Building Semantic Ontology Databases Based on Remote Sensing Images

Zhenfeng Shao 1, Jun Liu 2 and Xianqiang Zhu 1

1 State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

2 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

Abstract: This paper puts forward a novel method of automatically generating elements which is necessary for building sharing ontology database and studies the uncertainty problems. A concept called sharing ontology is proposed which is defined as a bridge of all kinds of spatial information systems involving domain spatial data acquiring, updating, transferring, storing, processing, analyzing, information extracting, knowledge discovering and Geo-services. By establishing the spatial information sharing ontology database, we can provide for the users an integrated spatial information service and application platform based on the semanteme, and then link all spatial information service platforms on internet at the semantic level, so as to provide a feasible way of semantic-based spatial information sharing and interoperability. Through data mining based on the remote sensing images and GIS database, we can build the sharing ontology database automatically, which achieves semantic sharing of isomerous spatial information. Some innovative ideas of the research work are as follows:(1) Putting the uncertainty problem about features extracted from remote sensing images up to a height of information theory and exploring the consolidate mathematics expression between information quantity and uncertainty about features mined from remote sensing images.(2) Building the sharing ontology database by combining remote sensing and GIS spatial data, and establishing the feasibility in practice for the construction of the sharing ontology database.(3) Establishing the stability foundation for the precise semantic sharing between isomerous spatial information systems.

Keywords: sharing ontology database, domain ontology, semantic-based spatial information sharing and interoperability, uncertainty

1. Introduction

With rapid development of the global information-sharing process, the establishment of a cross-sector network of geographic information systems, allowing users transparently use all the spatial information resources on the network has become an urgent problem. The emergence of Grid [1] technology provides the possibility of eliminating the obstacles of geographic information systems, but the question is that it is unable to achieve spatial information sharing services only by connecting the spatial information systems seamlessly. As people in different areas may have different cognition of the real world and may have description of the same geographical phenomenon or object from different angles, then semantic heterogeneity comes out, resulting in “semantic gap” or “islands of information”, which even occurs in spatial data collection stage [2]. How to eliminate the semantic heterogeneity is a problem that current spatial information sharing services must face.

Just as academician Deren LI once pointed out that there were four challenges in the sharing of different spatial information systems: inconsistencies of spatial data format, temporal datum, geodetic datum and semantics[3], while semantic conflict was one of the most important problems which cause the semantic challenge. To address this problem, this paper describes a framework to eliminate the semantic heterogeneity and establishes sharing ontology databases which can describe the intrinsic characteristics of spatial objects by mining the elements from remote sensing images.
1.1. Remote sensing image mining
Up to now relative researches on remote sensing image mining can be summarized as follows in general:

Simon Fraser at University of Canada has used the multi-dimensional analysis technology to create image data cube of mining the image’s association rules, which linked the information of image size, color and description of images [5]. A research team of California University at Santa Barbara in American, led by professor Manjunath have studied the association rules between the image objects based on special incident cube. Professor Li Qi, Du Ming-yi, Li Fan, Xu Ming-jie have researched the mining methods of spatial data warehouse [6] [7] [8] [9]. Prof. Deren LI of Wuhan University has developed the data mining methods based on the formula concept analysis [4] [10].

1.2. Geographic ontology
So far geographic ontology has been paid more and more attention by researches involving semantic sharing. The OGC has divided the integration and interoperability operation of geographic information system into three levels: the data layer, the syntax layer and the semantic layer [11]. The sharing of the first two levels can be realized by unifying data structure standards, modularizing service functions etc. But the integration and interoperability on the semantic layer is relatively more complex, and there are many studies by the means of ontology. Sheth has divided the heterogeneity of information system into four levels: system, grammar, semantic and structure[12]. Wiederhold has proposed the semantic analysis based on conceptualized tools for sharing and reusing the heterogeneous knowledge [13]. Henrique C.M.Andrade and Joel Saltz have used the ontology concept to express the relationship of the same object in different fields [14]. Zhao Zhongtian has adopted the RDF/RDFS ontology model for building the domain ontology [15]. F. Fonseca has introduced the basic principles of ontology-driven geographic information systems[16].

In this paper, the authors put forward a novel method which describes the construction of GIS sharing ontology databases by data mining from remote sensing images.

2. Framework for building the sharing semantic ontology database
A framework for building the sharing semantic ontology database is designed, as illustrated in figure 1, through spatial data mining from remote sensing images, we can automatically acquire the necessary characteristics for sharing ontology databases, and generate semantic sharing ontology databases based on which to construct domain ontology system, and provide a feasible way to the semantic-based information sharing and interoperability.

In the whole flow showed in figure 2, there are two aspects of research work:

- Mining characteristic information for sharing ontology database from remote sensing images
The mining process from remote sensing can be understood as a process of forming concepts from remote sensing images, a process of extracting concepts of different levels from images and GIS data, analyzing the relationships between concepts by specializing and generalizing concepts, thereby the potential and implicated knowledge can be mined from remote sensing images.

- Exploring of mechanism on constructing GIS ontology databases

The GIS ontology databases are composed of GIS application ontology database, GIS domain ontology database, GIS sharing ontology database and so on. A GIS sharing ontology can be defined as an ontology which is shared by all GIS systems in different fields. The elements of a sharing ontology can be acquired by spatial data mining. The specific process of building flow is composed of three parts: data mining, constructing sharing ontology database and GIS interoperability.

3. Concept Assignment and Methodology

The procedure of mining data from remote sensing images can be considered as a process of concept formation. The most abundant amount of information of remote sensing images is the basis of information mining. The purpose of remote sensing image data mining is to analyze and mine potential knowledge from massive data and other relevant data, to extract all level concepts and relationships between these concepts. In this paper, we extract characteristic information of geographic entities for further ontology database construction from the view of data mining.

3.1. Geographic entity concept

In psychology, a concept refers to cognition pattern which reflects the common characteristics and nature attributes of a certain object. It is the basic unit of higher cognitive activities. Each concept of the geographical entity includes the connotation and extension. Connotation refers to the essence characteristics of objects, and extension represents the scope of the concept.

In this paper, there are two ways to generate concepts:

The first one is artificial selection. Concepts can be chosen through the support of experts in the field, as well as the levels and relationships between concepts, so as to achieve the purpose of forming concept.

The second one is semi-automatic generation. As for the existing GIS data, such as land-use maps and other classification maps, different categories themselves are the basic concepts within the given field, while concepts of higher levels can be generated by mining the remote sensing images automatically. As the concepts all have corresponding semantic information, so supported by expertise, we can establish the levels and relationships between these concepts.

3.2. Ontology and ontology graphic

O’Leary defines ontology as “an explicit specification of a conceptualization” [18]. This knowledge-based specification typically describes a taxonomy of the relationships that defines the knowledge. Ontology structure is defined by ternary group (S, O, R), in which S represents subject, O represents object, and the relationship between subject and object is defined as R. S and O are collections of concepts. Ontology can be described by graphic which contains nodes and edges, in which node refers to S and O, and edge refers to R. Ontology graphic is described as binary Group G = (N, E), where N are collection of nodes, E are edges.

The relationship that needed to identify between ontology concepts includes synonyms, contains and no relationship. By empowering the relationship between the ontology, we can build ontology graphic to describe this relationship. After we acquire the relationship, we can describe the relationship between ontologies or between ontology systems, so as to build more complicated ontology relationship.

3.3. Generation of ontology databases

According to the concept of ontology, we define the ontology databases as a formula and explicit specification of a conceptualization shared by different application fields. The elements of ontology database can be mined from remote sensing images or GIS system by means of spatial data mining technology and ontology theory.

According to the characteristics mined from remote sensing images, supporting by expertise and GIS knowledge, we can get concept sets and attribute sets that can describe geographic entities in all fields:
{latitude and longitude, texture, features spectrum, area, points or shape or linear features, slope, aspect, elevation, soil type, climate…}. The sharing ontology databases are the crossover sets that are generated from these concept sets and attribute sets and used to describe the concepts of some geographic entity in different fields. As long as these crossover sets are not empty, that means, the subsets or full sets of GIS sharing ontology databases are found.

3.4. Domain ontology and application ontology
We can define the domain ontology according to the definition of sharing ontology database. The domain ontology is a sharing knowledge base that is common approved by all application in a specific area, is a concept with level structure, and every concept is described by a set of attributes in corresponding area. The application ontology is the combination of domain ontology and specific applications, and is bound in the domain ontology, which is a common concept framework. When different domain ontologies are used for different application, it is necessary to consider the nature of application, so we call it the application ontology. The application is the projection of domain ontology in a specific field and combined with the needs of users. The sharing ontology expresses the basic characteristics in different fields, and provides the basis for integration and interoperability.

4. Implementation and experimental results
4.1. Images preparation
In order to mine concepts which can be used in different fields, based on the size of the object, we select remote sensing images with appropriate resolution and corresponding GIS data, such as all kinds of maps. With these images, we can prepare for mining the nature characteristics of different objects, and prepare for building the ontology graphic.

4.2. Mining association rules of texture characteristics
The concrete flow is: normalizing, storing into database, generating Hasse graphic, building rules.
- Normalize the images into gray value 0,1,2.
- Form a database record with every pixel on the normalized images and its corresponding 8 neighborhoods pixels that are also be normalized, and then import this record into the texture feature database.
- According to the texture features and a fast algorithm based on tree index integrated with concept lattice and Hasse graph and so on, we can establish the concept lattice corresponding to the texture features, and generate the Hasse graph.
- According to the Hasse graph and the generated concept lattice, we can build the associated rules that reflect the texture features.

Fig. 3: Relationship between general ontology, domain ontology and application ontology
<table>
<thead>
<tr>
<th>Rule number</th>
<th>rule</th>
<th>Support</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>{(1,0,0)} =&gt; {(0,0,0)}</td>
<td>0.409832</td>
<td>0.801980</td>
</tr>
<tr>
<td>(2)</td>
<td>{(1,1,0)} =&gt; {(1,0,0)}</td>
<td>0.320984</td>
<td>0.801619</td>
</tr>
<tr>
<td>(3)</td>
<td>{(0,1,2)} =&gt; {(0,0,2)}</td>
<td>0.296737</td>
<td>0.923469</td>
</tr>
<tr>
<td>(4)</td>
<td>{(0,-1,2)} =&gt; {(0,0,2)}</td>
<td>0.428436</td>
<td>0.932292</td>
</tr>
<tr>
<td>(5)</td>
<td>{(1,-1,2)} =&gt; {(1,0,2)}</td>
<td>0.280903</td>
<td>0.811448</td>
</tr>
<tr>
<td>(6)</td>
<td>{(-1,-1,2)} =&gt; {(-1,0,2)}</td>
<td>0.300826</td>
<td>0.927711</td>
</tr>
<tr>
<td>(7)</td>
<td>{(1,-1,0),(1,1,0)} =&gt; {(1,0,0)}</td>
<td>0.259364</td>
<td>0.938650</td>
</tr>
<tr>
<td>(8)</td>
<td>{(0,-1,0),(0,1,0)} =&gt; {(0,0,0)}</td>
<td>0.370809</td>
<td>0.855491</td>
</tr>
<tr>
<td>(9)</td>
<td>{(1,-1,0),(-1,1,0)} =&gt; {-1,0,0}</td>
<td>0.237458</td>
<td>0.812925</td>
</tr>
<tr>
<td>(10)</td>
<td>{(1,1,2),(1,1,2)} =&gt; {(1,0,2)}</td>
<td>0.364783</td>
<td>0.926829</td>
</tr>
<tr>
<td>(11)</td>
<td>{(1,1,2),(-1,0,2)} =&gt; {(1,0,2)}</td>
<td>0.348655</td>
<td>0.932961</td>
</tr>
<tr>
<td>(12)</td>
<td>{(-1,-1,2),(-1,1,2)} =&gt; {(0,-1,2)}</td>
<td>0.367746</td>
<td>0.803859</td>
</tr>
<tr>
<td>(13)</td>
<td>{(0,-1,2),(0,1,2)} =&gt; {(0,0,2)}</td>
<td>0.348779</td>
<td>0.805243</td>
</tr>
</tbody>
</table>

Fig. 4: association rules

• Explain association rules of texture characteristics.

<table>
<thead>
<tr>
<th>rule 1</th>
<th>rule 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="rule 1" /></td>
<td><img src="image2" alt="rule 2" /></td>
</tr>
</tbody>
</table>

rule 1: S = 0.409832, C=0.801980
rule 12: S = 0.348655, C=0.932961

Fig. 5: explanation of rules

After mining these rules of texture characteristics, we need to explain the specific meaning of these rules according to the knowledge in different domains. For example, the explain of rule 12 is: in a 3*3 neighborhood image, after normalized, if the value of lower-left corner pixel is 2, and the value of upper-left corner pixel is 2, then we can reason that the value of its very left pixel is 2, and the support for this rule is 0.348655, the confidence is 0.932961.

4.3. Construction of sharing ontology databases

We define the hierarchy structure of sharing ontology databases as Fig 2. In the structure of sharing ontology databases, concepts and their attributes are divided into several categories in accordance with different functions, and managed by different ontologies which are named functional ontologies. Based on the difference in connotation and extension, the functional ontology in the sharing ontology database is divided as follows:

![hierarchy structure of ontology](image)

• time Ontology: include time concept and relationship, such as time for image acquisition, mapping, time sequence, time interval, etc.
• Location Ontology: involves the latitude and longitude of work area, the projection and so on.
• concept ontology: be composed of concepts from images mining.
• Topology Ontology: consists of adjacent, connection, and intersection relations.
• Resource Ontology: refers to concepts and attributes of resources, such as data types (TM, QuickBird, Spot, DEM, DOM, DLG, DRG…, scale (10000, 100000, 1000000…), and so on.
• Feature ontology: covers all features of geographical entities obtained by data mining, such as texture association rules.

In particular, the establishment of the ontology contains the following three parts of the work:

(1) Acquisition of ontology concept automatically

Mainly, we need to acquire the connotation of concept ontology above all. The bottom level concepts of some field are contained in the GIS data. Based on statistical data including remote sensing images, we can get a relation matrix between concepts in different domain, and build the relationships between concepts, which generate the elements of the sharing ontology database. Other concepts and attributes of functional ontology can be acquired by other means such as data mining, for example, the acquisition time of image can be the element of time ontology, image type and scale can be the elements of resource ontology, texture and shape characteristics can be the elements of feature ontology.

(2) Building sharing ontology databases based on ontology graphic

Firstly an ontology graphic should be designed. Secondly we can acquire the concepts and attributes of all functional ontology. Thirdly, supporting by some expertise, we can build ontology graphic. As the ontology concepts acquired are at the lowest level of the field, it is necessary to confirm the relationship between concepts or relationship between concepts.

Finally we can construct sharing ontology databases. Sharing ontology databases can be expressed as dual group (Os,G), in which Os is the set of ontology, and G is the ontology graphic. All the sub-ontology and all the elements for sub-ontology in the sharing ontology databases are included in Os, and all the relationships between ontology as well as reasoning relations are included in ontology graphic G. In a word, the sharing ontology databases stores ontologies and ontology graphics.

(3) Generating the domain ontology based on the sharing ontology

Supporting by expertise and GIS knowledge, for a specific field, we can form domain ontology system by extracting the elements of sub-ontology that needed from sharing ontology databases.

Repeating the above operations, we can extract corresponding relationships between ontology in domain ontology and sharing ontology databases, and then build domain ontology system based on sharing ontology databases. Mean while, as different domain ontology systems are built by extracting ontology from sharing ontology databases, through a specific ontology in sharing ontology databases, we can easily build the semantic mapping relations between concepts from different domain ontology systems, so as to provide the basis for interoperability to different domain ontology systems.

4.4. Accuracy analysis

In this paper, the uncertainty includes the following aspects:

• uncertainty from data

In the remote sensing imaging process, there exist two phenomenon that the same object may have different spectrum and the different objects may have the same spectrum, which make the spectrum and characteristics of objects show uncertainty. In the process of data acquisition, it is difficult to control some error, such as uniformity of atmospheric conditions, even if by spectral and geometric correction, there will still be residual. In the process of image matching by geometric correction, it is unable to eliminate the error caused by ground control point or image control points, which affect the next extracting texture characteristics. The classification maps used in this paper, for the technical means and human factors, may have some uncertainty, such as the unclear, variance class definition or no unified quantitative standards, or the classification error caused classifying itself. Although by precision controlling, they can basically meet the use of sectors, the gap between them and the actual is existing.

• uncertainty in extracting rules

In this paper, association rules in the formula concept analysis are used to extract texture nature features of objects. According to the support and confidence, there are five indicators used to control the uncertainty:
(1) Support degree

The support degree of association rule \( A \Rightarrow B \) means the percentage that transaction \( A \cup B \) (that means containing A and B) contained by transaction D, or, it is the probability \( P(A \cup B) \), that is, support(\( A \Rightarrow B \))=\( P(A \cup B) \). It shows the probability that union set C of item A and item B contained in transaction D.

(2) Confidence grade

The confidence of association grade rule \( A \Rightarrow B \) means the percentage that transaction A and transaction B contained by transaction D, or, it is the conditional probability \( P(B/A) \), that is, confidence(\( A \Rightarrow B \))=\( P(B/A) \). It shows the probability that item A and item B contained in transaction D at the same time.

Support degree is the measurement of importance of the association rules, while confidence is accuracy. The Support illustrates the representation of this rule in all the transaction, the greater the confidence is, the more important this rule is. Some association rules have great confidence but with low support, that means, the chance of these rules is very small, and their practical value is little, so, they are also not important.

(3) Expected confidence

It means the percentage that transaction B contained by transaction D, that is, Expected Confidence(B) = \( P(B) \). It shows the probability that item B contained in all transactions without any other conditions.

(4)Lift

It is the ration of confidence and expected confidence, that is, Lift(B/A) = confidence(\( A \Rightarrow B \))/ Expected Confidence(B). It shows the influence of the emergence of item A to the emergence of item B. The greater the lift is, the bigger influence item A to item B. Under normal circumstances, the lift of useful association rules is bigger than 1, that is to say, its confidence is bigger than expected confidence, or, the emergence of item A promotes the emergence of item B, also, they are relevant to a certain degree. If the lift is less than 1, the rules will be of no significance[17].

(5) Interesting rules

Support degree and expected confidence are two important indicators that reflect the extent of interest of rules. We define separately a minimum support threshold as min_sup and a minimum confidence threshold as min_conf, if the support and confidence of rule are both bigger then min_sum and min_conf, it is an interesting rule, or a strong rule. During the actual data mining analysis, we must choose an appropriate minimum support threshold and minimum confidence threshold. If they are too small, there will be a large quantity of useless rules, not only affecting the efficiency and wasting system resources, but also submerging the main target. On the other hand, if they are too big, there will no rules, or not enough rules, filtering out the meaningful rules. So we should set an appropriate threshold according to the general needs.

5. Conclusions

This paper presents an innovative method and its implementation to realize GIS semantic interoperability by data mining from remote sensing, and constructs sharing ontology databases combined with GIS data. Through matching GIS data with remote sensing images, we can obtain the lowest level concepts that are necessary for the sharing ontology databases. The texture characteristics are mined from images to generate association rules in formula concept analysis methods. Using ontology theory, combing with ontology graphic and intervene, we have constructed the sharing ontology databases successfully; the authors also analyze the uncertainty in the whole flow. In the future, we expect to improve the efficiency of data mining and strengthen the capacity of ontology interoperability.

6. Acknowledgements

This paper is supported by the National Key Basic Research and Development Program of China (No.2004CB318206) and the Basic Research of Survey and mapping bureau Project (No.1469990711111).
7. References


