Paper presented at the 22nd Asian Conference on Remote Sensing, 5 - 9 November 2001, Singapore.

THE VALIDITY MEASUREMENT OF FUZZY C-MEANS CLASSIFIER FOR REMOTELY SENSED IMAGES

Chi-Farn Chen and Jyh-Ming Lee
Center for Space and Remote Sensing Research
National Central University
Chung Li, TAIWAN
Tel: 886-3-4227151-7624, Fax: 886-3-4254908
e-mail: cfchen@csrsr.ncu.edu.tw

KEY WORDS: Validity Function, Unsupervised Classification, Number of Clusters

ABSTRACT: A fundamental problem of the unsupervised classification is often the determination of the valid number of the clusters. This study presents a series of testing procedures to investigate the application of a clustering validity function to the fuzzy c-means (FCM) algorithm. The main objective of the investigation is designed to test the performance of the validity criteria for optimal partition. The testing results from a series of remote sensing images indicate that the validity function indeed can be used as the optimal index for the choice of the cluster numbers for the unsupervised fuzzy classification.

1. INTRODUCTION

This study presents a series of testing procedures to investigate the application of the fuzzy c-means (FCM) algorithm (Bezdek, 1981) to remote sensing images. Since the mixture of the land clusters is normally found in the remote sensing image, it is important to assign the proper fuzziness index for the fuzzy objective function before the fuzzy partition can be performed. In addition, because FCM algorithm basically is an unsupervised fuzzy classification, which generally needs the determination of the number of clusters before the algorithm is executed. This study employs the fuzzy clustering validity function to estimate the applicable fuzziness index and measure the valid number of the clusters for remote sensing image. In general, the validity of the partition can be measured by the compactness within clusters and the separation among clusters (Pal and Bezdek, 1995 and 1997). This study uses the fuzzy clustering validity function introduced by Xie and Beni (1991) to measure the compact and separated clusters and estimate the applicable fuzziness index and valid number of the clusters for remote sensing image. In order to analyze this fuzzy clustering validity function, this study designs a series of measurements to compute the validity indices against different fuzziness indices and various cluster numbers. The measurements are implemented by using simulated multi-spectral data and SPOT image. The simulated image is used to obtain the applicable range of the fuzziness index. The result indicates that the fuzziness indices ranging from 2 to 2.5 are able to cope with the relatively fuzzy images. The testing of SPOT image demonstrates that the fuzzy clustering validity function is able to provide the proper guide for the determination of the number of clusters when the FCM algorithm is applied to satellite images.

2. FUZZY VALIDITY FUNCTION OF FCM CLASSIFICATION

The fuzzy validity function basically is designed to measure the overall average compactness and separation of the fuzzy partition. In general, the average compactness can be estimated by the deviation of pixels from the center of each cluster, and the separation of the partition can be represented by the distance between cluster centers. The fuzzy validity function considered in this study was introduced by Xie and Beni(1991).

$$S_{XB,m} = \frac{\sum_{i=1}^{c}\sum_{k=1}^{n}(m_{ik})^m \|x_k - v_i\|^2}{n\left(\min_{i,j} d_{wij}^2\right)} \qquad v_i \neq v_j \ and \ m > 1 \tag{1}$$

where c is the number of clusters, n is the number of pixels, m_{ik} is the membership of kth pixel about ith cluster, x_k is value of kth pixel, v_i is the center of the ith cluster, and d_{wij} is the weighted Euclidean distance, which is the metric distance modified by the standard deviations of cluster i and j.

From equ. 1, it is obvious that the smallest validity function (S) yields the most valid fuzzy partition and accordingly points toward the optimal number of the clusters.

In this study, the use of the fuzzy validity function to obtain the optimal number of the clusters for FCM can be described in the following two main steps.

1) The choice of a proper fuzziness index (m)
- 1.1 Find a test data with the number of clusters ($c^{\#}$) known;
- 1.2 Run FCM to perform the partition for different m with number of clusters (c) from 2 to a pre-defined number;
- 1.3 Calculate the fuzzy validity function (S) from each run of 1.2;
- 1.4 Find the optimal number of clusters (c^*_m) with the smallest S for each m;
- 1.5 Choose the proper m with $c^*_m = c^{\#}$;

2) The determination of the optimal number of the clusters
- 2.1 Choose a proper fuzziness index (m);
- 2.2 Use FCM to perform the partition for the number of clusters (c) from 2 to a pre-defined number;
- 2.3 Calculate the fuzzy validity function (S) for each number of clusters (c);
- 2.4 Determine the optimal number of the clusters(c^*) with the smallest S;

3. APPLICATION OF FUZZY VALIDITY FUNCTION TO MULTI-SPECTRAL IMAGES

The simulated and SPOT images are used to test the fuzzy validity function. For simulated images, the test is designed to select the proper fuzziness index (m) for satellite multi-spectral image. After m is selected, the optimal number of the clusters of SPOT image is tested from the analysis of the fuzzy validity function.

The simulated images, there are four of them, are originated from the statistical parameters such as the means (m) and standard deviations (σ) of a seven-cluster classified SPOT image. The pixel values of the first simulated image (Fig. 1, Image A) that has the same three bands of SPOT image are randomly selected in the range [m (corresponding class mean) $\pm$ 0.3σ(standard deviation)]. The other three simulated images are generated by the same way as the first simulated image but with different standard deviation of 0.5σ (Fig. 2, Image B), 0.8σ (Fig. 3, Image C), and 1σ (Fig. 4, Image D). It is noted that the images generated with larger standard deviation have more blurred appearance. The simulated images are tested with a series of runs of FCM for different values of fuzziness index m=1….5 and the number of clusters c=2….15. Table 1 describes the optimal number of the clusters (c^*_m) of each fuzziness index (m) for Images A, B, C and D. Since the simulated images is created from a seven-cluster SPOT image, Table 1 shows that all of four simulated images point to the correct number of the clusters (c = 7) only at m = 2.5. As a result, it is concluded that the proper fuzziness index (m) can be selected as 2.5 in FCM for the multi-spectral image classification.

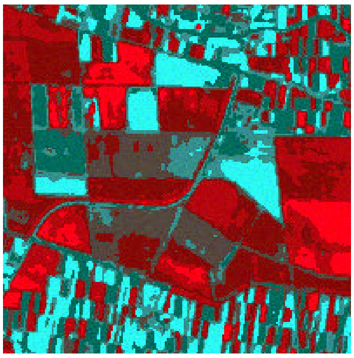


Fig.1 Image A(0.3σ)

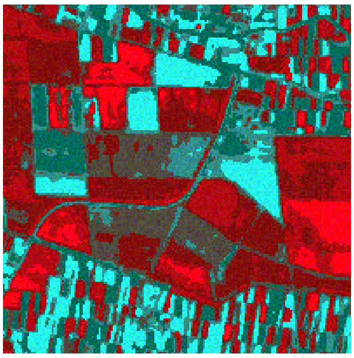


Fig.2 Image B(0.5σ)

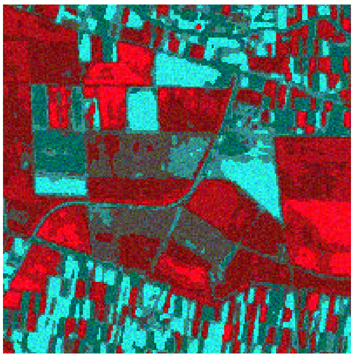


Fig.3 Image C(0.8σ)

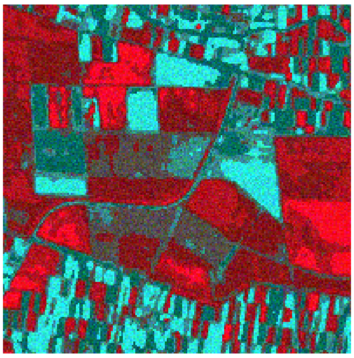
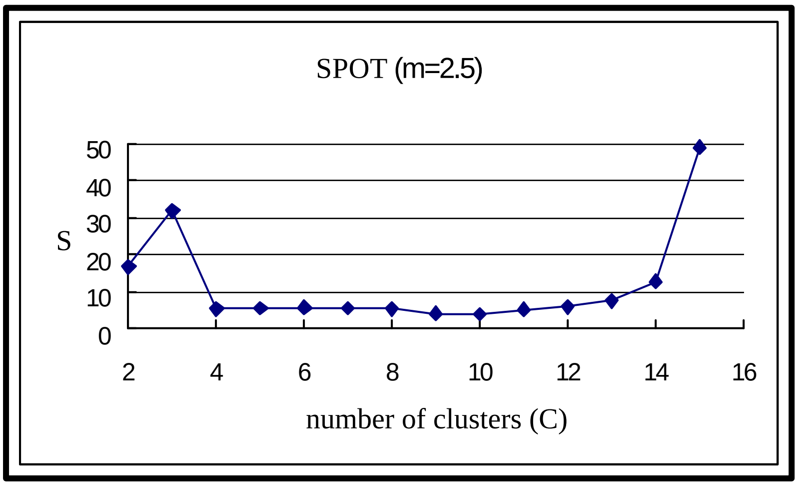


Fig.4 Image D(1.0σ)

Table.1 The optimal number of clusters for Image A,B,C,D at different fuzziness indices(The true number of clusters is 7).

		Fuzziness index								
Image		*m*=1.0	*m*=1.5	*m*=2.0	*m*=2.5	*m*=3.0	*m*=3.5	*m*=4.0	*m*=4.5	*m*=5.0
	A	7	7	7	7	4	5	5	5	3
	B	13	7	7	7	4	3	3	3	11
	C	10	15	7	7	14	15	10	12	12
	D	15	15	5	7	8	13	13	14	15

For a practical application, a SPOT image is used for the test. The test site locates at the typical rural area of southern Taiwan. Through the visual analysis of the image, there are about 10 major features can be found in the image. Moreover, the manual interpretation of the topographic maps also shows that the test site contains about 10 major land features that can be labeled as water, sugarcane field, paddy field, built-up land, two different types of bare soil, and four different types of vegetation. The analysis of the fuzzy validity function of the image shows that the optimal number of the clusters for FCM is 10 (Fig. 5). The visual comparison of the original SPOT image (Fig. 6) and 10-cluster classified image (Fig. 7) indicates that both images agree fairly well with the geometric locations and boundaries of the major clusters.



	m=2.5
C	S
15	48.918
14	12.677
13	7.653
12	6.038
11	5.118
10	3.877
9	4.147
8	5.478
7	5.582
6	5.828
5	5.568
4	5.371
3	32.041
2	16.678

Fig.5 Values of fuzzy validity index against different number of clusters. for SPOT image

Fig.6 SPOT image

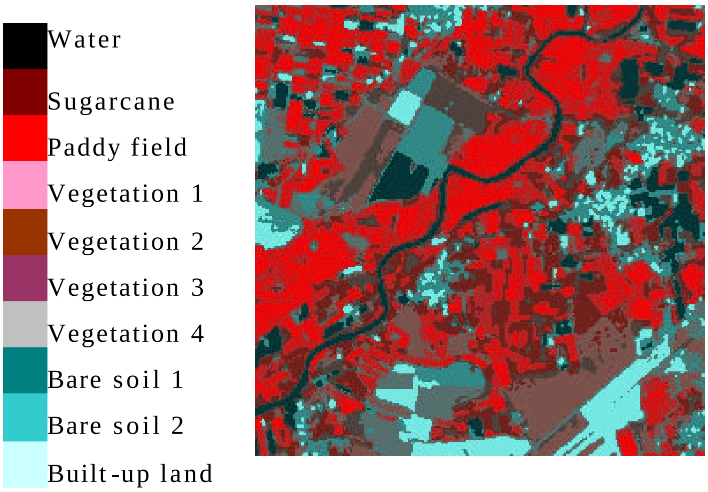


Fig.7 Classified image

4. CONCLUSION

A fuzzy validity function is used to validate the FCM partition for remote sensing images. In general, two important parameters for FCM algorithm to choose are the proper fuzziness index and the optimal number of clusters. This study uses a series of test to analyze the fuzzy validity function and demonstrate that the fuzziness index with value of 2.5 would be the preferred choice for FCM when applying to remote sensing images. The practical test for SPOT image indicates that the fuzzy validity function indeed provide a useful guidance to obtain the optimal number of clusters for FCM.

REFERENCES

Bezdek, J. C., 1981. Pattern recognition with fuzzy objective function algorithms, Plenum Press, New York.

Pal, N. R. and J. C. Bezdek, 1995. On Cluster Validity for the Fuzzy c-Means Model, IEEE Transactions on Fuzzy Systems, Vol.3, No.3, pp.370-379.

Pal, N. R. and J. C. Bezdek, 1997. Correction to "On Cluster Validity for the Fuzzy c-Means Model, IEEE Transactions on Fuzzy Systems", Vol.5, No.1, pp.152-153.

Xie, X. L. and G. Beni, 1991. A Validity Measure for Fuzzy Clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, No.8, pp.841-847.