A multi-subregions decision tree land cover classification approach using Landsat8 image

Li Feng[1], Liang Handong[2], Mi Xiaonan[3], Wei Aixia[1]

(1. Department of Disaster Prevention Engineering, Institute of Disaster Prevention, Sanhe 065201, China;
2. College of Geoscience and Surveying Engineering, China University of Mining and Technology, Beijing 100083, China;
3. Shanxi Climate Center, Taiyuan 030002, China)

Abstract: Coal fires burning caused serious environmental, economic and safety catastrophe in Wuda district, North China. The land cover change research helped to evaluate the extent of coal fire damage. The image data of Landsat8 satellite offered the possibility of detecting and studying land cover/use in coal fire area. Five subregions were divided from one Wuda image based on topographic, landform and land surface radiation characteristics. Corresponding to each subregion, five different decision tree models with different parameters were respectively constructed based on a general sole decision tree for the whole research area, which was built by spectral characteristics analysis, height, slope and infrared information. By contrasting with a general sole decision tree and other four common classification methods applied to the whole area, land cover accuracy of multi-subregions decision tree classification approach derived higher overall accuracy (87.63%) and Kappa coefficient (0.86) because subregions decreased land-cover confusions. In particular, the accuracy of building and coal ash classification mapping showed a marked increase.

Key words: land cover; classification; multi-subregions; decision tree; coal fire

CLC number: TP79 **Document code:** A **Article ID:** 1007-2276(2015)07-2224-05

Landsat8 卫星影像的多子区决策树土地覆被分类方法

李 峰[1],梁汉东[2],米晓楠[3],卫爱霞[1]

(1. 防灾科技学院 防灾工程系,河北 三河 065201;
2. 中国矿业大学 地球科学与测绘工程学院,北京 100083; 3. 山西气候中心,山西 太原 030002)

摘 要: 乌达矿区的煤火自燃造成了严重的环境、经济和安全灾害,对该地区的土地覆被变化研究有助于评估煤火灾害的影响程度和范围, 而 Landsat8 卫星影像为煤火区的土地覆被分类探测与研究提供了可能。依据乌达地区的地形、地貌和地表辐射特征划分 5 个子区域,基于通用单决策树模型,利用光谱特征分析、高程、坡度和热红外信息对每个子区域分别构建 5 种不同参数的决策树模型。相比通用单决策树模型以及其他 4 种普通分类方法,因减少了土地覆被的混淆度,多子区决策树模型土地覆被分类的整体精度和 Kappa 系数更高,分别达到 87.63%和 0.86,尤其是建筑物和煤灰的分类精度有较为明显的提升。

关键词: 土地覆被; 分类; 多子区; 决策树; 煤火

收稿日期:2014-11-05; 修订日期:2014-12-10

基金项目:河北省高等学校科学研究计划重点项目(ZD2014203);中央高校基本科研业务费资助项目(ZD2014203);
国家自然科学基金(41371449)

作者简介:李峰(1979-),男,讲师,博士,主要从事遥感、车载/机载 LiDAR 点云处理方面的研究。Email:lif1223@aliyun.com

0 Introduction

The Wuda coal field area in Inner Mongolia, China, had once been one of the largest coal fire regions worldwide[1]. Coal seam fires posed a serious threat to local environment and human health. In order to evaluate ecology environment influences induced by coal fire, it is necessary to study local land cover changes according to remote sensing measurements. A few references about land cover in coal fire field are exhibited as follows. Prakash and Gupta[2] acquired land cover map of Jharia coal field in India using GIS and remote sensing visual interpretation techniques, but this method limited the overall mapping accuracy. Gao et al[3] draw a conclusion that using the Object-Oriented Classification (OOC), the overall accuracy was higher than the accuracy obtained using the pixel-based classification, and the user′s and producer′s accuracy of almost all the classes were also improved. Zhang et al[4] obtained Wuda coal field good classification accuracy based on multi-level classification approach.

Land cover information could be received by commonly used remote sensing image classification method, such as Maximum Likelihood Classification (MLC), Object-oriented classification, Support Vector Machines (SVM) and Spectral Angle Mapping (SAM). MLC considering spectral similarities was widely applied to land cover classification and this technique was continuously improved [5-6]. Object-oriented analysis built on characteristic texture and contextual information enables produce higher land cover map accuracy [7]. Adopting the method of structural risk minimization for class member discrimination, SVM minimized the probability of misclassifying the previously unseen data point drawn randomly from a fixed but unknown probability distribution. SVM usually performs better in terms of classification accuracy[8]. SAM mainly is applied to recognize minerals employing spectral characters[9]. Additionally, in the light of region features and objects properties, a whole image region could be divided into several subregions, and then they could be compounded into a new classification region after completing subregions classification[5]. The aim of this research is to build a multi-subregion decision tree model so as to fast achieve land cover classification with high accuracy results in a sophisticated coal fire field.

1 Study area and data

1.1 Study area

The study area, Wuda coal mine, is located in the southern part of Inner Mongolia, North China. It covers approximately 27km×27km, and the latitude of the study area ranges from 39°23′ to 39°39′ N and longitude 106°30′ to 106°49′ (see Fig.1). Physically the area is bounded by the Gobi Desert in the north and west, the Yellow River on the east, and south by the Helan Mountain. In 2004, investigation report from Beijing Remote Sensing Corporation, BRSC Showed that the extent of the coal fire areas was 3.5 million m^2 where included sixteen coal fire zones and four coal gangue fire zones in the Wuda coalmine area. Until 2009, BRSC made another more detailed investigation and concluded that the total area affected by coal fires was 4.75 million m^2 covering 13.6% of the total area of the Wuda syncline. The prolonged combustion of Wuda coal fire induced terrain and landform of mine zone to change dramatically.

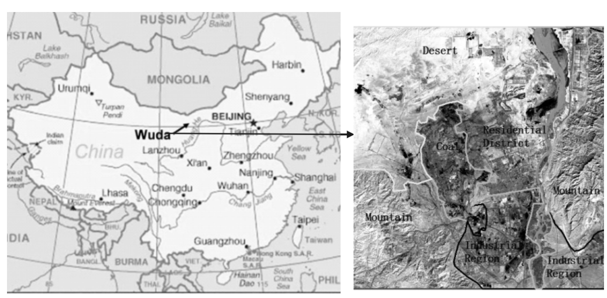

Fig.1 Location of the Wuda coalmine area and its subregions

1.2 Study data

Landsat8 satellite, launched by NASA in February, 2013, officially began normal operations on

May 30, 2013. Landsat8 carries two instruments: Operational Land Imager(OLI) sensor includes refined heritage bands(B2 to B7) with 30 m revolution, along with three new bands: a deep blue band (B1) for coastal/aerosol studies, a shortwave infrared band(B9) for cirrus detection, and a quality assessment band (B8). Thermal Infrared Sensor (TIRS) provides two thermal bands (B10 and B11) with 30m resolution, which is higher than TM and ETM+ sensors relatively with 120 m and 60 m resolution. Unfortunately, now TM and ETM+ sensors can't work any more because of aging and malfunction. The features provided by TIRS of Landsat8 enable more precisely identify high-energy radiation regions, such as desert and coal fire areas. Applying Wuda Landsat8 data acquired in September, 2013, this paper would study land cover classification in governance stage of Wuda coal fire.

2 Methodology

Figure 2 shows the methodology used in this study, which is explained in detail in the following sections, where decision tree classification is seen in section 2.4.

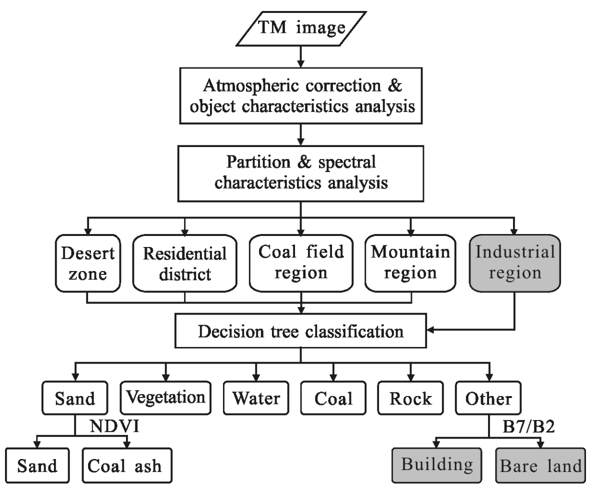

Fig.2 Flowchart of the methodology

2.1 Pre-processing of Landsat8 Data

The Landsat8 image was geometrically corrected with ground control points, resampled with bilinear method, and geo-referenced to UTM zone 48 with WGS84 spheroid. Radiometric calibration and atmospheric correction were applied to eliminate atmospheric affections. Then the good quality data was considered to perform land cover classification.

2.2 Subregions division

Land cover in the study area is highly heterogeneous, wind-blown sands and coal dusts are widely distributed in the study area. If merely one decision tree model is adopted, classification results' accuracy would be affected infinitely. Thus, according to topography, radiation and landform features (see Tab.1), split multi-subregions from one whole image with different decision tree models are expected to produce better classification effects. Split subregions include desert zone, residential district, coal field region, mountain region and industrial region (see Fig.1).

Tab.1 Subregions and their features

Region features	Subregions				
	Desert zone	Residen-tial district	Coal field region	Mountain region	Industrial region
Topogra-phy	Average height is 1 164 m, slope is lower	Average height is 1 093 m, slope is lower	Average height is 1 213 m, slope is higher	Average height is 1 428 m, slope is higher	Average height is 1 169 m, slope is lower
Radiation	Constant high thermal radiation	Moderate thermal radiation	High thermal radiation	Low thermal radiation	Moderate thermal radiation
Landform	Sand dunes	Buildings, vegetation and river	Bare burned coal	Bare rock	Coal piles, industrial equip-ments

2.3 Spectral characteristics analysis

Land cover types of Wuda district are classified into sand, vegetation, water, coal, rock, residential district (building) and bare land in terms of local land surface and spectral characteristics of Landsat8 image. (NDVI, Normalized Difference Vegetation Index), (MNDWI, Normalized Difference Water Index), (NDBI, Normalized Difference Building Index) are separately used to extract vegetation, water and partial building or coal types. Based on topographic height and its slope, Rock type will be quickly derived from DEM data in mountain region. Thermal infrared band (B10) can separate some bare land types from buildings according to radiation features. Except Band6 values are zero, other bands′ reflectance of Landsat8 image were counted from different training samples. Derived spectral profiles of eight land covers are shown in Fig.3, where the ratio between B7 (SWIR1) and Band2 (blue) can definitely identify Sand, Coal Ash, Rock and Bare Land types. The sum of Band4 (red), B5(Near infrared) and B7 can partly distinguish building and bare land types. Detailed classification rules are described as follows.

Vegetation:

NDVI=(B5−B4)/(B5+B4) >0.18;

Water: MNDWI=(B3−B7)/(B3+B7)<0.3;

Sand: 2.6<B7/B2<4;

Coal: NDBI=(B7−B5)/(B7+B5)>0.09;

Rock: Height>1 280 m,

Bare Land: B10>10.4 or (B10<10.4 and Not (4 800< (B4+B5+B7)<7 300)),

Residential district: 4 800 <(B4 +B5 +B7) <7 300 & Height>1 160 m.

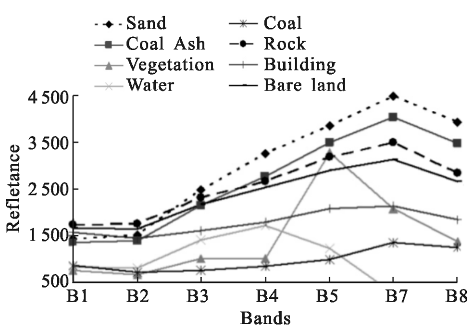

Fig.3 Spectral profiles of eight land covers in study area

2.4 Decision tree classification model

Firstly a general sole decision tree model (see Fig.4) is built using classification rules in section 2.3. Each subregion can use this decision tree model, but classification thresholds of each region need to slightly adjust respectively. The research is planned to implement by means of two schemes. One method is to directly apply a general decision model mentioned above to the whole study area; the other is to employ adjusted decision trees with different thresholds corresponding to multi-subregions. We found that it is difficult to identify coal ash from sand, as well as Building from bare land using a general decision tree. However, multi-subregions decision tree model can lightly fix this problem. For the first case, Coal ash and sand frequently occur on desert zone, residential district, coal field district and mountain region. The NDVI value between 0.09 and 0.15 is easily to detach coal ash from sand after previous sand type is determined. For the other case, the ratio between B7 and B2 from 0.4 to 1.4 can successfully recognize building from bare land after other type is confirmed only for industrial region. The both processes are demonstrated just in Fig.2.

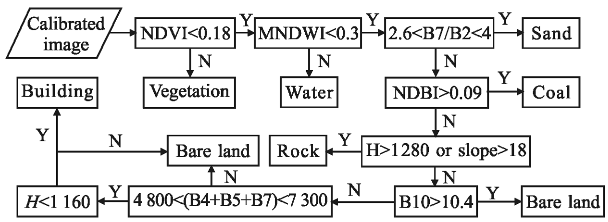

Fig.4 General decision model in study area

3 Accuracy assessments

Classification accuracy is evaluated using ground referenced data derived from field sampling, in which each category contains one hundred sample spots. The first method that only a general decision tree is applied to the entire region is tested using seven hundred sample data because no coal ash type is identified. The test results are displayed in Tab.2 and Fig.5(a). The second method applying different decision trees to corresponding subregions is verified by virtue of eight hundred sample data. The method′s results are shown in Tab.3 and Fig.5(b). Both tables represent two error matrixes which are formed with data from classification map and ground data. The diagonal entries represent correct classifications and the off-diagonal entries represent misclassifications. Important accuracy indexes such as the overall accuracy, Kappa coefficient and user's and producer's accuracy were calculated. In order to evaluate the accuracy of multi-subregions decision tree model, MLC, SVM, SAM and Object-oriented methods are also respectively applied to the same test region, their classification accuracies are listed in Tab.4.

Tab.2 Error matrix produced with a general decision tree classification results

Classification data	Reference data									User's accuracy/%
	Water	Bare Land	Sand	Coal	Rock	Building	Coal Ash	Vegetation	Total	
Water	85	0	0	0	0	0	0	0	85	100
Bare land	3	87	4	9	4	32	1	4	144	60
Sand	0	1	92	6	2	1	98	0	200	46
Coal	3	2	1	78	3	2	0	1	90	87
Rock	1	0	1	0	88	0	1	0	91	97
Building	2	7	0	0	0	59	0	1	69	86
Coal Ash	0	0	0	0	0	0	0	0	0	0
Vegetation	6	3	2	7	3	6	0	94	121	78
Total	100	100	100	100	100	100	100	100	800	
Producer's	85	87	92	78	88	59	0	94		
Overall accuracy: 72.88%, Kappa coefficient: 0.69										

Tab.3 Error matrix produced with multi-subregions decision tree classification results

Classification data	Reference data									User's accuracy/%
	Water	Bare Land	Sand	Coal	Rock	Building	Coal Ash	Vegetation	Total	
Water	85	0	0	0	0	0	0	0	85	100
Bare land	4	92	4	7	4	11	1	5	128	72
Sand	0	1	88	5	2	1	2	0	99	89
Coal	3	2	1	78	3	2	0	1	90	87
Rock	1	0	1	0	88	0	1	0	91	97
Building	1	2	0	2	0	80	0	0	85	94
Coal Ash	0	0	4	1	0	0	96	0	101	95
Vegetation	6	3	2	7	3	6	0	94	121	78
Total	100	100	100	100	100	100	100	100	800	
Producer's	85	92	88	78	88	80	96	94		
Overall accuracy: 87.63%, Kappa coefficient: 0.86										

Tab.4 Accuracies of other four classification methods

Methods	Overall accuracy	Kappa coefficient
MLC	76.00%	0.73
OOC	57.63%	0.52
SVM	73.50%	0.70
SAM	63.13%	0.58

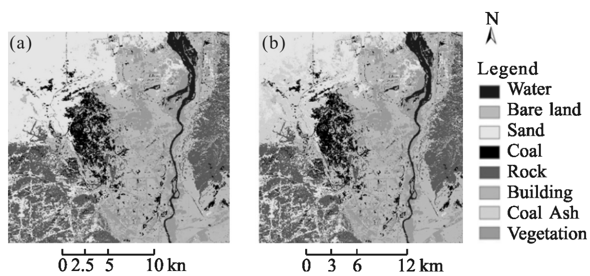

Fig.5 (a) Land cover produced by sole decision tree and (b) land cover produced by multi-subregions decision tree

4 Results and discussion

Based on the overall accuracies of the error matrices, approach of classification with different decision tree corresponding to subregions obtained an overall accuracy of 87.63% and Kappa coefficient is 0.86, which was obviously higher than classification approach with sloe decision tree corresponding to the whole region, with an overall accuracy of 72.88% and Kappa coefficient is 0.69. Therefore, one can say that this multi –subregions decision tree classification is superior to the sole decision tree classification.

The next comparison was on account of the user′s and the producer′s accuracies, which measure the commission and omission errors, respectively, for each land cover class. Evidently, sole decision tree classification failed to distinguish coal ash land-cover type. Thus this land-cover types had both 100% omission and commission error. While in multi-subregions decision tree classification, this type was distinguished, having 95% user′s and 96% producer′s accuracy. For building, multi-subregions decision tree classification had distinctly better accuracy than sole decision tree classification, having 94% user′s and 80% producer′s accuracy. For water, coal, rock and vegetation, two classifications kept the same good producer′s accuracy. For bare land, the producer′s accuracy of multi sub regions decision tree classification improved by 5% from 87% to 92% while that of Sand drop by 4% from 92% to 88%.

Combined with Tab.2, Tab.3 and Tab.4, MLC and SVM methods had better overall accuracies (76.00% and 73.50%) than sole decision tree classification which had 72.88% overall accuracy. However, multi-subregion decision tree model demonstrated best classification results with 87.63% overall accuracy and 0.86 Kappa coefficient. Setting multi subregions availably isolated confusing land cover types, reduced their false positive rates, and promoted their classification accuracies.

5 Conclusions

Land cover mapping and potential surface or subsurface coal fire detection is an important step to identify the distribution of surface coal fires, and furthermore to design fire-extinguishing schemes. By comparing the sole decision tree approach with multi-subregions decision tree approach, we found that multi-subregions decision tree approach can not only effectively improve building identification accuracy and overall accuracy of land cover, but recognize Coal Ash type more precisely. The multi-subregions classification reduces confusion of land surface, emphasizes thematic features in subregions, as a result, adopt different-threshold decision tree can enhance single class and overall land cover accuracy.

References:

[1] Kuenzer C, Zhang J, Sun Y, et al. Coal fires revisited: the Wuda coal field in the aftermath of extensive coal fire research and accelerating extinction activities [J]. ***International Journal of Coal Geology***, 2012, 102: 75–86.

[2] Prakash A, Gupta R P. Land-use mapping and change detection in a coal mining area–a case study in the Jharia coalfield, India[J]. ***International Journal of Remote Sensing***, 1998, 19(3): 391–410.

[3] Gao Y, Mas J F, Maathuis B H P, et al. Comparison of pixel-based and object-oriented image classification approaches–a case study in a coal fire area[J]. ***International Journal of Remote Sensing***, 2006, 27(18): 4039–4055.

[4] Zhang C, Guo S, Guan Y, et al. The approaches of land cover classification of the Wuda coal fire area[J]. ***Journal of Remote Sensing***, 2011, 15(1): 148–161. (in Chinese)

[5] Zhou Q, Robson M. Contextual information is ultimately necessary if one is to obtain accurate image classifications [J]. ***International Journal of Remote Sensing***, 2001, 22: 3457–3470.

[6] Prol-Ledesma R M, Uribe-Alcantara E M, Diaz M O. Use of cartographic data and Landsat TM images to determine land use change in the vicinity of Mexico City [J]. ***International Journal of Remote Sensing***, 2002, 23 (9):

1927–1933.

[7] Blaschke T, Strobl J. What's wrong with pixels? Some recent development interfacing remote sensing and GIS [J]. ***GeoBIT/GIS***, 2001, 6: 12–17.

[8] Song X F, Duan Z, Jiang X G. Comparision of aritificial neural networks and support vector machine classifiers for land cover classification in northern China using a SPOT–5 HRG image [J]. ***International Journal of Remote Sensing***, 2012, 33(10): 3301–3320.

[9] Rowan L C, Mars J C. Lithologic Mapping in the Mountain Pass, California area using advanced spaceborne thermal emission and reflection radiometer(ASTER) data [J]. ***Remote Sensing of Enviroment***, 2003, 84: 350–366.

[10] Wang K, Qi H P. Researching on the assorting methods of land utility and overburden remote sense[J]. ***Xi'an University of Architecture***, 2008, 34(5): 353–354. (in Chinese)