Land Cover Classification and Winter Wheat Mapping Using Data Mining Strategy*

He Guojin  Zhang Xiaomei  Jiao Weili  HU Deyong  Wang Wei  CHENG Bo, XIANG Bo
China Remote Sensing Satellite Ground Station, Chinese Academy of Sciences
No. 45 Bei San Huan Xi Road, P.O.Box 2434, Beijing 100086, P. R. China
gjhe@ne.rsgs.ac.cn

Abstract: With the increasing sensors available, the knowledge gap between satellite data and application needs is continuously increasing. Aimed at the information demands for mapping land resources, a study on the key technologies of intelligent information processing for satellite remote sensing is presented based on data mining and knowledge discovery strategy as well as information theory, which is suitable for Land cover classification and feature extraction. Significant results have been achieved by performing a test on land cover classification in Minshan area, which is the most important ecosystem region of panda distribution in China, together with an experiment on extracting the winter wheat distribution and the status of agricultural structure adjustment of Beijing from multi-temporal Landsat images.

Key words: Multi-Source data, land cover Classification, Feature Extraction, Data Mining.

1. Introduction

Earth observation sensors, especially those aboard satellites, continuously gather enormous amounts of data. The number of data satellites produced in a month now would have taken ten years to generate a decade ago. However, the most of the data is archived without ever being analyzed or used due to the ability of acquiring large quantities of data having greatly outstripped the ability of efficiently extracting information from it. This has resulted in the fact that, on the one hand, more and more satellite remotely sensed data is available, both optical and SAR data, which allows many applications, such as land cover classification and crop monitoring on earth surface to be practical, on the other hand increasing large amounts of satellite remotely sensed data leads to difficult questions about how the data is going to be processed to extract the real information necessary in terms of different kinds of applications. Current information processing capabilities are keeping up with the flow, however, the extraction of information from images is often experts based, which has contributed to keep a wide knowledge gap between the current data offering and the real need for information. This issue seems becoming even more challenging due to new missions and even constellations being carried out with varieties of higher resolution sensors to meet a wide range of application requirements [1]. Therefore, there is a strong need to develop new approaches, systems to deal with such problems, with applications driven focused, reliable, timely, understandable information extraction being particularly important. As a matter of fact, it has been becoming a topic with increasing international interest.

In recent years, data mining and knowledge discovery technology has become increasingly common in intelligent information processing domains. With its concepts and capabilities of digging out the knowledge from the flow data, it also gives potentials to treat with remote sensing information extraction, in which such technique seems to be particularly useful for land cover classification and thematic mapping applications. Aimed at meeting the information demands for mapping land resources, this paper addresses a study on the key technologies of intelligent information processing for satellite remote sensing based on data mining and knowledge discovery strategy as well as information theory, focusing on Land cover classification and feature extraction, and brings significant results by taking the local conditions into account.

This paper is organized as four sections. The information processing chain is firstly given and described in section 2. Section 3 presents two study cases on land use classification in Minshan area, and winter wheat monitoring in Beijing respectively, followed by discussion and conclusions in section 4.

2. Methodology

2.1. Information Processing Chain

“Data mining is a process concerned with uncovering patterns, associations, anomalies, changes and statistically significant structures and events in data. It not only can help us in knowledge discovery, that is, the identification of new phenomena, but also is useful in enhancing our understanding of known phenomena ”[2,3]. Different from relational or flat-file database, satellite remote sensing data is a kind of special scientific data, available as images, embodying information of spatial, spectral and temporal. Hence, the information mining process of such data should
have optimal chain and approaches.

Relying on information theory based analysis of remotely sensed data, a processing chain (shown in Fig.1) for remote sensing image information mining and knowledge discovery is presented in this paper. It is a process mainly involving data selection, data pre-processing, pattern recognition, and knowledge evaluation. Here, we give an emphasis on data preprocessing, and regard knowledge based data selection as a key step of the mining process.

Fig. 1. RS Data Mining Chain

2.2. Methodology Description

As mentioned above, remote sensing image contains information of spatial, spectral and temporal. An important character of remote sensing information different from or better than human eyes is that it can convert some “unseen” information to visible images. Taking domainial applications into account, the information in remote sensing is often time dependent. As addressed in information theory, information quantity is far related with time. On the one hand, information has uneven distribution in different period of time; on the other hand, the time longer, the information quantity larger. In other words, the measurement of remote sensing to a specific object will be much more precise when greater quantity of sequential temporal satellite images are used. For these reasons, J.Albertz(1990) presented a method of improving image spatial resolution by multi-temporal satellite data fusion; the research results from Badhwar(1982), Townshend(1987), Wagner(1993), Wolter(1995) and Mary(1997) indicated vegetation and agriculture classification accuracy would increase with more multi-temporal satellite images being applied. However, the more images used means the greater costs, both satellite data spending and image processing. This is often unpractical, and not allowed. Hence, further attentions should be paid to satellite data selection, which will help to gain the most information in limited temporal images with resulting in a precise
knowledge discovery. That is why we bring temporal data selection into the category of remote sensing data mining and knowledge discovery. By reasonable temporal data selection, satellite remote sensing images for mining are under target.

Given sets of target data, the preprocessing is a critical issue that should be firstly addressed, prior to the application of the pattern recognition process. The basic operation in this step is to make sure the gathering of sensed data, RS and Non RS ones, to be eventually transmitted to a form for information mining. The key challenge in such data transmitting is enhancing the weak information representation regarding to specific objects by fusing the data sets from different sensors, with physical and domainal understanding of the data sets being particularly important.

For land use/cover classification and thematic mapping, accuracy or precision is much concerned by users, although timely information providing is also extremely important. Hence, the pattern recognition step poses several challenges as well, especially in the earth observation era with satellite images of high temporal, spatial and spectral resolution. With respect to classification methods, statistical and data-driven methods, based on training datasets are usually used for remote sensing classification, of which Maximum-likelihood is the most robust pixel based classifier. However, these kinds of classifiers are becoming less powerful facing the new texture abundant sensor images like SAR and Quickbird data, as well as the condition when non RS data is involved. Note that many attempts are being performed worldwide in automatic, intelligent image processing, aimed at providing the required information to users with quick response, high quality. New algorithms, novel approaches from machine learning, artificial intelligence and other related domains have been considered and adopted to automatic image feature extraction. Rule based methods are proved to be significant in mining remote sensing information, thanks to decision tree approaches and related. Knowledge evaluation is to validate the information drawn from the data is sound to domain applications, which will affect final decision.

3. Experimental Cases

3.1. Land Cover Classification in Minshan Area

In this case, a land cover classification method in high mountain area that provides satisfactory results with multi-source data is addressed. Minshan area, locating in Sichuan province is the most important ecosystem region of panda distribution in China, where the mountains are very high and appear to be heavy shadows in Landsat images. The land cover classification accuracy is not satisfied from traditional image classification approaches like Maximum likelihood due to the effect of heavy shadows and the radiometric limitation of Landsat TM/ETM+ data.

![Fig.2. The classification map of Jiu-zhai-gou county](image-url)
The classification approach presented in this paper incorporates ancillary data such as the knowledge of vegetation distributing vertically with altitude, DEM and some land use maps to improve the classification accuracy. Here, how to integrate the different kind of ancillary information is the key step in the classification process. Note that Expert System (ES) has this kind of capabilities, but results from difficulty in obtaining knowledge used in classification. Thus, data mining strategy has been introduced for decision making of classification rules, with using the Quinlan’s ID3 decision tree algorithm [4,5].

Classification result shows that the rule-based land cover classification accuracy is much higher than that of Maximum Likelihood method. Fig.2 is a classification map of Jiu-Zhai-Gou county, one of the seven counties in the test site.

3.2. Winter Wheat Mapping and Agricultural Planting Structure Monitoring

In recent years, there appears to be a rapid adjustment of agricultural planting structure in Beijing. In order to provide the real-time information for monitoring the planting structure adjustment, a project regarding to intelligent information processing on satellite remote sensing images has been carried out, by following the processing chain given in Fig.1. A temporal data selection principle for target data decision was firstly addressed in terms of physical analysis on the sequential temporal Landsat images, followed by a study concerning data preprocessing for enhancing dynamic change information based on data fusion approaches, which is particularly suitable for detecting the information of planting structure adjustment. In addition, a data mining strategy based feature extraction method was also presented.

Significant results, both for change detection and thematic mapping, were achieved by performing a test on extracting the winter wheat distribution of Beijing from multi-temporal Landsat images, together with the status of agricultural structure adjustment. The red colour in Fig. 3 shows the winter wheat distribution of Beijing in 2004.

![Fig. 3. Winter wheat distribution of Beijing in 2004(red colour).](image)

4. Concluding Remarks

In 21st century, more and more earth observation data from different satellite sensors is available for our application
purposes, with multi-temporal, multi-spatial and multi-spectral. However, only small percentage of archived data has been analysed because of the limited information processing ability, which has resulted in a wide knowledge gap between satellite data and application needs. The capability and efficiency of information extraction is increasingly becoming one of the critical issues that prevents from further applications of satellite remote sensing. Therefore, it is a great important and challenging topic to develop new technologies for solving such a problem.

Data mining and knowledge discovery technology seems to be a potential option for exacting knowledge from flow satellite data. With attempt to meet the information demands for land use classification and thematic mapping, a research on intelligent information processing for satellite remote sensing is addressed based on data mining strategy. Significant results have been achieved by performing tests on both land use classification in Minshan area, southern-western China and agriculture monitoring in Beijing. Note that this is only the first step. Further study should be considered to focus on an image information mining system for satellite remote sensing applications, with content based searching of images being particularly challenging.

References: