change detection approach combining sparse fusion and constrained k means for multi-temporal remote sensing images," *The Egyp. J. Remote Sens.*, vol.18, no.2, pp. 279-288, Dec 2015.
- [3] M.Halkidi, Y.Batistakis and M.Vazirgiannis, "Cluster validity methods: part 1," *Sigmod Record.*, vol. 31, no.2, pp.39-45, Jun 2002.
- [4] H.Li, S.Zhang, X.Ding, C.Zhang and P.Dale, "Performance evaluation of cluster validity indices (CVIs) on multi/hyper spectral remote sensing datasets," *MDPI J. Remote Sens.*, vol.8, no.4, Mar 2016.

- [5] U.Fayyad, G.P-Shapiro and P.Smyth, "From data mining to knowledge discovery in databases," *AI Mag.*, vol.17, no.3, pp.37-54, 1996.
- [6] A. Ghosh, N.S Mishra and S.Ghosh, "Fuzzy clustering algorithms for unsupervised change detection in remote sensing images," *Info. Sci.*, vol.181, no.4, pp. 699-715, Feb.2011.
- [7] H. Cui, M.Xie, Y. Cai, X.Huang and Y.Liu, "Cluster validity index for adaptive clustering algorithms," *IET Comm.*, vol.8, no.3, pp. 2256-2263, Sep.2014.
- [8] W.Wang and Y. Zhang, "On fuzzy cluster validity indices," *Fuzzy Sets and Sys.*, vol.158, no.19, pp. 2095-2117, Oct.2007.
- [9] U. Maulik and I.Saha, "Automatic fuzzy clustering using modified differential evolution for image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3503-3519, Sep.2010.
- [10] X.L Xie and G.Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 13, no. 8, pp. 841-847, Aug.1991.
- [11] N.R Pal and J.C Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Trans. Fuzzy Sys.*, vol. 3, no. 3, pp. 370-379, Aug. 1995.
- [12] J.C Bezdek, "Cluster validity with fuzzy sets," *J. Cybernetics.*, vol.3, no.3, pp. 58-73, 1974.
- [13] C-H Wu, C-S Ouyang, L-W Chen and L-W Lu, "A new fuzzy clustering validity index with a median factor for centroid-based clustering," *IEEE Trans. Fuzzy Sys.*, vol.23, no. 3, pp. 710-718, Jun. 2015.
- [14] Gath and A.B Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Patt. Anal. Machine Intell.*, vol.11, no.7, pp. 773-781, Jul.1989.
- [15] S.Saha and S.Bandyopadhyaya, "Application of a new symmetry-based cluster validity index for satellite image segmentation," *IEEE Geosci. Remote Sens. Letters.*, vol. 5, no.2, pp. 166-170, Apr.2008.
- [16] T.Zhang, R.Ramakrishnan and M.Livny, "BIRCH: an efficient data clustering method for very large databases," *ACM SIGMOD.*, vol.25, no.2, pp. 103-114, June 1996.
- [17] S. Sharma, "Applied multivariate techniques," John Wiley and Sons. 1996.
- [18] S. Guha, R.Rastogi and K.Shim, "CURE: An efficient clustering algorithm for large databases," In *Proc. ACM SIGMOD Conf.*, pp. 73-84, 1998.
- [19] S. Guha, R.Rastogi and K.Shim, "ROCK: A robust clustering algorithm for categorical attributes," In *Proc. IEEE Conf. Data Engg.*, pp. 512-521, March 1999.
- [20] V.S Sidorova, "Automatic hierarchical clustering algorithm for remote sensing data," *Patt. Recogn. Image Analy.*, vol.21, no.2, pp. 328-331, June 2011.
- [21] V.S Sidorova, "Detecting clusters of specified separability for multispectral data on various hierarchical levels," *Patt. Recogn. Image Analy.*, vol.24, no.1, pp. 151-155, March 2014.
- [22] V.S Sidorova, "Hierarchical cluster algorithm for remote sensing data of earth," *Patt. Recogn. Image Analy.*, vol.22, no.2, pp. 373-379, June 2012.