Research Paper

Volume : 4 | Issue : 12 | December 2015 • ISSN No 2277 - 8179

Preliminary Study of Incorporating Kernels with Fuzzy Classifiers

Engineering

KEYWORDS : Remote Sensing, Land Use Land Cover, Mixed pixel, Kernel method, Fuzzy Classifier, Kernel Based Fuzzy Algorithms.

Ishuita Sengupta

Dr. Ramswarup Colony, Sripati Niwas, Behind Taj Banquet Hall, Moradabad (U.P.), India

ABSTRACT

In Remote Sensing, Land Use/Land Cover is prepared for different project planning purpose using different classification techniques. If there are mixed pixels, classification techniques should be based upon soft approach i.e. soft classification. This paper emphasizes over a preliminary approach of implementing fuzzy classification using the incorporation of kernels based methods. The kernels are used to handle non-linearity. The concept of kernels provides the basis for the development of more robust approaches to the remote sensing classification problem; it implicitly represents the mapping of the input space to the feature space.

The major objective of the paper is to demonstrate the role and effect of using kernels with fuzzy classifiers. An evaluation of kernel based fuzzy algorithms has been formulated for extracting material of interest at sub-pixel level. Output images procured after applying the classification techniques using 8 Kernels with the FCM as well as using standard FCM, every output respective to the feature class has been highlighted.

Introduction

Remote Sensing is used to prepare Land Use/Land Cover for different project planning purpose using different classification techniques. If there is a problem of mixed pixels, then the classification technique to be used must be soft classification. This research work has given a preliminary approach of implementing fuzzy classification using the concepts of kernels. The kernels are used to handle non-linearity. The concept of kernels provides the basis for the development of more robust approaches to the remote sensing classification problem; it implicitly represents the mapping of the input space to the feature space.

Image Classification Approach

Multi-spectral image classification techniques are used for the information extraction for various environmental studies. Image classification can be broadly classified into supervised and unsupervised classification as shown in Figure 1 and Figure 2. Supervised classification algorithms such as the Maximum Likelihood (ML) classifier and the Minimum Distance to mean classifier were introduced taking one-pixel-one-class approach. These supervised classification algorithms i.e. Maximum Likelihood and Minimum Distance to Means classifiers are based on the evaluation of various spectral response patterns when classifying an unknown pixel. The Minimum Distance to Means classifier is one of the simplest classification approaches but it is insensitive to different degrees of variance in the spectral response data.

Supervised classification is much more accurate for mapping classes than unsupervised classification, provided that there should be prior knowledge of the conventional classes. This familiarity allows the specialist to choose and set up discrete classes (thus supervising the selection) and then, assign them category names. For each selected class, mean values and variances of Digital Numbers (for each band), used to classify them are calculated from all the pixels enclosed within the site. For any class, more than one polygon can be established. When Digital Numbers are plotted as a function of the band sequence (increasing with wavelength), the result is a spectral signature for that class.

Fuzzy Based Approach

Land cover classes can be defined as fuzzy sets, and pixels as elements of sets in a fuzzy representation for remote sensing image analysis. Each pixel is attached with a group of membership grades to indicate the extent to which the pixel belongs to certain classes. Pixels with a mixture of classes or in intermediate conditions can be described by membership grades. For example, if a ground contains two cover types, "soil" and "vegetation", it may have two membership grades indicating the extent to which it is associated with the two classes.

Fuzzy logic is an effective method in image classification and collateral data can also be classified well. The FCM algorithm used by Bezdek et al., (1984) is unsupervised in nature. When the information about the classes of interest is known a priori, supervised image classification techniques are most widely used (Campbell, 1996).

Wang (1990) introduced fuzzy supervised classification of remote sensing images with higher classification accuracy. Supervised FCM classification was used for the estimation and mapping of sub-pixel land cover composition (Foody, 2000, Atkinson et al., 1997).

Kernel Based Approach

The data in an image exhibit different pattern that may or may not be clearly visible. Pattern analysis refers to a class of machine learning algorithms that classifies data based on the properties of different patterns. Linearly separable classes are the simplest case that can appear in the pattern of a data (Isaacs et al., 2007). If the data appear to be non-linearly separable the classification will be computationally intricate in the original input space. For separating these non-linear data many kernels based methods were introduced in recent years (Girolami, 2002; Camps-Valls and Bruzzone, 2009). These methods map the input data to a higher dimensional space where the data turn out to be linearly separable (Awan and Sap, 2005).

Mostly kernel based algorithms were used in Support Vector Machines (SVM). It is a statistical learning approach which uses kernels for remote sensing classification (Pal, 2009). Kernel methods are used in wide range of applications. Kernel functions maps sample data from the initial sample space into a higher dimensional space where the sample data are linearly separable and allows interpreting data in feature space.

Possibilistic c-Means Classifier

The main advantage of PCM is related to the relaxation of the probabilistic constraint of FCM. Therefore, the formulation of PCM is based on a modified FCM objective function whereby an additional term called as regularizing term is also included. FCM has been successful in assigning the membership (uij) of a pixel to multiple classes however; the assignment is relative to total number of classes defined (Krishnapuram and Keller, 1993). This is due to the constraint on the membership values given by the Eqn. (1).

$$\sum_{j=1}^{c} \mu_{ij} = 1 \text{ for all i} \quad (1)$$

Where, i varies from 1 to n, n is the total number of pixels in the image, c is the total number of classes defined by the analyst. Eqn. (1) can be interpreted as the sum membership values of a pixel for all the classes and should be equal to one (Bezdek, *et al.,* 1984, Krishnapuram and Keller, 1993). Krisnhapuram and Keller (1993) introduced a relaxation in FCM, such that the sum of membership may exceed beyond 1. This variation of PCM is known as Possibilistic *c*-Means (PCM), where the constraint on membership value is as per Eqn. (1). Thus, similar to FCM, PCM classification is also an iterative process where the class membership values are obtained by minimizing the generalized least square error objective function (Krishnapuram and Keller, 1993), given by Eqn. (2).

$$J_{PCM}(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{c} \mu_{ij}^{m} \left\| x_i - v_j \right\|_A^2 + \sum_{j=1}^{c} \eta_j \sum_{i=1}^{n} 1 - \mu_{ij}^{m} \quad (2)$$

Ensuring that the conditions given below are satisfied:

$max\ \mu_{ij} > 0,$ (3)

$\sum_{i=1}^{n} \mu_{ij} > 0, for\ all\ j$ (4)

$0 \leq \mu_{ij} \leq 1, for\ all\ i, j$ (5)

where, j is a parameter that depends on the distribution of pixels in the cluster '*j*' and is assumed to be proportional to the mean value of the intra cluster distance.

In Eqn. (2), the first term demands that distances between feature vectors and its prototypes be as low as possible, while the second term forces the *uij* to be as large as possible, thus avoiding the trivial solution. Generally, η_j depends on shape and average size of the cluster '*j*' and its value may be computed as:

$$\eta_j = K \frac{\sum_{i=1}^{n} \mu_{ij}^{m} d_{ij}^{2}}{\sum_{i=1}^{n} \mu_{ij}^{m}} \quad (6)$$

where K is a constant and is generally kept as 1.

After this the class membership *uij* is obtained as:

$$\mu_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^2}{\eta_j} \right)^{1/m-1}} \quad (7)$$

In case of PCM, this membership value represents the "degree of sharing" which is contrary in FCM which represents the "degree of belongingness or compatibility or typicality" (Krishnapuram and Keller, 1993). Overcoming this constraint gives higher accuracy of supervised classification using PCM as compared to that of FCM (Kumar and Ghosh, 2006, 2007). Also, PCM, as a supervised classifier, works better in case of untrained classes, when compared to FCM (as supervised classifier) (Foody, 2000). Untrained classes are those classes which are present in the image but are not known to the analyst; hence, the classifier is not trained with that unknown class.

Thus, the advantages of PCM over FCM are the motivation behind selecting PCM as soft classification approach in this research. Further, PCM can handle noise and outliers (Krishnapuram and Keller, 1996). Noise and outliers affect the prototype parameters i.e. cluster means.

Kernels

We can characterize, identify and classify the land covers with improved accuracy and robustness with the help of information contained in multispectral data. In the literature of remote sensing, there are a number of supervised and unsupervised methods that have been developed for multispectral image.

In recent years, few kernel methods, like Support Vector Machines (SVMs) or Kernel Fisher discriminant analysis, have demonstrated excellent performance in multispectral data classification in terms of accuracy and robustness. The kernel methods can handle large input space efficiently; work with a relatively low number of labeled training samples, and deal with noisy samples in a robust way. Such properties of kernel methods make them well-suited to handle the problem of multispectral image classification.

In case of non-linear surfaces, a feature vector is mapped into a higher dimensional Euclidean space via a non-linear vector function which was quite expensive. To cope up with this problem, the concept of Kernel function is being introduced.

In this research work, 4 kernels have been incorporated. The kernels used in this work are grouped under local kernels. A brief elaboration about this category is as follows:

Local Kernels:

Only the data that are close in the proximity of each other's have an influence on the kernel values. All local kernels are based on a distance function. In this work following local kernels are used:

Gaussian Kernel with Euclidean Norm:

$K(x, x_i)=exp(-0.5(x-x_i)A^{-1} (x-x_i)^T$

where A is a weight matrix known as norm and represented as:

Euclidean Norm A=I

Radial:

$K(x, x_i)=exp(-||x-x_i||^2)$

KMOD:

$K(x, x_i)=exp(1/1+(-||x-x_i||^2)$

Inverse Multiquadric:

$K(x,x_i)= 1/(||x-x_i||^2 +1)^{0.5}$

Methodology

In this research work, kernel based fuzzy algorithms have been evaluated for extracting material of interest at sub-pixel level. As well as the output generated from these algorithms in the form of land use and land cover map have been evaluated.

Study area for this research work has been identified in the surroundings of **Radaur,** a town in Yamuna Nagar district in the Indian state of Haryana.

The study area has the following land use and land cover classes as shown in Table 1.

Table 1: Land Use/Land Cover Classes present in study area

S. No.	Land Use/Land Cover Categories
1	Mango Orchards
2	Urban Areas
3	Water Bodies
4	Wheat

Volume : 4 | Issue : 12 | December 2015 • ISSN No 2277 - 8179

The block diagram, as shown in Figure 1, states the procedure that is being adapted in image classification. The multispectral geocoded data incorporated is an 8-band image. Further using the training data, soft classification technique is being used. Firstly image is classified using Possibilistic c-mean approach and secondly using the kernels classification is being implemented, to measure the accuracy of membership values.

The classification has been done using an in-house application program. Using this, in every pixel click an output is being generated that is membership value of that particular pixel, to a specific signature class; with this a preliminary assessment has been formulated for measuring the belongingness of the pixel to a particular class.

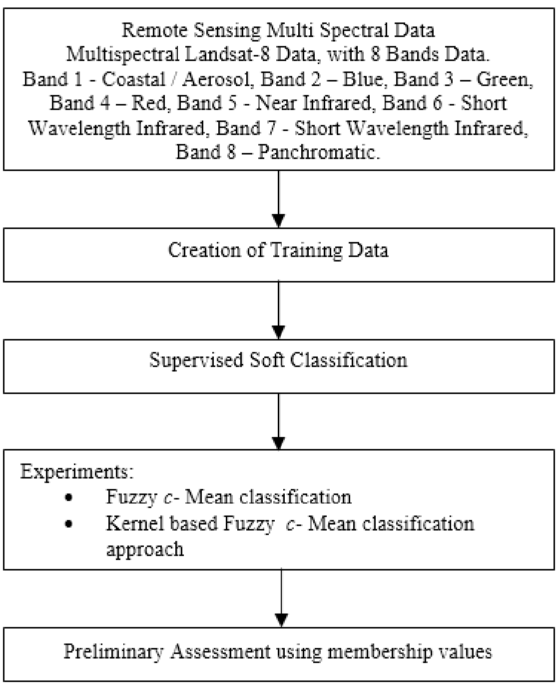


Figure 1: Flow diagram for methodology to be adopted

Result and Discussion

In this preliminary study of kernel based classifiers, we have taken four feature classes for classifying the image. Here, we have implemented and got the different output images for every classifier with a combination of four output images, where every image represents only single feature.

With the help of visual interpretation and ground reality, some deductions and assessments have been formulated to identify the belongingness of the particular pixel to the particular class, as shown in Figure 2. Output images procured after applying the classification techniques using 4 Kernels with the PCM as well as using standard PCM, every output respective to the feature class has been highlighted.

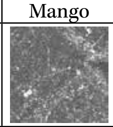
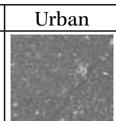
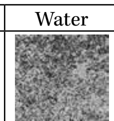
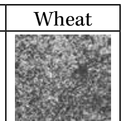
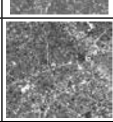
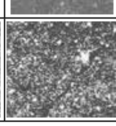
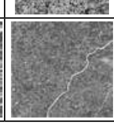
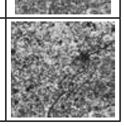
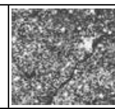
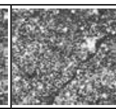
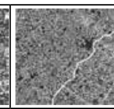
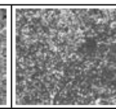
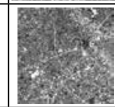
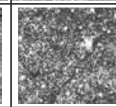
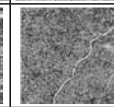
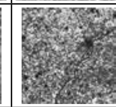
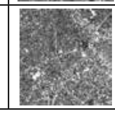
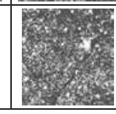
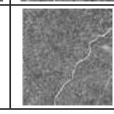
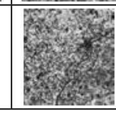
Kernels	Mango	Urban	Water	Wheat
PCM				
Gauss (Euclidean)				
Inverse Multi-quadratic				
KMOD				
Radial				

Figure 2: Tabular representation of output images corresponding to the type of classifiers used.

We have also assessed the data, quantitatively, by calculating the membership values for every feature class, and measuring the degree of belongingness to that feature class. As shown in Table 2, following conclusions have been dropped:

For the feature class Wheat, the maximum membership value is 1.00 of the kernel Inverse Multiquadratic i.e. the output image is showing more accurate results of finding the patches of wheat in the imagery. On the other hand, for the same feature the membership value of standard PCM is much less, leading to low degree of belongingness.

For the feature class Mango the maximum membership value is 0.99 of the kernel Inverse Multiquadratic i.e. the output image is showing more accurate results of finding the patches of mango in the imagery. On the other hand, for the same feature the membership value of standard FCM is much less, in comparison to other remaining kernels, leading to low degree of belongingness.

Similarly, for the feature class Urban the maximum membership value is 1.00 of the kernel Inverse Multiquadratic i.e. the output image is showing more accurate results of finding the patches of wheat in the imagery. On the other hand, for the same feature the membership value of standard FCM is much less, in comparison to other remaining kernels, leading to low degree of belongingness.

For the feature class Water the maximum membership value is 0.99 of the kernel Inverse Multiquadratic i.e. the output image is showing more accurate results of finding the patches of wheat in the imagery. On the other hand, for the same feature the membership value of standard FCM is much less, leading to low degree of belongingness.

Table 2: Membership Values of different classifiers related to the classes

Kernels	Mango	Urban	Water	Wheat
PCM Classifier	0.93	0.09	0.98	0.86
Gaussian (Euclidean norm based on PCM Classifier)	0.95	0.93	0.84	0.98
Inverse Multiquadric based on PCM Classifier	0.99	1.00	0.99	1.00
KMOD based on PCM Classifier	0.70	0.6	0.60	0.63
Radial based on PCM Classifier	0.96	0.94	0.86	0.99

Conclusions

In this paper, we propose a new methodology that integrates the fuzzy approach with kernel methods.Fuzzy Classifiers were more effective than the earlier approaches, but noise detection and accuracy assessment was not that much robust. Incorporation of kernel methods have rectified such issues.The paper has covered up the implementation of local kernel methods only,resulting to major inclination towards more accuracy.

REFERENCE

[1] Anil Kumar, S. K. Ghosh, V. K. Dadhwala, "Study Of Mixed Kernel Effect On Classification Accuracy Using Density Estimation", Proceedings of the ISPRS Commission VII Symposium 36 (Part 7). | [2] Bezdek, J. C., Ehrlich, R., & Full, W., 1984. FCM: The Fuzzy c-Means Clustering Algorithm. Computers & GeoSciences, Vol. 1, pp.191–203. | [3] Burges, C.J.C. 1996, Simplified support vector decision rules. In Lorenza Saitta, editor, Proceedings of the Thirteenth International Conference on Machine Learning, pages 71–77, Bari, Italy, Morgan Kaufman. | [4] Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A. and Vapnik, V. 1997, Support vector regression machines. Advances in Neural Information Processing Systems, 9:155–161. | [5] Foody, G. M., 2000. Estimation of sub-pixel land cover composition in the presence of untrained classes, Computers & Geosciences, Vol. 26, No. 4, pp.469–478. | [6] Girosi, F. 1998, An equivalence between sparse approximation and support vector machines. Neural Computation; CBCL AI Memo 1606, MIT. | [7] Krishnapuram, R. and Keller, J.M., 1993. A possiblistic approach to clustering. IEEE Transactions on Fuzzy Systems, Vol. 1, pp.98-108. | [8] Krishnapuram, R. and Keller, J.M., 1996. The possibilistic c-means algorithm: Insights and recommendations, IEEE Transactions on Fuzzy Systems, Vol.4, No. 3, pp.385-393. | [9] Kumar, A., Ghosh, S.K., and Dadhwal, V.K., 2006a. ―Study of Mixed Kernel Effect on Classification Accuracy Using Density Estimation‖, in Proceedings of Mid-term synposium ISPRS, ITC- The Netherlands, May 8-11, 2006. | [10] Mukherjee, S., Osuna, E. and Girosi, F. 1997, Nonlinear prediction of chaotic time series using a support vector machine. In Proceedings of the IEEE Workshop on Neural Networks for Signal Processing 7, pages 511–519, Amelia Island, FL. | [11] Osuna, E. and Girosi. F. 1998, Reducing the run-time complexity of support vector machines. In International Conference on Pattern Recognition (submitted). | [12] SchÖlkopf, B., Burges, C. and Vapnik, V. 1995, Extracting support data for a given task. In U. M. Fayyad and R. Uthurusamy, editors, Proceedings, First International Conference on Knowledge Discovery & Data Mining. AAAI Press, Menlo Park, CA. | [13] SchÖlkopf, B., Smola, A. and Müller, K. R. 1998, Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation. In press. | [14] Schmidt, M. 1996, Identifying speaker with support vector networks. In Interface 96 Proceedings, Sydney. | [15] P.H.Swain and S.M.Davis,Remote Sensing :The Quantitative Approach, New York: Mc Graw Hill | |