available at www.sciencedirect.com
SCIENCE DIRECT®
www.elsevier.com/locate/ecolinf

Linear mixture model approach for selecting fuzzy exponent value in fuzzy c-means algorithm

*Francis Okeke, Arnon Karnieli**

The Remote Sensing Laboratory, Jacob, Blaustein Institute for Desert Research Ben Gurion University of the Negev, Sede Boker, Campus, 84990, Israel

ARTICLE INFO

Article history:
Received 31 July 2005
Received in revised form 27 September 2005
Accepted 5 October 2005

Keywords:
Fuzzy classification
Image processing
Linear mixture model

ABSTRACT

The implementations of both the supervised and unsupervised fuzzy c-means classification algorithms require a priori selection of the fuzzy exponent parameter. This parameter is a weighting exponent and it determines the degree of fuzziness of the membership grades. The determination of an optimal value for this parameter in a fuzzy classification process is problematic and remains an open problem. This paper presents a new and efficient procedure for determining a local optimal value for the fuzzy exponent in the implementation of fuzzy classification technique. Numerical results using simulated image and real data sets are used to illustrate the simplicity and effectiveness of the proposed method.

1. Introduction

Fuzzy c-mean (FCM) is an unsupervised classification or clustering algorithm that has been applied successfully to a number of problems involving feature analysis, clustering and classifier design, such as in agricultural engineering, remote sensing, astronomy, chemistry, geology, image analysis, medical diagnosis and shape analysis. It is a method that is especially suitable for ecological modeling (Foody, 1996a,b) and plant ecological mapping (Brown, 1998). Fuzzy c-means is a clustering algorithm that has commonly been adapted for supervised classification of remotely sensed imagery (Atkinson et al., 1997; Deer, 1998; Deer and Eklund, 2003; Foody, 1996a,b). The idea to adapt the FCM for use in a supervised mode in the classification of remotely sensed images was first proposed by key et al. (1989) and later implemented by Foody (1992). The modification from unsupervised to supervised classification involves the specification of the means and covariance matrix as determined from training data, and requires only a single pass of the data through the algorithm (Deer, 1998; Deer and Eklund, 2003; Foody, 2000).

The FCM algorithm is based on an iterative optimization of an objective function J_m, which is the weighted sum of squared errors within groups and is defined as follows (Bezdek, 1981):

$$J_m(U,V;X)=\sum_{k=1}^{n}\sum_{i=1}^{c}U_{ik}^{m}\|x_k-v_i\|_A^2,\quad 1<m<\infty, \tag{1}$$

where U is a fuzzy c-partition of $X \cdot X=\{x_1, x_2, \ldots, x_n\}$ is a finite data set in the pattern space $\mathbf{R}^S$, c is a fixed and known number of cluster, $\mathbf{V}=(v_1, v_2, \ldots, v_c)\in \mathbf{V}^{cs}$ with $v_i\in\mathbf{R}^S\forall i$ is the cluster prototype or cluster center of class i, $1\le i\le c$. The parameter m is the weighting exponent for each fuzzy membership, which determines the "degree of fuzziness", $1\le m<\infty$. If $m=1$, then the algorithm reduces to the hard c-means algorithm. $U=[u_{ik}]\in\mathbf{R}^{cn}$ is referred to as the grade of

* Corresponding author. Tel.: +972 8 6596855; fax: +972 8 6596805.
E-mail addresses: francisokeke@yahoo.com (F. Okeke), karnieli@bgu.ac.il (A. Karnieli).

1574-9541/$ - see front matter
doi:10.1016/j.ecoinf.2005.10.006

membership of x_k to the cluster i. u_{ik} satisfies the following constraints;

$$u_{ik}\in[0,1];\ 1\le i\le c,\ 1\le k\le n; \sum_{i=1}^{c} u_{ik}=1;\ 1\le k\le n;\ 0<\sum_{k=1}^{n} u_{ik}<n;\ 1\le i\le c. \quad (2)$$

$\|x_k-v_i\|_A^2$ is an inner product induced norm on $\mathbf{R}^S$; A can be any $(s\times s)$ symmetric positive definitive matrix. Bezdek (1981) shows that, for $\|x_k-v_i\|_A>0$, $\forall i$, $\forall k$, then J_m is minimized only when $m>1$, and

$$v_i=\frac{\sum_{k=1}^{n}(u_{ik})^m x_k}{\sum_{k=1}^{n}(u_{ik})^m},\quad 1\le i\le c, \quad (3)$$

$$u_{ik}=\frac{1}{\sum_{j=1}^{c}\left(\frac{\|x_k-v_i\|_A^2}{\|x_k-v_j\|_A^2}\right)^{1/m-1}},\quad 1\le i\le c,\ 1\le k\le n. \quad (4)$$

The convergence of the fuzzy c-partition and its representative fuzzy cluster centers has been demonstrated in Bezdek (1981). The optimal solution is obtained by Picard iteration through Eqs. (3) and (4). The iteration is terminated when it reaches a stable condition. This can be defined, for example, when the changes in the cluster centers or the membership values at two successive iteration steps is smaller than a predefined threshold value.

Note that in Eq. (1) any positive definite matrix A induces the inner product norm via the weighted inner product expression (Bezdek, 1981). Thus, $\|x_k-v_i\|_A^2=(x_k-v_i)^T A(x_k-v_i)$. The matrix A determines the norm used in Eq. (1) and effectively controls the shape of the clusters (Bezdek, 1981; Bezdek et al., 1984; key et al., 1989). Therefore, the FCM algorithm requires, a priori, the choice of a norm (A matrix) for its execution, and A is neither determined nor optimized during the execution of the FCM algorithm. Because every norm on $\mathbf{R}^S$ is inner product induced, there are infinitely many norms available for use in Eq. (1). However, only three of these norms enjoy widespread use (Bezdek, 1981; Bezdek et al., 1984). These three norms are the Euclidean norm, the Diagonal norm and the Mahalonobis norm. Let CC_x be the sample covariance matrix of the data set X; let $\{a_i\}$ denote the eigenvalues of CC_x; let D_x be the diagonal matrix with diagonal elements $(d_x)_{ii}=a_i$; and let I be the identity matrix. Then, Bezdek et al. (1984) defines these three norms as:

$$A=I\sim\text{Euclidean norm}, A=D_x^{-1}\sim\text{Diagonal norm},$$
$$A=CC_x^{-1}\sim\text{Mahalonobis norm}.$$

A detail discussion of the geometric and statistical implications of these norms can be seen in Bezdek (1981). Geometrically, each norm induces a topology for $\mathbf{R}^S$, which has a differently shaped unit ball. The Euclidean norm generates hyper spherical balls, so is appropriate when clusters in X have the general shape of "spherical" clouds, while the Mahalonobis norm accounts for distortions in the same way as the Diagonal norm and, additionally, tries to mitigate the effect of statistical dependence between pairs of measured features (Bezdek, 1981). In some cases, emphasis is placed on the dependence of J_m on A. For example, in the case of the modification of the original FCM algorithm for automatic adaptation for different cluster shapes as in Gustafson and Kessel (1979) and its further extension in Gath and Geva (1989). In this paper, only the Euclidean norm is applied and further reference to the matrix A is suppressed.

In addition to the number of classes c, the FCM clustering algorithm and its various extensions require a priori choice of the "degree of fuzziness" parameter m, also called fuzzy exponent, or the weighting exponent or smooth factor. When using the FCM in the supervised mode, the number of classes is known a priori and, in the unsupervised mode, the optimal number of classes is determined by fuzzy validity measures or from pre knowledge. This leaves the value of the fuzzy exponent m to be determined and its determination is problematic (Pal and Bezdek, 1995; Pal and Bezdek, 1997; Yu et al., 2004). As m approaches 1, the clustering becomes harder. As m becomes very large (i.e. $m\ge100$), membership becomes almost constant and so fuzzy that virtually no cluster would be distinguished (McBratney and Moore, 1985). Since the parameter m is not constrained at the upper end, it poses the question as to what value m should take and whether there is a value that optimizes classification (McBratney and Moore, 1985).

In order to determine an optimal value for m, Bezdek proposed heuristic guidelines for m in Bezdek (1981). Over the years, different ranges and values for the optimal choice of m have been proposed and used by different researchers. Bezdek suggested the range 1–30, with the range 1.5–3 as giving good result. He also gave an interesting interpretation of the special case where $m=2$ (Deer and Eklund, 2003). It was noted, however, that there was no strong theoretical justification or empirical evidence for these choices (Deer and Eklund, 2003).

McBratney and Moore (1985) investigated the choice of m for the FCM algorithm. They observed that the objective function value, J_m, decreases monotonically with increasing number of groups and increasing values of m, and that its rate of change with changing m is not constant. They then observed that the greatest change occurred around $m=2$. Based on the intuitive use of $m=2$ by previous researches, they devised a measure, $\phi=-\left[\left(\frac{dJ_m}{dm}\right)\sqrt{c}\right]$, to be used in determining the optimal value for m, where $\frac{dJ_m}{dm}$ is the derivative of objective function, J_m, over the fuzzy exponent m. They took the best value of m for a class to be at maximum of the curve when plotting ϕ versus class. The procedure described in McBratney and Moore (1985) involves a combined determination of optimal number of classes c and optimal choice of fuzzy index m. Numerical test carried out in their study found $m=2$ to be optimal. Choe and Jordan (1992) proposed a method for determining m based on the concept of fuzzy decision theory. After defining a fuzzy goal as "good" cluster criteria and a fuzzy constraint as minimizing the sum of square errors, they choose m such that m has the maximum membership value obtained by the intersection of the fuzzy goal and fuzzy constraint. Using a numerical example, they found that the FCM algorithm was relatively insensitive to the value chosen for m in the range 8–30. They further suggested that the value $m=12$ was optimal. Deer and Eklund (2003) presented an investigation of the value of the fuzzy exponent, m, in a supervised Mahalanobis distance fuzzy classifier by requiring that the fuzzy class memberships reflect proportions of contributing classes in the pixels of a remotely sensed image. Their empirical investigation using a specific data set revealed conflicting requirements. For pure pixels, they favored fuzzy exponent of $m=1.6$ or higher, giving a high membership in the pure class that the pixel is actually a member of. For mixed pixels, they favored $m=3$ (and preferably lower), thereby distributing the memberships between the classes that actually comprise the pixel. Therefore, they observed that the range 1.6–3 suggest a physical interpretation of the intuitions about "good values" for m expressed by previous authors. Their theoretical investigation, which was also based on the condition that the fuzzy memberships reflect the proportions of contributing classes, revealed some special cases. Most notably, that $m=3$ returns the ideal fuzzy memberships for all class proportions in the case of identical distributions. They stated that this thereby provided a measure of theoretical support for previous findings derived from empirical use of the FCM clustering algorithm. Their study also reported cases where there were no value m that returns fuzzy memberships equaling class proportions. Also some past studies used validity indices to determine appropriate choice for the fuzzy index m. For instance, Chen and Lee (2001) used the separation index S of Xie and Beni (1991) and series of test to determine that $m=2.5$ would be the preferred choice for remote sensing images. However, Pal and Bezdek (1995, 1997) already reported that the S can become unpredictable for very high or low values of exponent, m, and that the index S can be strongly influenced by m because it utilizes the FCM centroids to calculate S.

Most recently, Yu et al. (2004) presented detailed theoretical and numerical analysis of the fuzzy exponent. They gave not only a new local optimality test on fixed points of the FCM, but also proffered a new approach to selecting the weighting exponent in the FCM. They gave two theoretical rules for selecting the fuzzy exponent as follows:

$$\text{Rule } \alpha: \ m \le \frac{\min(s, n-1)}{\min(s, n-1) - 2}, \ \text{if } \min(n-1, s) \ge 3.$$

$$\text{Rule } \beta: \ m \le \frac{1}{1 - 2\lambda_{\max}(F\overset{*}{u})}, \ \text{if } \lambda_{\max}(F\overset{*}{u}) < 0.5.$$

Where $F\overset{*}{u} = H^T H/n$; s and $\lambda_{\max}(Fu^*)$ are the number of nonzero eigenvalues and the maximum eigenvalue of the matrix $F\overset{*}{u}$, respectively; and

$$H = \left[\frac{(x_1 - \bar{x})}{\|x_1 - \bar{x}\|}, \frac{(x_2 - \bar{x})}{\|x_2 - \bar{x}\|}, \ldots, \frac{(x_n - \bar{x})}{\|x_n - \bar{x}\|}\right].$$

These rules provide theoretical upper bounds for the valid fuzzy exponent in the FCM. However when $\lambda_{\max}(F\overset{*}{u}) \ge 0.5$, both rules become invalid and the selection of appropriate fuzzy exponent then depends on the discretion of the user.

Generally, there has not been any unified approach or universally accepted procedure for choice of optimal value for the fuzzy exponent, m. In a review study, Xinbo and Weixin (2002) observed that all the proposal, of interval values and single values for m, come from experiments or experience, which are heuristic, but do not provide a straightforward way for optimal choice of m. In addition, they stated that all the proposals lack appropriate test approaches for selecting optimal m-value and that the open problem of choice for optimal m-value still call for further investigation. In this paper, we present a new procedure for the choice of optimal m-value for FCM algorithm and its various extensions. We show that our algorithm determine good m-value within a range of valid m-values. Also for the case where the rules of Yu et al. (2004) failed to provide an upper bound for the selection of suitable m-value (i.e. when $\lambda_{\max}(Fu^*) \ge 0.5$), our algorithm still provided precise value for the FCM.

2. Proposed new procedure for choice of optimal fuzzy exponent

Different ranges and single values for optimal fuzzy exponent, m, used by different researchers, have been stated in the previous section. For these ranges of values, there are still possibilities of making more precise choices of m-value within the range. Really, there must be an m-value that is the best within a range of m-values, based on certain criteria. This best value is desirable in order to provide a common procedure for comparing different classification tasks. On the other hand, the various proposals or the use of a single m-value ($m=2$ being the most popular) for the FCM processing of any particular data type (i.e. remote sensing data) would not be good enough and may be misleading. This is because every data set from any specific data type has unique data structure. This has been demonstrated in the detailed empirical and theoretical investigation of Deer and Eklund (2003), where it was found that pure pixels favored the use $m=1.6$, while mixed pixels within the same data set favored the use of $m=3$. Also Choe and Jordan (1992) pointed out that the appropriate fuzzy exponent would depend on the complexity of the data structure. These are collaborated by the detailed theoretical and numerical analysis of Yu et al. (2004). The structure of the data set is unknown a priori and may vary considerably from one point to the other within the same data set. Therefore, since every data set has a unique data structure, the optimal m-value should be peculiar to each particular data set and should be sought from within the data structure of each data set.

To do that, let us assume that a finite data set, $X=\{x_1, x_2, \ldots, x_k\}$ in the pattern space $\mathbf{R}^S$, is to be processed by an unsupervised FCM algorithm. Let us also take the products of the classification procedure on X to be $\overset{*}{X}(\overset{*}{U}, \overset{*}{V})$ from the same space; $\overset{*}{c}$ as the resultant number of classes; $\overset{*}{V} = (\overset{*}{v}_1, \overset{*}{v}_2, \ldots, \overset{*}{v}_{\overset{*}{c}}) \in \mathbf{R}^{\overset{*}{c}s}$, with $\overset{*}{v}_i \in \mathbf{R}^S \forall i$ as the resultant class prototype or center of class i, $1 \le i \le \overset{*}{c}$; and $\overset{*}{U} = [\overset{*}{u}_{ik}] \in \mathbf{R}^{cn}$, $\overset{*}{u}_{ik}$ as the resultant grade of membership of x_k to the class i. Assuming we want to reconstruct or predict the original data set X, using the output of the FCM classification procedure, we obtain $\overset{*}{X}(\overset{*}{U}, \overset{*}{V})$. There will always be a difference between the product, $\overset{*}{X}(\overset{*}{U}, \overset{*}{V})$, and the original data set. This is simply because the FCM is a type of generalization of reality. Thus

$$\| X - \overset{*}{X} \| = \sigma, \ \sigma > 0, \ \forall m, \tag{5}$$

where $\|.\|$ is any suitable similarity or distance measure. We define the optimal fuzzy exponent m_o to be the m-value corresponding to min(σ), the minimum value (over some range of m-values) of the difference between the original data set and the reconstructed or predicted data set. Optimality in this case is referred to the favorable condition of being capable of reconstructing the original data set using product of fuzzy classification. This property is desirable since it provides a unifying and common criterion useful for comparing the suitability of different classification outputs.

In order to reconstruct or predict the original data set from the outputs of the FCM algorithm, we employ the idea of linear mixture modeling in remote sensing. Consider the linear mixture model for sub-pixel land cover estimation, where reflectance of a pixel in each spectral band is expressed as linear combination of the characteristic reflectances of its component endmembers weighted by their respective areal proportion within the pixel (Ichoku and Karnieli, 1996). In the linear mixture model, the reflectance r is expressed as

$$\mathbf{r} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{6}$$

where $\mathbf{A}$ represents the endmembers, $\mathbf{x}$ represents the proportion of respective endmembers contained in each pixel and $\mathbf{e}$ represents an error term. In linear mixture modeling, the objective is to solve Eq. (6) for the proportion $\mathbf{x}$ of the endmembers in individual pixels, given the reflectance $\mathbf{R}$ and the endmembers $\mathbf{A}$. In our own case, we are interested in the reverse. If we ignore the error term, we can restate Eq. (6), for a one-dimensional data set (i.e. aerial photographs) using the output of FCM algorithm as follows

$$\overset{*}{X} = \overset{*}{V}\overset{*}{U}, \tag{7}$$

where $\overset{*}{X}$ represents the $(1\times k)$ vector of predicted DN values of the data set, $\overset{*}{V}$ represent the $(1\times\overset{*}{c})$ vector of the FCM outputted prototypes (centers) and $\overset{*}{U}$ represents the $(\overset{*}{c}\times k)$ matrix of the FCM outputted degree of membership of each prototype to the DN values. In this case, $\overset{*}{V}$ is similar to $\mathbf{A}$, the endmembers in Eq. (6), and $\overset{*}{U}$ is similar to $\mathbf{x}$, the proportion of respective endmembers in Eq. (6), while $\overset{*}{X}$ is similar to $\mathbf{R}$, the reflectance in Eq. (6). Thus, we predict X with $\overset{*}{X}$ using the outputs $\overset{*}{V}$ and $\overset{*}{U}$. What follows is a description of a heuristic strategy for the determination of an m-value for an unsupervised FCM classification procedure using the linear mixture model.

2.1. Algorithm for the determination of optimal m-value

This algorithm assumes that the optimal number of classes in the data set must have been determined. Optimal number of classes might be known a priori as in the case of supervised classification in remote sensing or determined by classification validity measures. Thus, assuming we have determined the optimal number of classes $\overset{*}{c}$, then m is determined as follows:

i. Initialization: Set $m = 1.1$; $\overset{*}{c}$, and choose appropriate maximum possible value for m, $m_{max}\cdot m_{max}$ can be chosen to be 3, 5 or 7.
ii. Use any FCM algorithm to compute fuzzy prototypes $\overset{*}{V}$ and fuzzy membership grades $\overset{*}{U}$.
iii. Reconstruct the original data set using Eq. (7).
iv. Compute and record σ, the difference between the original data set and the predicted data set using Eq. (5).
v. Increment the m-value, $m=m+$increment; if $m\le m_{max}$ go to ii. m_{max} has been chosen in step i.
vi. End and determine the optimal m-value, which is the m-value corresponding to the least σ value, min(σ), referred to here as best predicted value (BPV).

Note that, in order not to miss the best m-value, initial runs of the algorithm might be made by setting m_{max} and increment to large values. For example, m_{max} might be set to 30 and the increment set to 5. With this initial run, a coarse best m-value will be obtained from where a more refined m_{max} and increment will be set. Thus, initial runs will act as guidance in the selection of final m_{max} and increment value for the algorithm. Alternatively, rules α and β of Yu et al. (2004) can be used to determine m_{max}. Numerical results of this procedure follow.

3. Experimental validation

3.1. Simulated data set

In order to compare the true cover proportion and the fuzzy membership grades obtained from fuzzy classification, the actual composition of the reference data must be known. To get this information for real data set is problematic due to various practical problems including spatial inaccuracies (Maselli et al., 1996). Because of these problems, we simulated an image of 1-m resolution (Sm1) comprising of four classes (trees, shrubs, herbaceous plants, and bare soil), and with digital numbers (DN) values ranging from 64 to 255. Trees appear in red, shrubs in blue, herbaceous plants in light green and bare soil in cyan. Further, we degraded the resolution of part of this data set by means of spatial degradation technique. The degradation is performed in a block at the center of the image. The resultant image (Sm2) is shown in Fig. 1 and is the image to be used for the analyses. The spatial degradation was done by simple averaging of DN values of 4×4 pixels for the part of the original data set degraded, following the procedure in Maselli et al. (1996). Finally, we computed the proportion of the classes (trees, shrubs, herbaceous plants, and bare soil) in the simulated image by simply considering the number of the classes contained in each degraded pixel based on Sm1. This we termed the true class proportion. For instance, if the degraded pixel is made up of 8 original pixels of trees and 8 original pixels of shrubs, then the proportion of trees and shrubs in this pixel is 0.5 each, while the proportion of herbaceous plant and bare soil would be zero each. However, it should be noted that this method of synthesizing data sets is without regard to the so-called point spread function (PSF) of imaging devices and may lead to data not particularly representative of what an actual imaging device would have recorded (Maselli et al., 1996). This method of computation of the proportion, though not rigorous, is considered adequate for the illustrations in this work.

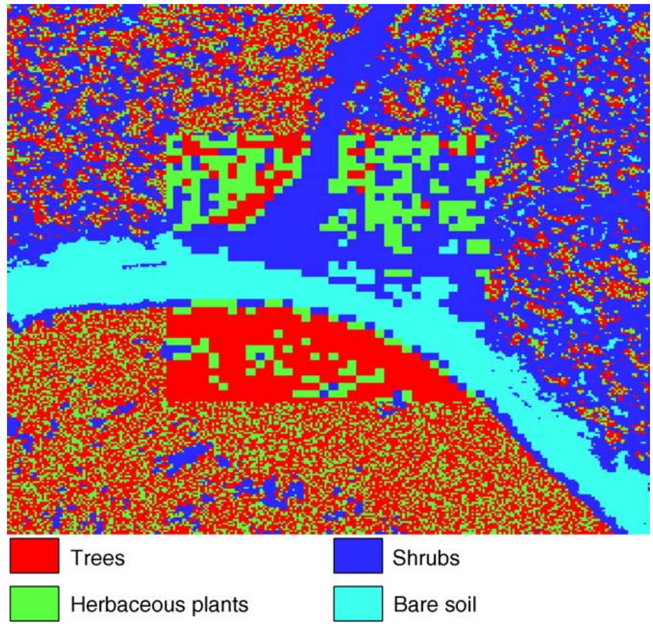

Fig. 1 – Simulated image (Sm2) showing trees, shrubs, herbaceous plants, and bare soil (center of the image is degraded).

3.2. *Results of fuzzy classification of simulated image*

We first determined the optimal number of classes in the data set (Sm2). Varying the fuzzy exponent (*m*) from 1.1 to 3, we performed fuzzy classification for number of classes from 2 to 7. The result is shown in Fig. 2, where the optimal number of classes is confirmed to be 4 using 7 different fuzzy validity criteria. These are: partition coefficient (*F*) and partition entropy (*H*) of Bezdek (1981); non-fuzzy index (NFI) of Roubens (1978); fuzzyness performance index (FPI) and modified partition entropy (MPE) (Roubens, 1982); separation index S of Xie and Beni (1991); V_k of Kwon (1998). The optimal number of classes for the fuzzy classification based on the 7 validity measures is 4. The partition coefficient (*F*) and the non-fuzzy index (NFI) take the maximum value as the optimal number of classes, while the rest take minimum. As expected, Fig. 2 is similar for all values of the fuzzy exponent from 1.1 to 3. Practically, the number of classes does not change significantly for a small variation of *m*-values such as from 1.1 to 3.

With the optimal number of classes now fixed at 4, we then determined the optimal value of the fuzzy exponent using the technique of computing BPV as described in Section 2. In this study, four distance metrics are used to find the difference between the predicted data and the original data set. These metrics are Euclidean distance (*E*), Divergence distance (Div), Bhattacharyya distance (Ba) and Angular separation (AS). We performed fuzzy classification of the data set (Sm2) with number of clusters fixed at 4 and the fuzzy exponent varying from 1.4 to 2.5 at 0.1 increment. Initial runs show that the best value must lie between 1.4 and 2.5. Frobenius matrix norm (Euclidean norm) and the matrix two-norm (spectra norm) were used to determine the difference between the matrix of true

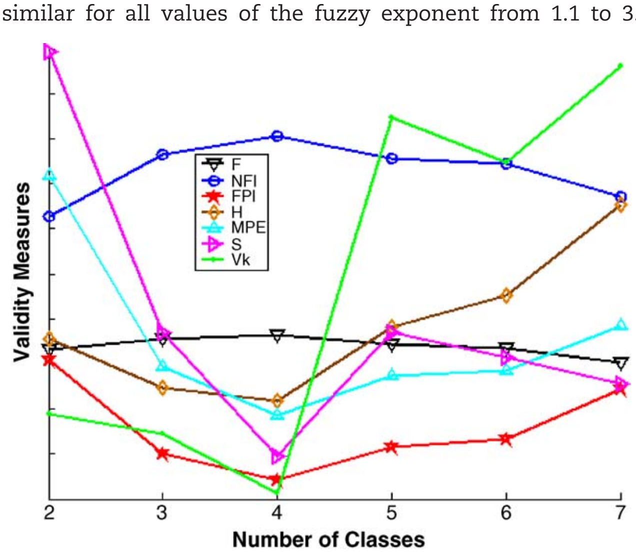

Fig. 2 – Validity measures for number of classes from 2 to 7.

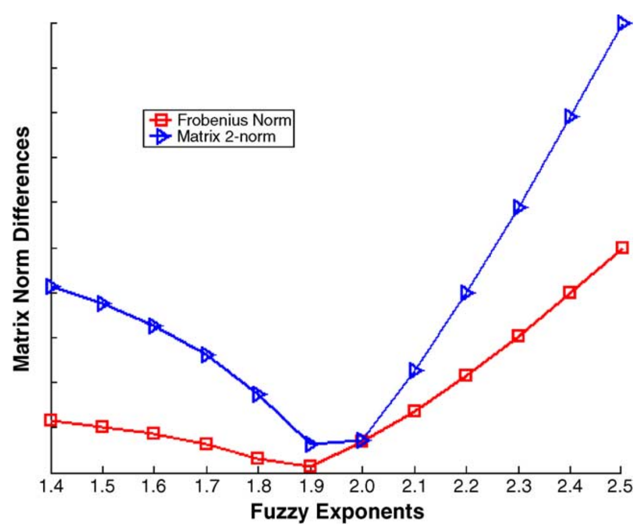

Fig. 3 – Differences of matrix norms for fuzzy exponents from 1.4 to 2.5.

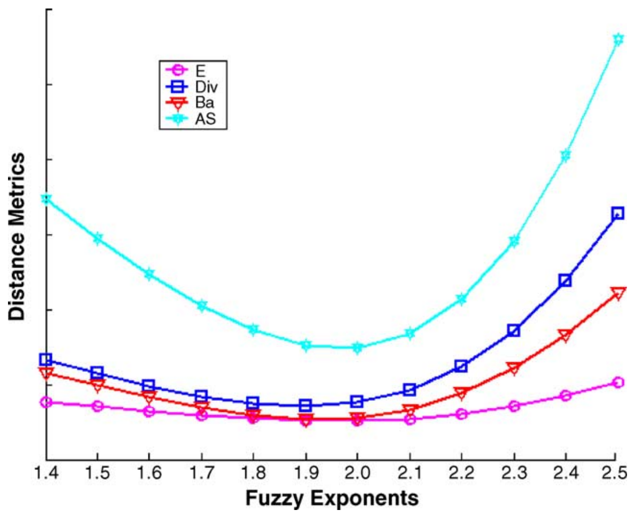

Fig. 4 – Distance metrics for fuzzy exponents from 1.4 to 2.5.

proportion and the membership grades. (The norm of a matrix is a scalar that gives some measure of the magnitude of the elements of the matrix. The spectra norm of matrix *A*, for instance, is the largest singular value of the singular value decomposition of matrix *A*, while the Euclidean norm of matrix *A* is the square root of the sum of the diagonal elements of the product of matrix *A* by itself). The determined differences are plotted against the values of fuzzy exponents. The result is shown in Fig. 3. The figure shows that the fuzzy exponent ($m=1.9$) generated membership grades that have the least matrix distance from the true class proportion, for *m*-values between 1.4 and 2.5. The differences between the original data set and the predicted data set in terms of distance metrics are plotted against fuzzy exponents between 1.4 and 2.5 and shown in Fig. 4. The figure shows that ($m=1.9$) is the optimal value. Here the experiment conducted with the use of four metric indices is only for the sake of comparing results from different possibilities. One metric index will be adequate to determine BPV. These results show the effectiveness of the algorithm in Section 2.1 as well as collaborate earlier reports by researchers that fuzzy membership grades reflect the proportion of contributing classes.

3.3. *Results of fuzzy classification of real data set*

Real data set processed for the numerical experiments in this study were those used by Yu et al. (2004) and can be obtained from the UCI Repository of Machine Learning Databases (UCI, 2005). The data sets have known number of features and number of classes and we did not have to determine the optimal number of classes. For all the data sets, the FCM classification was performed using fuzzy exponent values from 1.1 to 3.0 at an increment of 0.1, and the Euclidean distance was used to compute the reconstruction error between the original data set and the predicted data set. Column 7 in Table 1 shows the optimal fuzzy exponent values as determined by our algorithm of Section 2.1, while column 6 gives the upper bound for the selection of fuzzy exponent values as illustrated in Yu et al. (2004). The results show that the algorithm of Section 2.1 provided precise fuzzy exponent values that fall within the upper bounds as determined by the rules in Yu et al. (2004) for the Sonar, Vowel, Isolet5, PimalIndianDiabetes, Waveform and Glass data sets. Remarkably, for the case of the Iris data set where there was no upper bound based on rules described in Yu et al. (2004), our algorithm of Section 2.1 still determined an optimal fuzzy exponent value.

4. Discussion

The fuzzy *c*-means algorithm converges to a solution when provided with all suitable parameters, which include appropriate number of iterations and or convergence tolerance values, optimal number of classes and optimal fuzzy exponent. Appropriate number of iterations and tolerance values are mainly set during computations. Optimal number of classes is either known a priori from pre knowledge as in supervised classification tasks or determined by using fuzzy validity criteria. Suitable and precise fuzzy exponent values must also be set in FCM computations especially when comparing different fuzzy classification schemes.

Once the optimal number of classes are known, the value for the centroids do not change significantly with changing values of the fuzzy exponent, but the membership grades do (McBratney and Moore, 1985). Therefore, changes in membership grades correlate highly with changes in fuzzy exponents. The linear mixture model of Eq. (6) is used for the reconstruction of the original data set based on the assertion that fuzzy membership grades reflect the proportion of contributing classes. Many past studies have suggested, with varying degrees of empirical support, that there is a strong relationship between fuzzy memberships and true proportions of contributing classes. These include Atkinson et al. (1997),

Table 1 – Result of processing of real data set obtained from the UCI Repository of Machine Learning Databases

Name of data set	No. of samples	No. of features	No. of classes	$\lambda_{max}(F_{Data})$	$\frac{1}{1-2\lambda_{max}(F_{Data})}$	BPV
Isolet5	1559	617	26	0.1926	1.6265	1.2
Sonar	208	60	2	0.1949	1.6388	1.1
Vowel	990	10	11	0.2189	1.7787	1.2
PimaIndian diabetes	768	8	2	0.2558	2.0475	2.0
Waveform	5000	21	3	0.3272	2.8935	1.2
Glass	214	9	6	0.3424	3.1726	1.3
Iris	150	4	3	0.6652	$+\infty$	1.7

BPV refers to the "best predicted value" of Section 2.

Canters (1997), Deer (1998), Deer and Eklund (2003), Fisher and Pathirana (1990), Foody (1992, 1996a,b), Foody and Cox (1994), Maselli et al. (1996) and Wang (1990a,b). However, Schowengerdt (1996) cautioned against the use of likelihood indicators (fuzzy memberships derived partly using hard statistical pattern classification approaches) as a global measure of class mixing proportions, especially for classes having relatively low separability and considerable distribution overlap. Nevertheless, these likelihood indicators are not equivalent to the fuzzy membership grades described in this work, and therefore the above caution does not counter the analysis made in this work. Also Deer and Eklund (2003) pointed out some cases where fuzzy memberships did not equal class proportion. Excluding such special cases, we assert in this paper and in collaboration with results from other past researchers that fuzzy partitions have strong correlation with true class mixing proportions. The centroids are similar to the fixed endmembers in linear mixture model. Though the reconstructed points based on Eq. (7) are convex combination of the fixed centroids, they are also products of centroids and the membership grades and membership grades depend on fuzzy exponent values. Therefore, if the optimal number of classes of a data set has been correctly determined, then the best fuzzy clustering output will actually be the one that reproduced approximately the original data set or nearest to the original data set based on the linear mixture model. The fuzzy exponent that leads to this reproduction of the original data set becomes by implication and optimal value.

Deer and Eklund (2003) in a recent paper proposed an approach for the determination of optimal m-value requiring that the (supervised) fuzzy classification return fuzzy memberships reflecting class proportions in a remotely sensed data set. Our work in this paper extends the notion of Deer and Eklund (2003) by describing an approach that chooses m-value that minimizes the "error" between the output of the FCM and the true class proportions via the original data set. Since in practice it would not always be possible to obtain the true class proportion from the sub-sampled test data or "ground truth", the true data set is used to reconstruct this "error". Future work will try to use true class proportions of different multidimensional data sets to test the algorithm of Section 2.1. Presently for practical purposes, the rules of Yu et al. (2004) (or trial runs of the algorithm of Section 2.1) can be used to set the upper bounds for m-values, and the algorithm of Section 2.1 used with Euclidean distance to derive the best m-value within a range of values.

The method described in this work is similar to the approach adopted in the work of Metternicht (1999), where the sharpness and inflection values of the Semantic Import Model fuzzy function, are manipulated in such a way that the resulting fuzzy membership function is in accordance with the shape of the data to be analyzed for change detection. Sharpness in the fuzzy Semantic Import Model is an indicator of increased membership to a fuzzy set and the inflection is the turning point of the function. Both quantities in the Semantic Import Model correspond to the fuzzy exponent in the FCM algorithm since they act to determine the fuzziness of the membership function.

The algorithm presented in this work for the determination of optimal fuzzy exponent has been programmed in Matlab version 6.5 and tested in a PC (Pentium 4, 2GB RAM). The algorithm runs very fast, for example, approximately 3 min for an average data set of about 1,000,000 points, for one-dimensional data set (aerial photo), using Euclidean distance metric.

5. Conclusion

The fuzzy exponent (m-value) is required for the implementation of the fuzzy classification algorithm. The question of choice of optimal value for the fuzzy exponent has remained an open problem. This paper presents a new and suitable procedure of determining an optimal fuzzy exponent for fuzzy classification. In the proposed procedure, the output of the fuzzy classification is used to predict the original data set. Then the differences between the predicted data and the original data set, for a range of fuzzy exponent values, are inspected. The fuzzy exponent value that corresponds to the least distance between the predicted data set and the original data set becomes the optimal value. Numerical results using simulated image indicate that the proposed method was able to identify a fuzzy exponent value, which gave fuzzy membership grades closest to the true class proportion. Also numerical results with real data sets show the simplicity and effectiveness of the algorithm. It was able to provide precise optimal fuzzy exponent values in all cases and even in the worst case where there is no theoretical upper bound for the fuzzy exponent value.

Outside being able to generate fuzzy membership values that are closest to the true class proportion, the proposed method has the benefit of providing a unified method of identifying optimal fuzzy exponent. This is achieved only with the products of classification and without considering the algorithm of the fuzzy classification itself. Thus the method is applicable to all cases (original fuzzy c-means algorithm and its extensions) wherever the value of the fuzzy exponent is required.

REFERENCES

Atkinson, P.M., Cutler, M.E.J., Lewis, H., 1997. Mapping sub-pixel proportional land cover with AVHRR imagery. International Journal of Remote Sensing 18, 917–935.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.

Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: the fuzzy c-means clustering algorithm. Computers and Geosciences 10, 191–203.

Brown, D.G., 1998. Mapping historical forest types in Baraga County Michigan, USA as fuzzy sets. Plant Ecology 134, 97–111.

Canters, Frank, 1997. Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification. Photogrammetric Engineering and Remote Sensing 63, 403–414.

Chen, Chi-Farn, Lee, Jyh-Ming, 2001. The validity measurement of fuzzy c-means classifier for remotely sensed images. 22nd Asian Conference on Remote Sensing (Singapore).

Choe, H., Jordan, J.B., 1992. On the optimal choice of parameters in a fuzzy c-means algorithm. IEEE International Conference on Fuzzy Systems 349–354.

Deer, Peter J., 1998. Change Detection using Fuzzy Post Classification Comparison, Department of Computer Science. The University of Adelaide, Adelaide, Australia.
Deer, Peter J., Eklund, Peter, 2003. A study of parameter values for a Mahalanobis distance fuzzy classifier. Fuzzy Sets and Systems 137, 191–213.
Fisher, P.F., Pathirana, S., 1990. The evaluation of fuzzy membership of land cover classes in the suburban zone. Remote Sensing of Environment 34, 121–132.
Foody, G.M., 1992. A fuzzy sets approach to the representation of vegetation continua from remotely sensed data: an example from Lowland Heath. Photogrammetric Engineering and Remote Sensing 58, 221–225.
Foody, G.M., 1996a. Approaches for the production and evaluation of fuzzy land cover classification from remotely-sensed data. International Journal of Remote Sensing 17, 1317–1340.
Foody, G.M., 1996b. Fuzzy modelling of vegetation from remotely sensed imagery. Ecological Modelling 18, 3–12.
Foody, Giles M., 2000. Estimation of sub-pixel land cover composition in the presence of untrained classes. Computers and Geosciences 26, 469–478.
Foody, G.M., Cox, D.P., 1994. Sub-pixel land-cover composition estimation using a linear mixture model and fuzzy membership functions. International Journal of Remote Sensing 15, 619–630.
Gath, I., Geva, A.B., 1989. Unsupervised optimal fuzzy clustering. IEEE Transactions of Pattern Analysis and Machine Intelligence 11, 773–781.
Gustafson, D.E., Kessel, W., 1979. Fuzzy clustering with a fuzzy covariance matrix. Proceedings IEEE-CDC 2, 761–766.
Ichoku, Charles, Karnieli, Arnon, 1996. A review of mixture modeling techniques for sub-pixel land cover estimation. Remote Sensing Reviews 13, 161–186.
Key, J.R., Maslanik, J.A., Barry, R.G., 1989. Cloud classification from satellite data using a fuzzy sets algorithm: a polar example. International Journal of Remote Sensing 10, 1823–1842.
Kwon, S.H., 1998. Cluster validity index for fuzzy clustering. Electronic Letters 34, 2176–2177.
Maselli, F., Rodolf, A., Conese, C., 1996. Fuzzy classification of spatially degraded Thematic Mapper data for the estimation of sub-pixel components. International Journal of Remote Sensing 17, 537–551.
McBratney, A.B., Moore, A.W., 1985. Application of fuzzy sets to climatic classification. Agricultural and Forest Meteorology 35, 165–185.
Metternicht, Graciela, 1999. Change detection assessment using fuzzy sets and remotely sensed data; an application of topographic map revision. ISPRS Journal of Photogrammetry and Remote Sensing 54, 221–233.
Pal, N.R., Bezdek, J.C., 1995. On cluster validity for the fuzzy *c*-means model. IEEE Transactions on Fuzzy Systems 3, 370–379.
Pal, N.R., Bezdek, J.C., 1997. Correction to "On cluster validity for the fuzzy *c*-means model". IEEE Transactions on Fuzzy Systems 5, 152–153.
Roubens, M., 1978. Pattern classification problems and fuzzy sets. Fuzzy Sets and Systems 1, 239–253.
Roubens, M., 1982. Fuzzy clustering algorithms and their cluster validity. European Journal of Operational Research 10, 294–301.
Schowengerdt, R.A., 1996. On the estimation of spatial-spectral mixing with classifier likelihood functions. Pattern Recognition Letters 17, 1379–1387.
UCI, 2005. Online: http//www.ics.uci.edu/~mlearn/MLRepository.hml, last consulted: April, 2005.
Wang, Fangju, 1990a. Fuzzy supervised classification of remotely sensed images. IEEE Transactions on Geoscience and Remote Sensing 28, 194–201.
Wang, Fangju, 1990b. Improving remote sensing image analysis through fuzzy information representation. Photogrammetric Engineering and Remote Sensing 56, 1163–1169.
Xie, X.L., Beni, G., 1991. A validity measure for fuzzy clustering. IEEE Transactions of Pattern Analysis and Machine Intelligence 13, 841–847.
Xinbo, G.A.O., Weixin, X.I.E., 2002. Advances in theory and applications of fuzzy clustering. Chinese Science Bulletin 45, 961–970.
Yu, Jiang, Cheng, Qiansheng, Huang, Houkuan, 2004. Analysis of the weighting exponent in the FCM. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics 34.