Paper presented at the 22nd Asian Conference on Remote Sensing, 5 - 9 November 2001, Singapore.

THE STUDY OF RELEVANT TECHNIQUES FOR INTELLIGENT INTEGRATION OF GIS AND RS

Chunyu Zhao , Lingkui Meng, Changqing Huang, Heping Pan
School of Remote Sensing Information Engineering
Wuhan University, 129 Luoyu Road, Wuhan 430079, China
Tel: (86-27)-8765-6862
E-mail: zcy7@263.net

KEYWORDS: GIS, RS, mathematical morphology (MM), knowledge discovery from databases (KDD), spatial data mining (SDM), Rough Set theory

ABSTRACT: In recent years remote sensing (RS), geographic information systems (GIS) and global positioning systems (GPS), have been three main technologies for spatial information acquisition, spatial data storage management, spatial data updating, spatial analysis and applications. However, the integration of the three technologies is an important issue and a big challenge to the geo-information community. This paper addresses three technical aspects relevant for intelligent integration of RS with GIS: (1) possible database structures for integration, (2) application possibilities of the rough set theory in spatial data mining, (3) an approach of image clustering based on mathematical morphology in remote sensing.

1. INTRODUCTION

The integration of remote sensing (RS) and geographic information system (GIS) is a topic of general interest in the field of photogrammetry, remote sensing and GIS. It mainly contributes to two kinds of applications. One is GIS database updating by remote sensing imagery, and the other is remote sensing analysis by the support of GIS data. These two aspects complement each other in the continuous updating of GIS databases.

Advances in databases and data collection technologies including remote sensing, satellite telemetry, etc., have accumulated huge amounts of data in large databases. This explosively growing data source creates the necessity of knowledge and information discovery from data, which leads to a promising emerging field, called knowledge discovery and data mining in databases (KDD). It has been long acknowledged that GIS data can be used as prior or auxiliary information to improve remote sensing image classification. In previous studies, GIS data were often used in training area selection, and post processing of classification results or acted as additional bands.

Generally, image classification in remote sensing is accomplished in a statistical and interactive manner, so that it is difficult to use the auxiliary data automatically and intelligently. On the other hand, expert system techniques were incorporated in remote sensing image classification to make use of domain knowledge and logical reasoning. But building an expert system was very difficult, because of the knowledge acquisition "bottleneck". The traditional way of knowledge acquisition is that the knowledge engineer talks with the domain expert and the represents and inputs acquired knowledge to computer in a symbolic format. This is usually a long and repeated process where missing of information often cannot be avoided. Consequently, it is difficult to put an expert system into practical use in remote sensing image classification.

In fact, large amounts of knowledge that can be used in image classification exist, but are hidden in GIS databases. The simple method of extracting knowledge from GIS database is the GIS query. For example, " What is the maximum area of all the cities?" Some other knowledge is " deep" under surface, such as spatial distribution rules, spatial association rules, shape discrimination rules, etc., that is not stored explicitly in the database but can be mined out by computation and learning. Spatial data mining and knowledge discovery in database (SDMKDD) can be defined as the discovery of interesting, implicit, and previously unknown knowledge from large databases. Data mining represents the integration of several fields, including machine learning, database systems, and information theory. SDMKDD provides a new way of knowledge acquisition for remote sensing image classification. Some researchers have already done valuable work in the field. For example, Eklund et al. extracted knowledge from TM images and geographic data in soil salinity analysis using inductive learning algorithm C4.5 (Eklund, et al., 1998).

In this paper, the relevant techniques for intelligent integration of GIS with RS are studied. The paper is organized as follows. Section 2 describes the database structure for the integration of RS with GIS. Section 3 addresses two methods for SDMKDD. One is mathematical morphology-based clustering algorithm (MMC). The other is Rough set theory. Finally, we come to the conclusions.

2. DATABASE STRUCTURE FOR THE INTEGRATION OF RS WITH GIS

The integration of RS with GIS is a hot spot in the field of geographic information science. A good database structure is very important to the integration of RS with GIS. Such a structure should accelerate the intelligent integration, and also help the knowledge discovery and exploitation. It should be able to deal with the disagreement between the resolution of remotely sensed images and the precision of spatial data in GIS.

At present, the spatial information in GIS is still not exploited adequately and effectually. Some designed systems do not fit to the integration of RS with GIS. There are two important reasons. Firstly, spatial data in GIS is object-oriented and controllable. On the contrary, the data in remotely sensed images is on the signal level and uncontrollable. Therefore, a simple approach for integration between RS and GIS would be performed by data format transform, but it would make system performance come down. Secondly, there are amounts of spatial data with complicated relationship in GIS databases, and it brings the difficulty for knowledge discovery and exploitation. For realizing data updating integrally between RS and GIS, the enabling database structure for GIS should be able to meet three requests as follows.

? Supporting integration between RS with GIS. Data structure must support the function of GIS, and also be able to process imagery for GIS data directly. The data structure should not require transformation of data formats and should be able to integrate the image processing and the functions of GIS into one system (Li, et al., 1999).

? Resolving the disagreement between the classification precision of remotely sensed images and data precision in GIS.

? Being convenient to knowledge discovery from GIS databases and exploitation. The data structure should be capable of expressing complicated objects, spatial relationships and semantic relationships.

Recently there are several database structures for the integration of RS with GIS practically, such as mixed models based on raster-vector transform, vector storage model with spatial raster data index, spatial data model based on raster data, object-oriented spatial-temporal data storage model, feature-based data storage model, etc. The commonality and the difference of several spatial data structures mentioned above are compared and showed in table 1 (Yu, et al., 2000).

Tab. 1 Comparison of Several Spatial Data Structures

Data Structure	Data Type	Index Mechanism	Image Process Function	Support integration of system
Mixed model based on raster-vector transform	Vector	Index File	Vector-raster transform interface	Realizing the integration of RS with GIS seamlessly
Vector storage model with spatial raster data index	Vector	Raster index	Vector-raster transform interface	Realizing the integration of RS with GIS seamlessly
Spatial data model based on raster data	Raster	Index grid multi-value quad-tree	Supporting image processing directly	Realizing the integration of RS with GIS entirely

To the intelligent integration, the database structure for GIS should support spatial analysis and image processing simultaneously. Now some researchers are interested in studying spatial data storage model based on raster data. It is convenient to support image processing in GIS and reduce the discrepancy between remotely sensed images and spatial data in GIS. On the other hand, vector data can be zoomed in or zoomed out infinitely. There are quantitative errors in raster data. So the database structure for the integration of RS with GIS should be able to control the precision of raster data.

For expressing complicated geographic objects, some researchers proposed an object-oriented data structure. Using this data structure, the geographic entities can be viewed as objects, and it is also possible to aggregate geometry data, attribute data, relevant processing in an object-oriented representation. The database structure allows representation of particular knowledge about a single object and can improve the knowledge expressiveness of the system. But several existing object-oriented data structures lack the ability of expressing spatial semantics. To tackle the existing problem of database structures for the intelligent integration, some assumptions about building the spatial data structure for GIS are made below.

Spatial data in GIS is often organized in an unstructured pattern. The geographic objects stored in computer exist alone and lack of the ability of expressing the relationship among objects, so it is difficult to discover knowledge from GIS databases. It requires a new semantic-oriented data structure in GIS. The semantic-oriented model should have several characteristics as follows. ? the object nodes, the knowledge nodes and the index nodes are stored integrally. ? simple spatial objects are aggregated into complex objects. The data are stored hierarchically according to the different scales and precisions. ? a strong index mechanism should be constructed for depicting the position relationship of spatial objects and expressing the spatial semantic information of geographic objects. We can adopt index tree mechanism to depict the relationship of aggregation and use 2-dimensional table to express the topological structure of the objects. These data features improve the ability of system in knowledge expression, intelligent inference and association. The database structure benefits from using the information of GIS for the interpretation of remotely sensed images.

3. KNOWLEDGE DISCOVERY FROM GIS SPATIAL DATABASES

As discussed above, a good database structure for the integration of RS with GIS is convenient to discover knowledge from spatial databases. In this section, we study several methods of knowledge discovery from spatial database and spatial data mining (KDDSDM). At first, the structure framework of a spatial knowledge discovery system is shown as Fig.1 (Di, et al., 1998).

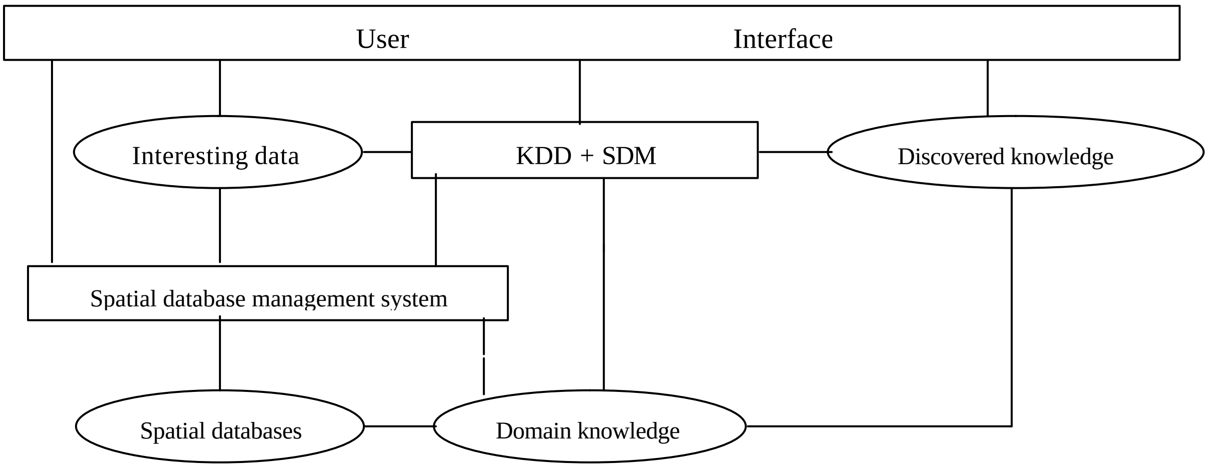

Fig.1 the architectural framework of SDM system

In general, geographic data consist of spatial objects and non-spatial descriptions of objects. Non-spatial descriptions of spatial objects can be stored in a traditional relational database. Spatial data can be described using two different properties, geometric and topological. For example, geometric properties can be spatial location, area, perimeter, etc., whereas topological properties can be adjacency, inclusion and others. Thus, the methods for discovering knowledge can be focused on the non-spatial and/or spatial properties of spatial objects. The algorithms studied for a long time for spatial data mining include inductive learning, generalization-based methods for mining spatial characteristics and discrimination rules, aggregate proximity technique for finding characteristics of spatial clusters, etc. In the following section we will categorize and describe two of these algorithms.

3.1 Attribute Analysis and Knowledge Discovery Based on Rough Set Theory in GIS

Data precision and uncertainty in GIS are important issues in geo-information science. Much attention in previous studies had been paid to the data precision and uncertainty of geometry information and quantitative data. However, no much study is done on attribute, special quantitative data. In the existing GIS, attribute data are only queried in simple ways, and it is difficult to analyze and discover knowledge from GIS databases. Rough set theory provides an intelligent decision analysis tool that can be used in classification analysis and knowledge discovery of inaccurate or uncertain data. The theory has performed successfully in decision analysis, machine learning, information index and likelihood reasoning, etc. It becomes an important tool for data mining and knowledge discovery (Pawlak, 1982). According to the Rough set theory, the attribute data in GIS organized as relational tables can be viewed as a kind of attribute-value system. In GIS, attribute data is stored in relational databases and depicted in relational tables. The rows of tables describe the objects and the volumes of tables describe attributes. Obviously, the theory and methods of the rough set can be used in analysis and processing of attribute data. Furthermore, we can extract the methods based on rough set theory that are suitable to attribute analysis and knowledge discovery in GIS, including consistency of attribute table, attribute significance, attribute dependence, reduction of attribute table, minimum decision algorithm generation and classification algorithm. In these algorithms, all of the attribute data are viewed as qualitative data. If there is a need for process the qualitative and the quantitative data synthetically, quantitative data has to be converted to qualitative data. A suitable transform method segments automatically the quantitative data according to the statistical distribution rules. Each segment will be viewed as qualitative data. These methods based on rough set theory are valuable in theory and practice and provide new solutions for GIS attribute analysis and knowledge discovery.

3.2 A Mathematical Morphology-based Clustering Algorithm

Cluster analysis is a branch of statistics that has been studied extensively for many years. The main advantage of using this technique is that interesting structures or clusters can be found directly from the data without using any background knowledge. The research results of clustering techniques can be naturally exploited in spatial data mining. In this section, we introduce a clustering algorithm based on mathematical morphology (MMC).

Based on the set theory, mathematical morphology provides an approach for processing of digital images. Using appropriate sets known as structural elements, mathematical morphology operations can simplify image data while maintaining their shape characteristics and eliminating irrelevancies. Serra (1982) introduced the basis and theory of mathematical morphology. Haralick, et al. (1987) discussed the basic mathematical morphology operations and their relations in N-dimensional properties of the basic binary and multi-level morphological operations with both 1D and 2D structural elements. A mathematical operation is presented as follows.

Binary Mathematical Operations (Saradjian, 2000)

Dilation, Erosion, Opening, and Closing. Consider a discrete binary image set X in an N-D distance grid Z^N. Let T? Z^N denote a structural element. $B = \{-t \mid t \in T\}$ denotes the symmetric set of T with respect to the origin, and *f* denotes the empty set.

The translation of X by a point *z?* Z^N is denoted by X_z and defined by $X_z = \{x + z \mid x \in X\}$. Then the two basic binary mathematical morphological operations of X by Y are defined as follows:

Dilation $$X \oplus T = \{z \mid B_z \cap X \neq f\} = \bigcup_{t \in T} X_t$$

Erosion $$X \Theta T = \{z \mid T_z \subseteq X\} = \bigcap_{t \in T} X_{-t}$$

Dilation is used to fill small holes and fill narrow gaps in objects or expand image objects, whereas erosion shrinks the image objects. If we want to find the contours of objects in an image very quickly this can, for instance, be achieved by the subtraction from the original picture of its eroded version. Opening is used to eliminate specific image details smaller than the structural element while closing connects objects that are close to each other, fill up small holes, and smoothes the object outlines by filling up narrow gaps. Unlike dilation and erosion, opening and closing are invariant to translation of the structural element. It means, if an image is eroded and then dilated the original image is not recoverable. The binary morphological operations of dilation, erosion, opening and closing can be extended to gray scale imagery (Sternberg, 1982; Haralick, et al., 1987). For such images, the minimum and maximum values are found within neighborhoods represented by the structural element. Let F and T be the domain of the gray scale image *f* and the gray scale structural element *t*, respectively. The gray scale dilation and erosion can be computed by

Dilation $$(f \oplus t)(x, y) = \max_{(x-m, y-n) \in F, (m,n) \in T} \{f(x-m, y-n) + t(m,n)\}$$

Erosion $$(f \Theta t)(x, y) = \min_{(x+m, y+n) \in F, (m,n) \in T} \{f(x+m, y+n) - t(m,n)\}$$

Now we will give the mathematical morphology-based clustering algorithm as follows (Li, et al., 2000).

Input: interesting data X (0 denotes background, 1 denotes objects)

Output: clustering result Y (uniform clustering with the same properties)

Procedure: initialized value i =1

? building circular structural element B, and its radius is i

? closing: $Y_i = X \bullet B_i$

? counting the clustering number of connected region $n_{i,}$
if n_i>1,then i = i +1 ,return to (1), otherwise continue (4);

? count the optimum clustering number n_i according to $n_{i,}$ then get the structural element's radius k

? $Y = X \bullet B_k$;

? coloring the connected regions, different clustering regions with different colors.

In this algorithm, it is important to define the optimum clustering number. The algorithm clusters by the distance, that is the objects of the uniform clustering have little distances among them. The algorithm is suitable to the 2D environment. We can extend to the 3D environment. MMC algorithm processes mainly digital image, however, to the large amounts of vector data, we can transform the vector data to raster data, but lots of work have to be done for data transform. We are still investigating clustering algorithms for vector data based on mathematical morphology.

4. CONCLUSION

The integration technique of RS with GIS presented in this paper can be used for quick detection of land-cover changes. KDD/SDM techniques allow us to derive large amounts of knowledge hidden in GIS databases previously, and the knowledge will be helpful in improving the quality of image classification. Experiments on land use classification using this method show the positive increase in the overall accuracy through using spatial data mining. However, the intelligent integration of RS with GIS remains a difficult problem. We will continue our investigation into better solutions.

REFERENCES

1. Han,J., 1996. Data Mining Techniques. ACM-SIGMCD'96 Conf. Tutorial.
2. Serra,J.,1982.Image analysis and mathematical morphology. Academic press, London, 610p.
3. Li Deren, Guan Zequn, 1999. A Study GIS Data Directly into Image Processing. Journal of WTUSM Vol.24 NO.1, pp.1-5.
4. Saradjian,M.R., Mini.J.A.,2000 Imagemap simplification using mathematical morphology. International Archives of Photogrammetry and Remote sensing, Vol. XXXIII. Part B3. Amsterdam 2000,pp.36-42.
5. Eklund,P.W., Kirkby,S.D., Salim,A., 1998. Data mining and soil salinity analysis, Int. J. Geographical Information Sciences Vol.12 NO.3, pp.247-268.
6. Li Deren, Di Kaichang, Li Deyi, 2000. Land Use Classification of Remote sensing Image With GIS Data Based on Spatial Data Mining Techniques. International Archives of Photogrammetry and Remote sensing, Vol. XXXIII. Part B3. Amsterdam 2000, pp.238-245.
7. Di Kaichang, Li Deren, Li Deyi, 1999. Rough set Theory and Its Application in Attribute Analysis and Knowledge Discovery in GIS, Journal of WTUSM Vol.24 NO.1, pp.6-10.
8. Di Kaichang, Li Deren, Li Deyi, 1998. A Mathematical Morphology Based Algorithm for Discovery Clusters in Spatial Databases, Journal of image and graphics Vol.3 NO.3, pp.173-178.
9. Yu Nenghai, Wang Xiaogang, Liu Zhengkai, 2000. Database Structure for the Integration of RS with GIS, Journal of WTUSM Vol.25 NO.4, pp.305-311.