AN IMPROVED ALGORITHM FOR SUPERVISED FUZZY C-MEANS CLUSTERING OF REMOTELY SENSED DATA

ZHANG Jingxiong  
Roger P Kirby

KEY WORDS remotely sensed data(images); classification; fuzzy c-means clustering; fuzzy membership values(FMVs); Mahalanobis distances; covariance matrix

ABSTRACT This paper describes an improved algorithm for fuzzy c-means clustering of remotely sensed data, by which the degree of fuzziness of the resultant classification is decreased as comparing with that by a conventional algorithm; that is, the classification accuracy is increased. This is achieved by incorporating covariance matrices at the level of individual classes rather than assuming a global one. Empirical results from a fuzzy classification of an Edinburgh suburban land cover confirmed the improved performance of the new algorithm for fuzzy c-means clustering, in particular when fuzziness is also accommodated in the assumed reference data.

1 Introduction

Classification plays an important role for remotely sensed data to be integrated into geographical information systems(GISs), and is increasingly computerized with sophisticated hardware and software (Campbell 1987; Lillesand and Kiefer 1994). Products of classification are usually represented in form of contiguous patches of pixels, with each being labelled as belonging to a discrete and dominant class. Such type of classification is termed as crisp or discrete. The accuracies of classification are often evaluated via error matrices, thus are also discrete as each pixel ground equivalent is assumed to belong to one of the candidate classes, which are made available by means of ground truthing (Congalton 1991). However, because of the complex interactions among remote sensors, the atmosphere and the underlying geographical phenomena, pixels are often mixed. Therefore, a more sensible and suitable alternative to discrete classification is fuzzy classifier, which allows for the resultant classified maps to have partial and multiple memberships at individual pixels (Fisher and Pathirana 1989; Wang 1990; Foody 1995).

Fuzzy classification can be performed in both supervised and unsupervised modes. In a supervised mode, fuzzy membership values (FMVs) are derived from intermediate outputs indicating relative strengths of class memberships of a pixel belonging to all the candidate classes, depending on a specific classifier that would be used in a discrete classification. That is to say that fuzzy classification is achieved by defining suitable fuzzy membership functions. In an unsupervised mode, on the other hand, fuzzy classification can be derived from the fuzzy c-means clustering (Bezdek et al., 1984). Fuzzy c-means clustering is a form of cluster analysis as discussed in general statistics. Unlike a supervised method that relies on a suitable distribution assumption and predefined class statistics, the fuzzy c-means clustering seeks to explore the coherence in the underlying data.

Fuzzy c-means clustering may be performed in a
supervised mode when clusters centres are known. This characteristic makes it a very attractive technique (Key et al., 1989). In a usual or conventional implementation of supervised fuzzy c-means clustering, all the data set is used to calculate a global covariance matrix, which is, in turn, used to define the distance norm, usually the Mahalanobis distance for remotely sensed images (Bezdek et al., 1984).

The degree of fuzziness of the resultant classification depends on the number of clusters aimed at, the norm used and the distribution of the underlying data set. When the number of clusters and the norm used are prescribed, the possibility for adjusting the degree of fuzziness of the resulting classification seems to be in the way the covariance matrix is defined.

From this viewpoint, an improved fuzzy c-means clustering algorithm of supervised mode is described in this paper. This is actually to supply not only class centres as in the usual way, but also the class covariance matrix, thus tightening the control over the evaluation of fuzzy membership values (FMVs) and hence the result of classification. This original improvement was based on the reasoning that specifying a per-class covariance matrix rather than a single covariance matrix for the whole image data set would be more sensible in order to produce a fuzzy classification closer to expectations implied in the class statistics on a per-class basis.

2 Fuzzy c-means clustering: the conventional and the improved algorithms

Let $X = \{x_1, x_2, \cdots, x_n\}$ be a sample of $n$ observations in $n$-dimensional Euclidean space ($c \leq n$). A fuzzy clustering is represented by a fuzzy set $M_{fc}$ with reference to $n$ observations and $c$ clusters, which is defined as:

$$M_{fc} = \{U_{c \times n} \mid \mu_{ik} \in [0.0, 1.0]\}$$

where $U$ is a real $c \times n$ matrix consisted of elements denoted by $\mu_{ik}$, and $\mu_{ik}$ is a fuzzy membership function expressing the FMV of an observation $x_k$ to the $i$th cluster. The value of FMV ranges between 0.0 and 1.0 and is positively related to the degree of similarity or strength of membership for a specified cluster. Besides, it is required that the sum of FMVs for an observation $x_k$ should be 1 across all clusters.

There is a variety of algorithms aiming at an optimal fuzzy c-means clustering. One method works by minimising a generalised least-square error function $J_m$:

$$J_m = \sum_{i=1}^{c} \sum_{i=1}^{n} (\mu_{ik})^m (d_{ik})^2$$

where $m$ is the weighting exponent which controls the degree of fuzziness (increasing $m$ tends to increase fuzziness; usually, the value of $m$ is set between 1.5 and 3.0); $d_{ik}$ is the distance between each observation $x_k$ and a fuzzy cluster centre (Bezdek et al. 1984). Usually, the Mahalanobis distance is used in remote sensing, which is calculated by the following equation

$$d_{ik}^2 = (x_k - v_i)^T \text{cov}(x_k - v_i)$$

where $v_i (i = 1, 2, \cdots, c)$ is fuzzy cluster centre for class $i$, $\text{cov}$ stands for the covariance matrix of the sample $X$, symbol $T$ indicates transpose of matrix.

Fig. 1 and Table 1 in combination provide an illustration of the process of a fuzzy c-means clustering, where three clusters are aimed at, and the value of parameter $m$ is 2.0, using a simple example with 18 pixels in a two-dimensional multispectral space. Pixels are in the order from left to right and bottom to top and the FMV are in percentage. Initial fuzzy cluster centres $v_1$, $v_2$ and $v_3$ are indicated by solid squares shown in Fig. 1(a), while final fuzzy cluster centres $v_1$, $v_2$ and $v_3$ are indicated by solid squares, with three clusters being formed by dashed boundaries, as shown in Fig. 1(b). The initial $U$ matrix (i.e., $U^0$) is generated randomly, whereas the final $U$ matrix is derived from the clustering process.

A supervised approach may be taken if class means are known. In other words, if $v_i$ are available by consulting training data, the fuzzy c-means clustering algorithm described previously becomes simply a one step calculation, by which the FMV for each pixel in each of the known classes can be calculated straightforwardly. Thus, fuzzy c-means clustering can be applied in both an unsupervised and a supervised mode, making itself particularly useful for fuzzy classification of remote sensing im-
An Improved Algorithm for Supervised Fuzzy c-means Clustering of Remotely Sensed Data

It is clear by checking the procedure for the conventional fuzzy c-means clustering that it is possible to adjust the degree of fuzziness of the resulting classification by tuning the number of clusters aimed at, the value of \( m \), and the norm for calculating distances. In a supervised fuzzy classification, the number of clusters is known. Thus, when a certain choice of the value of \( m \) and norm for calculating distances is made, the only possible way to adjust the degree of fuzziness in fuzzy classification is to decide a different strategy for using the covariance matrix. That leads to incorporating the per-cluster covariance matrices instead of a global one. Arguably, such a technique may be more able to produce a fuzzy classification with lowered degree of fuzziness (thus higher accuracies) as opposed to a conventional implementation, because each target cluster is subjected to tighten control over its distribution, thus leaving more pixels with higher FMVs to their dominant clusters.

The proposed fuzzy c-means clustering is tested using the simulated data set shown in Table 1 and Fig. 1. Suppose there are three classes, with which the whole data set will be classified, as indicated by the three ellipses shown in Fig. 2. In Fig 2, dots represent the individual pixels, while centres of classes are represented by small squares. Firstly, a global covariance matrix relevant to the whole data set is used, as in the unsupervised mode, where the global covariance matrix was supplied to calculate FMVs. As would be expected, the resulting classification is identical to that indicated in Fig. 1(b).

Then, the three classes are assumed to be defined by the three ellipses shown in Fig. 2, where centres and covariance matrices can be calculated for each individual class. By such a per-class control over the calculation of FMVs, the classified result is exactly the same as defined by the three ellipses, which appears more suitable than that shown in Fig. 1(b).

In order to evaluate the degree of fuzziness for a fuzzy classification, measure of entropy may be used (Foody, 1995). Measures of entropy express the way in which the probability of class membership is
partitioned between the classes. It is based on the assumption that in an accurate classification each grid cell will have a high probability of membership in only one class. Large entropy values indicate low accuracy in classification, while small values indicate high accuracy in classification. Entropy $H(p(x))$ is measured by

$$H(p(x)) = - \sum_{i=1}^{c} p_i(x) \log_2 p_i(x)$$

(4)

where $p_i(x)$ is the fuzzy membership value of grid cell $x$ belonging to class $i$, and the index $i$ ranges from 1 to $c$ (the total number of classes).

The improvement of the proposed technique over the conventional algorithm in terms of degree of fuzziness can be verified by calculating measures of entropy as described above for the example in Fig. 2. When a global covariance matrix is used, the resulting FMVs have an average measure of entropy of 0.86, while a lower measure of entropy of 0.69 is produced when using per-class statistics. Therefore, the improved method of the fuzzy c-means clustering will reduce the degree of fuzziness in the resulting fuzzy classification, thus making it more compatible with those fuzzy classifications derived from photogrammetric data, which often possess lower degrees of fuzziness (Zhang and Kirby 1997).

It is, however, worth noting that the interpretation of measures of entropy on an individual grid cell basis is not straightforward. In situations where both classified data and reference data are fuzzy, entropy measures will not be suitable. For example, when reference data are fuzzy, any entropy could be associated with an accurate representation; the interpretation of entropy values is therefore difficult. Foody (1995) suggested that cross-entropy should be used to illustrate how closely a fuzzy classification represents the geographical reality when multiple and partial class membership is a feature of the reference data as well as the classified data. The smaller the measure of cross-entropy, the closer the classified data are to the reference data. Cross-entropy $H_c(p(x), g(x))$ is measured using equation:

$$H_c(p(x), g(x)) = - \sum_{i=1}^{c} p_i(x) \log_2 g_i(x) + \sum_{i=1}^{c} p_i(x) \log_2 p_i(x)$$

(5)

where, $p_i(x)$ is the fuzzy membership value of a grid cell $x$ belonging to class $i$, and $g_i(x)$ is the probability of finding class $i$ at grid cell $x$ as defined on the reference data layer.

As there does not exist an independent layer of reference data for the simulated data set shown in Fig. 1, it is not possible to use the measure of cross-entropy to evaluate the closeness between them. So, it becomes necessary to undertake an empirical test in which real data are incorporated.

3 An empirical test

A suburban Edinburgh area near Blackford Hill was chosen as the test site. A variety of thematic and topographic features includes a wooded valley, residential buildings and non-residential complexes, road networks and footpaths, recreational areas, a small lake, agricultural fields and worked allotments, hills and flat ground. It appears that the area chosen provides a good environment with significant mixture of well-defined and poorly defined features to test the alternative approaches mentioned earlier.

For the purpose of this test, Landsat TM data (pixels of 30m) of the test site was used, which was imaged in mid-May 1988. As the source for deriving reference data, one stereo-model of aerial photographs at 1:24 000 scale was used. The aerial photographs (in natural colour) were flown in mid-June 1988, as part of the Scottish national aerial photographic initiative (Kirby 1992). As both the Landsat TM data and aerial photographs were acquired in the summer of 1988, it is assumed that there are no significant effects of temporality in land cover.

Simple classes were used, appropriate for the scene: (1) park and grassland, (2) built-up and barren land, (3) woodland, (4) shrubland, and (5) water bodies. This five-class scheme was used for classification of the Landsat TM data and the aerial photographs.

A fuzzy c-means clustering program with switches to both the conventional and the proposed algo-
An Improved Algorithm for Supervised Fuzzy c-means Clustering of Remotely Sensed Data

43

rithms was written in FORTRAN 77 on VAX/VMS in accordance with the algorithms described in Section 2. Due to the unsatisfactory results from the unsupervised classification, the supervised fuzzy clustering was applied, with bands 3, 4 and 5 used for the Landsat TM image (Zhang 1996). This makes the test of two alternative methods for supervised fuzzy classification logical.

For this, training data were acquired by identifying, for each class, blocks of representative pixels from the image displayed. These pixels were selected from outside of the test site if possible. The numbers of pixels selected are listed in Table 2. The class statistics (including class means and covariance matrices) were read out from signature data compiled from training samples listed in Table 2. The supervised fuzzy c-means clustering program was then performed to generate fuzzy classifications with the conventional and the proposed algorithms respectively.

To generate fuzzy reference data from the photographic model which was reconstituted on an AP190 analytical plotter, available in house at the university of Edinburgh, indicator kriging was used to interpolate class membership spatially. Indicator kriging is a specific approach in geostatistics, which estimates probabilities of finding individual classes at a location given a set of classified samples (Burrough 1986). It is supported by a geostatistical software system GSLIB (Deutsch and Journel 1992). For this, a set of classified samples was identified from screen-displayed photogrammetric data, as shown in Table 3. These sample points were carefully selected to ensure that each could be considered as a pure point, thus they have a full membership (100%) to the named class and zero memberships to all other classes. They were then transformed to a grid coordinate system with grid cell size of 2.5m x 2.5m and the semivariograms were calculated. The kriging procedure ran eventually with the output grid cell sizes equal to Landsat TM data pixel sizes (Zhang and Kirby 1997).

The evaluations of how FMVs are partitioned among candidate classes in the fuzzy classified Landsat TM data and how close the fuzzy classification based on Landsat TM data is to the fuzzy reference data are based on measures of entropy and cross-entropy, respectively. Average values for a total of 847 pixels are listed in Table 4. It is shown that consistently lower values of entropy and cross-entropy (thus better accuracies) are obtained by using the proposed algorithm than by using the conventional algorithm. This confirms the superiority of the proposed algorithm over the conventional algorithm.

| Table 2 | Sample sizes for deriving fuzzy classification from Landsat TM data |
|---------|-------------------------|
| Land cover | Training samples |
| grassland | 72 |
| built-up land | 109 |
| woodland | 39 |
| shrubland | 47 |
| water bodies | 16 |
| total | 283 |

| Table 3 | Sample sizes for deriving fuzzy reference data from aerial photographs |
|---------|-------------------------|
| Land cover | Classified samples |
| grassland | 234 |
| built-up land | 291 |
| woodland | 203 |
| shrubland | 76 |
| water bodies | 11 |
| total | 815 |

| Table 4 | Measures of entropy and cross-entropy (number of pixels: 847) |
|---------|-------------------------|
| Algorithm | Entropy | Cross-entropy |
| The conventional | 1.87 | 2.43 |
| The proposed | 1.76 | 1.81 |

4 Conclusions

The proposed algorithm for fuzzy c-means clustering was compared with the conventional algorithm by using simulated and empirical data. It was confirmed that the proposed algorithm is more suitable for deriving fuzzy classification from remotely sensed data than the conventional algorithm, in particular where there is a continuum of well-defined and poorly defined entities in the mapped phenomena and when fuzziness must be accommodated in both the test data and the assumed reference data. The findings in this paper will also be of interests
5 Acknowledgements

Research described in this paper represents part of Zhang Jingxiong doctoral studies at The University of Edinburgh, which was funded by a Sino-British Friendship Scholarship Scheme (SBFSS) award. The SBFSS is administered by The State Education Commission of China (Beijing) and The British Council (Manchester).

References

1 Besdek J C, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm. Computers and Geosciences, 1984, 10: 191 – 203
2 Burrough P A. Principles of Geographical Information Systems for Land Resources Assessment. Oxford: Clarendon, 1986
3 Campbell J B. Introduction to remote sensing. New York: The Guilford Press, 1987
4 Congalton R G. A review of assessing the accuracy of classification of remotely sensed data. Remote Sensing of Environment, 1991, 37(1): 35 – 46
5 Deutsch C, Journel A G. GSLIB: Geostatistical Software Library and User Guide. New York and Oxford: Oxford University Press, 1992
6 Fisher P F, Pathirana S. Urban boundary detection from Landsat imagery: a GIS and knowledge-based approach to MSS data. Technical Papers. ASPRS/ACSM Annual Convention. 1989, 93 – 101
7 Fodey G M. Cross-entropy for the evaluation of the accuracy of a fuzzy land cover classification with fuzzy ground data. ISPRS Journal of Photogrammetry and Remote Sensing, 1985, 50(5): 2 – 12
8 Goodchild M F, Sun G Q, Yang S R. Development and test of an error model for categorical data. International Journal of Geographical Information Systems, 1992, 6: 87 – 104
9 Key J R, Maslanik J A, Barry R G. Cloud classification from satellite data using a fuzzy sets algorithm; a polar example. International Journal of Remote Sensing, 1989, 10(12): 1823 – 1842
10 Kirby R P. The 1987 – 1989 Scottish national aerial photographic initiative. Photogrammetric Record, 1992, 14: 187 – 200
11 Lillesand T M, Kiefer R W. Remote Sensing and Image Interpretation. 3rd ed. New York: John Wiley and Sons, 1994
12 Wang F. Improving remote sensing image analysis through fuzzy information representation. Photogrammetric Engineering and Remote Sensing, 1990, 56(8): 1163 – 1169
13 Zhang J X. A surface-based approach to the handling of uncertainties in an urban-orientated spatial database. [Ph. D Thesis]. The University of Edinburgh, 1996
14 Zhang J X, Kirby R P. An evaluation of fuzzy approaches to mapping land cover from aerial photographs. ISPRS Journal of Photogrammetry and Remote Sensing, 1997