VIKAS SRIVASTAVA
G S P SINGH
and
S K SHARMA

Optimization of pixels with data mining and image segmentation for landuse land cover maximum likelihood classification algorithms

The image segmentation is a classification issue where every pixel is classified into different groups. A variety of image segmentation methods have been developed for image processing and computer applications of pixels of satellite data. The data mining of data of remote sensing data in which it assumed that optimization of pixels is typically labelled as single land cover and land use class. The pixel level and texture features are selected from the transformed colour image. The pixels are classified the spectral variables and informations. These pixels are classified by two methods of unsupervised and supervised classifiers algorithms. Pixels satellite images are natural grouping of digital value using maximum likelihood and self-organizing data analysis (ISODATA) algorithms. An analyst selects training sample sites with known class types and representative samples. The pixels are labelled by decision rules by their spectral properties with maximum likelihood classifier (MLC) algorithms.

1. Introduction

The quantitative analysis of the data sets in data mining, image segmentation and pattern are mix methods of traditional method and advanced algorithms. The aims of segmentation grouping into classify the image into homogenous regions based upon pixel characteristics. The data mining deploys many different algorithms techniques. These algorithms are classified into six types namely:

i. Level slicing
ii. Thresholding
iii. Edge detection
iv. Region based segmentation
v. Template matching
vi. K-mean
vii. Maximum likelihood

The problems are solved by data mining in special domain through optimization for arranging the allocation of resources. The analysis and optimization by the classification are image segmentation important in the study of remote sensing in decision orient applications. The spatial patterns of coal fires models have been developed by terrain, soil and estimated with the use in the data sets that have been chosen and can further explore analytically with an optimization in giving an efficient decision making by classifications. The spatial patterns of coal fires models have been developed by terrain, soil and vegetation estimated with the help of RS and GIS. Surface attributes are delineated from the satellite imagery, DEM and toposheet. Land use vegetation data characteristics are influenced by infiltration, erosion and evapotranspiration. The classification arranges pixel of image on the basis of spectral digital numbers within different bands. The pixel indices comprise the greenness, rainfall, infiltration, water availability and underground conditions and they provide texture or pattern associated with objects. The paper discusses the integrated approach to classify the landuse spatial data in the form of pixels. The data mining of pixels is the process for the classification of pixels radiation or the reflectance property. In pixel based approach, each pixel is classified on the basis of the digital numbers. These digital numbers are the associations of these texture statics derived from the satellite imagery as attributes of the land cover classifications for the coal seam fires.

2.0 Classification

Image segmentations or pattern recognitions are applications for the study of remote sensing data in decision oriented applications. The classification techniques are parametric in nature. In parametric classification techniques, maximum likelihood algorithms have been used. The software involves training set and ancillary data to define rules of classifications. The LISS III sensor data of Resources at I is used for K-means and Maximum Likelihood algorithms for the unsupervised and supervised classifiers for pixel class alignment. The image classification is assigning segments of classes to images. Image segmentation is the classiûcation of an image into different groups using clustering methods. Quantitative analysis is one of ways to make a decision and it divides an image into discrete regions where the pixels have high similarity or the contrast. Image segmentations or pattern

Messrs. Vikas Srivastava, G S P Singh and S K Sharma, Department of Mining Engineering, IIT (BHU), Varanasi, U.P., India

recognitions require many specializations including civil engineering analysis, diagnosis, image processing, etc. Maximum likelihood and K-means is simple and computationally faster than the hierarchical clustering and it can be done in ERDAS imagine software of Leica. It can also work for large number of variables. Different values of the initial centroids would result in the different clusters. Thus, the selection of a proper initial centroid is an important task. The image segmentation is an important tool for sensing of data and it is used to extract the region of interest from the background data. Land parcels are the segments with the form of pattern recognitions of different techniques and process outputs.

2.1 ML CLASSIFICATION (MLC)

MLC derived from the Bayes theorem eq (1), states that the a posteriori distribution P(i|¢), is often treated as a normalization i.e., the probability that a pixel with feature vector ¢ belongs to classification, is given by:

$$P(i|¢) = P(¢|i)\ P(i)/P(¢)/P(¢) \quad \ldots 1$$

where, P (¢|i) is the likelihood function,

P(i) is the a priori information

P(¢) is the probability

The pre-classified values of the target variable are training set of data. A data mining pattern is constructed using the training samples provided in the training data set. The value of the target variable is hidden temporarily from the models, performs classification according to the patterns and structure it learnt from the training set. The efficacies of the classifications are then evaluated by comparing them against the true values of the target variable.

MLC is applications of supervised models. The pixel values are assumed for various classes. The mean of feature values of each subject and the feature values as two groups of dependent random variables. Methods apply to simulate image data and our experimental results to the proposed model could improve the supervised MLC segmentation results when there are considerable differences across subjects. ML classification method is based on the Bayes theorem. The vector and the matrix are the input functions and can be estimated from the training pixels of class. We used ML to classify a diverse tropical land covers recorded from the Resources at I satellite through maximum likelihood classifier.

2.2.1 ML classification using ERDAS imagine software

Image classified in ERDAS by supervised classification. Table 1 shows percentage of pixels in 25 classes

2.2.2 ML classification through SPSS software

The pixel values are classified into supervised classifications (Fig.1) and the SPSS software is used for the supervised classification. The lowest mean is Band 4 and the lowest standard deviation is in the Band 3.

TABLE 1

	Class	Pixels	Percentage (%)
1	Class1	1129	0.05621135
2	Class2	98321	4.89526714
3	Class3	28875	1.43764647
4	Class4	2520	0.12546733
5	Class5	22527	1.1215883
6	Class6	18938	0.94289693
7	Class7	75	0.00373415
8	Class8	127056	6.32594321
9	Class9	31098	1.54832658
10	Class10	13545	0.67438689
11	Class11	28724	1.43012839
12	Class12	69211	3.44592035
13	Class13	67003	3.33598707
14	Class14	12586	0.6266396
15	Class15	44185	2.19991028
16	Class16	26125	1.30072776
17	Class17	691039	34.4058798
18	Class18	6078	0.30261525
19	Class19	154536	7.69413455
20	Class20	10435	0.51954427
21	Class21	17521	0.87234645
22	Class22	8550	0.42569272
23	Class23	55017	2.73922064
24	Class24	402112	20.0206025
25	Class25	71285	3.54918195
		2008491	100

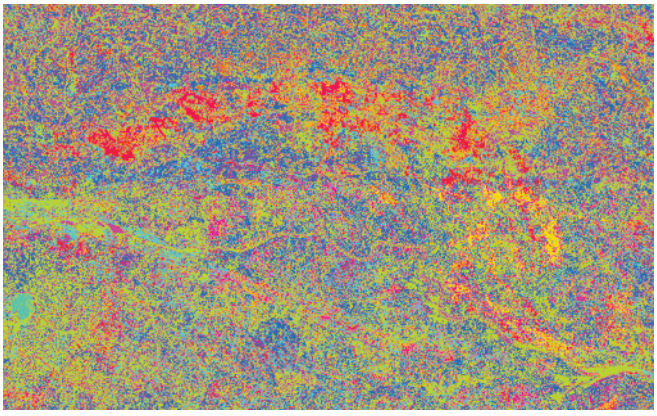

Fig.1 Supervised image after applying maximum likelihood algorithms

A scree plot (Fig.2) displays the Eigen values associated with a component or factor in descending order versus the number of the component or factor. A scree plots can be used in principal components analysis and factor analysis to visually assess which components or factors explain most of the variability of the data mining of the data.

TABLE 2: DESCRIPTIVE STATISTICS OF DATA

	Descriptive statistics			
	Mean	Std. deviation	Analysis N	Missing N
Band 1	1.12E+04	27926.712	255	76
Band 2	1.02E+04	21666.372	255	58
Band 3	7898.56	16733.643	255	1
Band 4	7867.58	21604.899	255	0

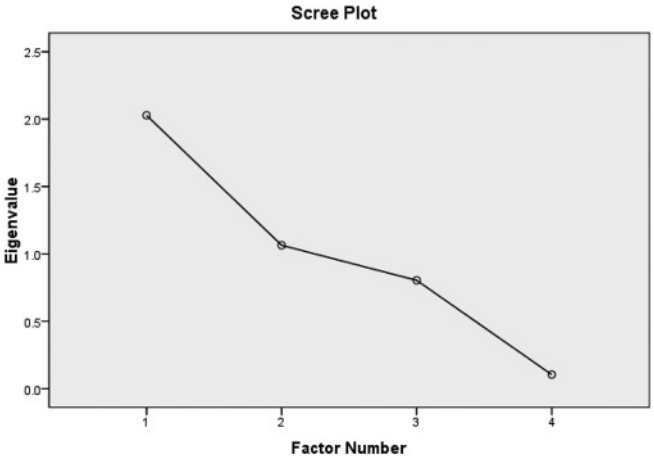

Fig.2 Scree plot between eigen value and factor number

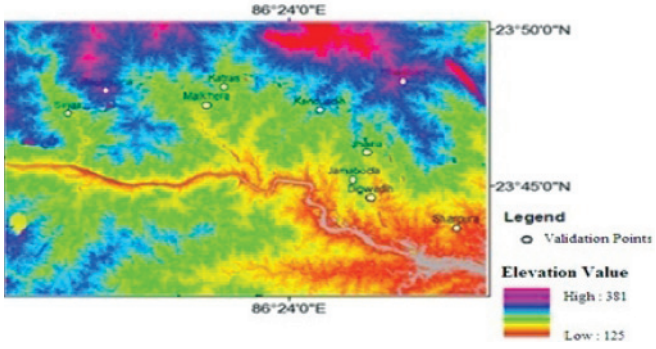

Fig.3 DEM (digital elevation model) of SRTM (shuttle radar yopography

3.0 Validation

Ground truths and pixel sampling by high resolution image with point data are used to validate the classification result with the help of hand held GPS (global positioning system).

TABLE 4: CORRELATION MATRIX

		Band 1	Band 2	Band 3	Band 4
Correlation	Band 1	1	0.175	-0.066	0.849
	Band 2	0.175	1	-0.15	0.415
	Band 3	-0.066	-0.15	1	0.028
	Band 4	0.849	0.415	0.028	1
Sig.(1-tailed)	Band 1		0.009	0.19	0
	Band 2	0.009		0.018	0
	Band 3	0.19	0.018		0.328
	Band 4	0	0	0.328	

a. Determinant = 0.181

TABLE 5: REPRODUCED CORRELATIONS

		Band 1	Band 2	Band 3	Band 4
Reproduced correlation	Band 1	.759a	0.175	0.058	0.849
	Band 2	0.175	.999a	-0.15	0.415
	Band 3	0.058	-0.15	.032a	0.028
	Band 4	0.849	0.415	0.028	.999a
Residual	Band 1		-2.50E-05	-0.124	1.39E-05
	Band 2	-2.50E-05		0	1.09E-07
	Band 3	-0.124	0		0
	Band 4	1.39E-05	1.09E-07	0	

Extraction method: Maximum likelihood.

ka. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 1 (16.0%) no redundant residuals with absolute values greater than 0.05

The models of pixels matrix values are simulated by data mining segmentation and validated in field (model in Fig.3). The accuracy assessment of the map was done by the ground truths control points. The ground truth positions were imposed over DEM (digital elevation model) of SRTM (shuttle radar topography mission) USGS (United States Geological Survey). The accuracy of landuse type to the classification map is 92%. The accuracy is increased about 12% by the applying maximum likelihood classification method on data.

Mission, USGS - United States Geological survey) data with color coding of elevations.

TABLE 3: VARIANCE ANALYSIS OF PIXELS

Total Variance explained						
Factor	Initial eigen values			Extraction sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.028	50.698	50.698	1.791	44.782	44.782
2	1.064	26.594	77.292	0.997	24.936	69.718
3	0.804	20.092	97.384			
4	0.105	2.616	100			

Extraction Method: Maximum Likelihood

4. Conclusion

In our analysis the spectral image provides the classified landuse land cover of coal seam fires area. The optimization of the pixels by the data mining classifications for the decision making is advantageous in coal seam fire mapping with vegetation growth. The data mining scrutinizes the facts of coal seam fire for advantageous requisite informations. Recode algorithms were applied to hidden patterns to obtain the class recognitions from the pixels by segmentation classifier. Decision making process derived from data sets should use to accelerate the transformation of information in different stages. It observed that the segmentation of data by the Maximum Likelihood improves the interpretations from the original image. The segmented result is validated in field and it proposed that MLC algorithm has better segmentation processes.

References

1. Danneels, G., and Pirard, E., Haveith, H. B., (2007): Automatic Landslide, Detection from Remote Sensing Images using Supervised Classification Methods, IGARSS, pp. 3014-3017.
2. Liu, Xin. , Yetik, I. S. , (2010): A maximum likelihood classification method for image segmentation considering subject variability Medical Imaging Research Center, Illinois Institute of Technology, Chicago, IL, USA.
3. Prakash. A., Gupta R. P., (1999): Surface fires in Jharia Coal field, india their distribution and estimation of area and temperature from TM data, *Int. J. Remote Sens* 20, 1935-1946.
4. Reddy, C. S. S., Srivastava, S. K., and Bhattacharya, A. (1993): Application of Thematic Mapper short wavelength infra red data for the detection and monitoring of high temperature related geo environment features, *International Journal of remote sensing*,14 (17): 3125-3132.
5. Xie, J., and Tsui, H.T. (2004): Image segmentation based on maximum likelihood estimation and optimum entropy-distribution (MLE–OED), Pattern Recognition Letters, 25, 1133-1141.

Printed by Pradip Kumar Chanda at The Indian Press Pvt. Ltd. 93A Lenin Sarani, Kolkata 700 013 and published by him for Books & Journals Pvt. Ltd. from 62 Lenin Sarani, Kolkata 700 013.