CONCEPTS, PRINCIPLES AND APPLICATIONS OF SPATIAL DATA MINING AND KNOWLEDGE DISCOVERY

Deren LI[a] Shuliang WANG[b, c]

[a] National Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[b] International School of Software, Wuhan University, Wuhan 430072, China
[c] School of Economics and Management, Tsinghua University, Beijing 100084, China2
E-mail: drli@whu.edu.cn Telephone: +86-27-87865402, Fax: +86-27-87865402

ISPRS, XXXVI-2/W25

KEY WORDS: spatial data mining, principles, cloud model, date fields, applications

ABSTRACT:
A growing attention has been paid to spatial data mining and knowledge discovery (SDMKD). This paper presents the principles of SDMKD, proposes three new techniques, and gives their applicability and examples. First, the motivation of SDMKD is briefed. Second, the intension and extension of SDMKD concept are presented. Third, three new techniques are proposed in this section, i.e. SDMKD-based image classification that integrates spatial inductive learning from GIS database and Bayesian classification, cloud model that integrates randomness and fuzziness, data field that radiate the energy of observed data to the universe discourse. Fourth, applicability and examples are studied on three cases. The first is remote sensing classification, the second is landslide-monitoring data mining, and the third is uncertain reasoning. Finally, the whole paper is concluded and discussed.

1. MOTIVATIONS

The technical progress in computerized data acquisition and storage results in the growth of vast databases. With the continuous increase and accumulation, the huge amounts of the computerized data have far exceeded human ability to completely interpret and use. These phenomena may be more serious in geo-spatial science. In order to understand and make full use of these data repositories, a few techniques have been tried, e.g. expert system, database management system, spatial data analysis, machine learning, and artificial intelligence. In 1989, knowledge discovery in databases was further proposed. In 1995, data mining also appears. As both data mining and knowledge discovery in databases virtually point to the same techniques, people would like to call them together, i.e. data mining and knowledge discovery (DMKD). As 80% data are geo-referenced, the necessity forces people to consider spatial characteristics in DMKD and to further develop a branch in geo-spatial science, i.e. SDMKD (Li, Cheng, 1994; Ester et al., 2000).

Spatial data are more complex, more changeable and bigger that common affair datasets. Spatial dimension means each item of data has a spatial reference (Haining, 2003) where each entity occurs on the continuous surface, or where the spatial-referenced relationship exists between two neighbor entities. Spatial data includes not only positional data and attribute data, but also spatial relationships among spatial entities. Moreover, spatial data structure is more complex than the tables in ordinary relational database. Besides tabular data, there are vector and raster graphic data in spatial database. And the features of graphic data are not explicitly stored in the database. At the same time, contemporary GIS have only basic analysis functionalities, the results of which are explicit. And it is under the assumption of dependency and on the basis of the sampled data that geostatistics estimates at unsampled locations or make a map of the attribute. Because the discovered spatial knowledge can support and improve spatial data-referenced decision-making, a growing attention has been paid to the study, development and application of SDMKD (Han, Kamber, 2001; Miller, Han, 2001; Li et al, 2001; 2002).

This paper proposes the concepts, techniques and applications of SDMKD. In the following, section 2 describes the concept of SDMKD, paying more attention to the knowledge to be discovered, discovery mechanism, and mining granularity. Section 3 presents the techniques to be used in SDMKD. After the existing techniques are overviewed, three new techniques are further proposed. Section 4 gives the applicability and examples of SDMKD. Finally we come to the conclusions in section 5.

2. CONCEPTS

Spatial data mining and knowledge discovery (SDMKD) is the efficient extraction of hidden, implicit, interesting, previously unknown, potentially useful, ultimately understandable, spatial or non-spatial knowledge (rules, regularities, patterns, constraints) from incomplete, noisy, fuzzy, random and practical data in large spatial databases. It is a confluence of databases technology, artificial intelligence, machine learning, probabilistic statistics, visualization, information science, pattern recognition and other disciplines. Understood from different viewpoints (Table 1), SDMKD shows many new interdisciplinary characteristics.

2.1 Mechanism

SDMKD is a process of discovering a form of rules plus exceptions at hierarchal view-angles with various thresholds, e.g. drilling, dicing and pivoting on multidimensional databases, spatial data warehousing, generalizing, characterizing and classifying entities, summarizing and contrasting data characteristics, describing rules, predicting future trends and so on (Han, Kamber, 2001). It is also a supportable process of spatial decision-making. There are two mining granularities, spatial object granularity and pixel granularity (Li, Wang, Li, 2005).

It may be briefly partitioned three big steps, data preparation (positioning mining objective, collecting background knowledge, cleaning spatial data), data mining (decreasing data dimensions, selecting mining techniques, discovering knowledge), and knowledge application (interpretation, evaluation and application of the discovered knowledge).

In order to discovery the confidential knowledge, it is common to use more than one technique to mine the data sets at the same time. And it is also suitable to select the mining techniques on the basis of the given mining task and the knowledge to be discovered.

2.2 Knowledge to be discovered

The knowledge is more generalized, condensed and understandable than data. The common knowledge is summarized and generalized from huge amounts of spatial data sets. The amount of spatial data is huge, while the volume of spatial rules is very small. The more generalized the knowledge, the bigger the contrast. There are many kinds of knowledge that can be mined from large spatial data sets (Miller, Han, 2001; Wang, 2002) (See Table 2). In table 2, these kinds of rules are not isolated, and they often benefit from each other. And various forms, such as linguistic concept, characteristic table, predication logic, semantic network, object orientation, and visualization, can represent the discovered knowledge. Very complex nonlinear knowledge may be depicted with a group of rules.

Table.1 Spatial data mining and knowledge discovery in various viewpoints

Viewpoints	Spatial data mining and knowledge discovery
Discipline	A interdisciplinary subject, and its theories and techniques are linked with database, computer, statistics, cognitive science, artificial intelligence, mathematics, machine learning, network, data mining, knowledge discovery database, data analysis, pattern recognition, etc.
Analysis	Discover unknown and useful rules from huge amount of data via a sets of interactive, repetitive, associative, and data-oriented manipulations
Logic	An advanced technique of deductive spatial reasoning. It is discovery, not proof. The knowledge is conditional generic on the mined data.
Cognitive science	An inductive process that is from concrete data to abstract patterns, from special phenomena to general rules.
Objective data	Data forms: vector, raster, and vector-raster Data structures: hierarchy, relation, net, and object-oriented Spatial and non-spatial data contents: positions, attributes, texts, images, graphics, databases, file system, log files, voices, web and multimedia
Systematic information	Original data in database, cleaned data in data warehouse, senior commands from users, background knowledge from applicable fields.
Methodology	Match the multidisciplinary philosophy of human thinking that suitably deals with the complexity, uncertainty, and variety when briefing data and representing rules.
Application	All spatial data-referenced fields and decision-making process, e.g. GIS, remote sensing, GPS (global positioning system), transportation, police, medicine, transportation, navigation, robot, etc.

Table.2 Main spatial knowledge to be discovered

Knowledge	Description	Examples
Association rule	A logic association among different sets of spatial entities that associate one or more spatial objects with other spatial objects. Study the frequency of items occurring together in transactional databases.	Rain (x, pour) => Landslide (x, happen), interestingness is 98%, support is 76%, and confidence is 51%.
Characteristics rule	A common character of a kind of spatial entity, or several kinds of spatial entities. A kind of tested knowledge for summarizing similar features of objects in a target class.	Characterize similar ground objects in a large set of remote sensing images
Discriminate rule	A special rule that tells one spatial entity from other spatial entity. Different spatial characteristics rules. Comparison of general features of objects between a target class and a contrasting class.	Compare land price in urban boundary and land price in urban center
Clustering rule	A segmentation rule that groups a set of objects together by virtue of their similarity or proximity to each other in the unknown contexts what groups and how many groups will be clustered. Organize data in unsupervised clusters based on attribute values.	Group crime locations to find distribution patterns.
Classification rule	A rule that defines whether a spatial entity belongs to a particular class or set in the known contexts what classes and how many classes will be classified. Organize data in given/supervised classes based on attribute values.	Classify remote sensed images based on spectrum and GIS data.
Serial rules	A spatiotemporal constrained rule that relates spatial entities in time continuously, or the function dependency among the parameters. Analyze the trends, deviations, regression, sequential pattern, and similar sequences.	In summer, landslide disaster often happens. Land price is the function of influential factors and time.
Predictive rule	An inner trend that forecasts future values of some spatial variables when the temporal or spatial center is moved to another one. Predict some unknown or missing attribute values based on other seasonal or periodical information.	Forecast the movement trend of landslide based on available monitoring data.

Exceptions	Outliers that are isolated from common rules or derivate from other data observations very much	A monitoring point with much bigger movement.

Knowledge is rule plus exception. A spatial rule is a pattern showing the intersection of two or more spatial objects or space-depending attributes according to a particular spacing or set of arrangements (Ester, 2000). Besides the rules, during the discovering process of description or prediction, there may be some exceptions (also named outliers) that derivate so much from other data observations (Shekhar, Lu, Zhang, 2003). They identify and explain exceptions (surprises). For example, spatial trend predictive modeling first discovered the centers that are local maximal of some non-spatial attribute, then determined the (theoretical) trend of some non-spatial attribute when moving away from the centers. Finally few deviations are found that some data were away from the theoretical trend. These deviations may arouse suspicious that they are noise, or generated by a different mechanism. How to explain these outliers? Traditionally, outliers' detection has been studies via statistics, and a number of discordancy tests have been developed. Most of them treat outliers as "noise" and they try to eliminate the effects of outliers by removing outliers or develop some outlier-resistant methods (Hawkins, 1980). In fact, these outliers prove the rules. In the context of data mining, they are meaningful input signals rather than noise. In some cases, outliers represent unique characteristics of the objects that are important to an organization. Therefore, a piece of generic knowledge is virtually in the form of rule plus exception.

3. TECHNIQUES FOR SDMKD

Because SDMKD is an interdisciplinary subject, there are various techniques associated with the abovementioned different knowledge (Li et al., 2002). They may include, probability theory, evidence theory (Dempster-Shafer), spatial statistics, fuzzy sets, cloud model, rough sets, neural network, genetic algorithms, decision tree, exploratory learning, inductive learning, visualization, spatial online analytical mining (SOLAM), outlier detection, etc. , main techniques are briefed in Table.3.

Some of techniques (Table.3) are further developed and applied, for example, the algorithms in spatial inductive learning include AQ11 and AQ15 by Michalski, AE1 and AE9 by Hong, CLS by Hunt, ID3, C4.5 and C5.0 by Quinlan, and CN2 by Clark, etc (Di, 2001). And the implementation of data mining in spatial database is still needed to be further studied. The following is our proposed techniques, SDMKD-based image classification, cloud model, and data field.

Table.3 Techniques to be used in SDMKD

Techniques	Description
Probability theory	Mine spatial data with randomness on the basis of stochastic probabilities. The knowledge is represented as a conditional probability in the contexts of given conditions and a certain hypothesis of truth (Arthurs, 1965). Also named probability theory and mathematical statistics.
Spatial statistics	Discover sequential geometric rules from disorder data via covariance structure and variation function in the contexts of adequate samples and background knowledge (Cressie, 1991). Clustering analysis is a branch.
Evidence theory	Mine spatial data via belief function and possibility function. It is an extension of probability theory, and suitable for stochastic uncertainty based SDMKD (Shafer, 1976).
Fuzzy sets	Mine spatial data with fuzziness on the basis of a fuzzy membership function that depicts an inaccurate probability, by using fuzzy comprehensive evaluation, fuzzy clustering analysis, fuzzy control, fuzzy pattern recognition etc. (Zadeh, 1965).
Rough sets	Mine spatial data with incomplete uncertainties via a pair of lower approximation and upper approximation (Pawlak, 1991). Rough sets-based SDMKD is also a process of intelligent decision-making under the umbrella of spatial data.
Neural network	Mine spatial data via a nonlinear, self-learning, self-suitable, parallel, and dynamic system composed of many linked neurons in a network. The set of neurons collectively find out rules by continuously learning and training samples in the network (Gallant, 1993).
Genetic algorithms	Search the optimized rules from spatial data via three algorithms simulating the replication, crossover, and aberrance of biological evolution (Buckless, Petry, 1994).
Decision tree	Reasoning the rules via rolling down and drilling up a tree-structured map, of which a root node is the mining task, item and branch nodes are mining process, and leaf nodes are exact data sets. After pruning, the hierarchical patterns are uncovered (Quinlan, 1986)
Exploratory learning	Focusing on data characteristics by analyzing topological relationships, overlaying map-layers, matching images, buffering features (points, lines, polygon) and optimizing road (Dasu, 2003).
Spatial inductive learning	Comes from machine learning. Summarize and generalize spatial data in the context of given background that is from users or a task of SDMKD.The algorithms require that the training data be composed of several tuples with various attributes. And one of the attributes of each tuples is the class label (Muggleton, 1990).
Visualization	Visually mine spatial data by computerized visualization techniques that make abstract data and complicated algorithms change into concrete graphics, images, animation etc., which user may sense directly in eyes (Soukup, Davidson, 2002).
SOLAM	Mine data via online analytical processing and spatial data warehouse. Based on multidimensional view and web. It is a tested mining that highlights executive efficiency and timely responsibility to commands (Han, 1998)
Outlier detection	Extract the interesting exceptions from spatial data via statistics, clustering, classification, and regression besides the common rules (Shekhar, Lu, Zhang, 2003).

3.1 SDMKD-based image classification

Based on the integration of remote sensing and GIS (Li, Guan, 2002), this subsection presents an approach to combine spatial inductive learning with Bayesian image classification in a loose manner, which takes learning tuple as mining granularities for learning knowledge subdivide classes into subclasses, i.e. pixel granularity and polygon granularity, and selects class probability values of Bayesian classification, shape features, locations and elevations as the learning attributes. GIS data are used in training area selection for Bayesian classification, generating learning data of two granularities, and testing area selection for classification accuracy evaluation. And the ground control points for image rectification are also chosen from GIS data. It implements inductive learning in spatial data mining via C5.0 algorithm on the basis of learning granularities. Figure 1 shows the principle of the method.

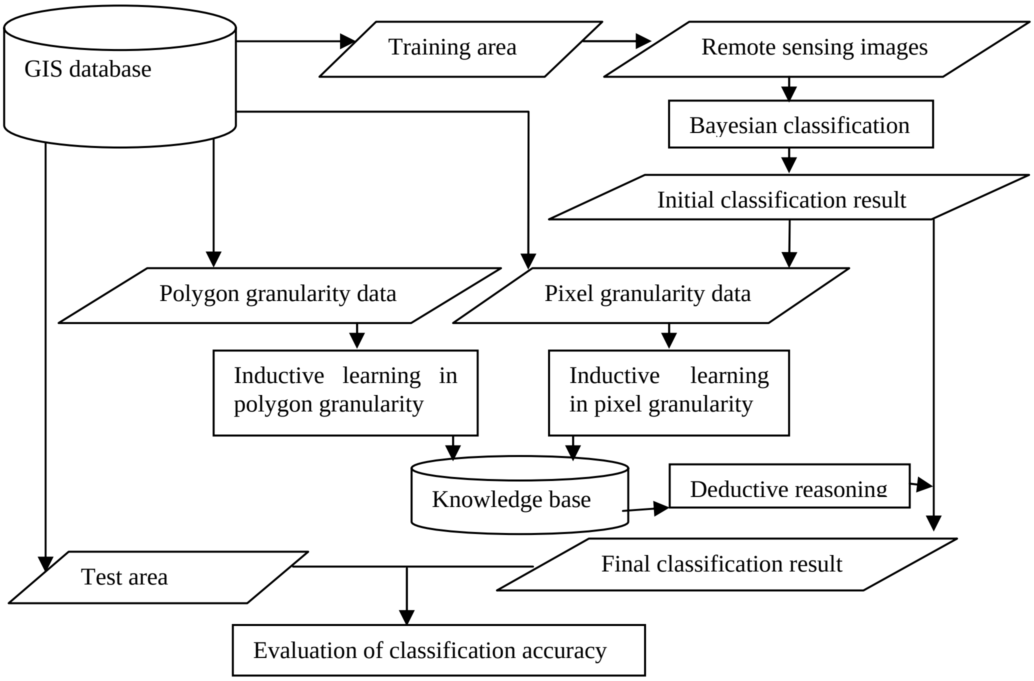


Figure 1. Flow diagram of remote sensing image classification with inductive learning

In Figure 1, first, the remote sensing images are classified initially by Bayesian method before using the knowledge, and the probabilities of each pixel to every class are retained. Second, inductive learning is conducted by the learning attributes. Learning with probability simultaneously makes use of the spectral information of a pixel and the statistical information of a class since the probability values are derived from both of them. Third, the knowledge on the attributes of general geometric features, spatial distribution patterns and spatial relationships is further discovered from GIS database, e.g. the polygons of different classes. For example, the water areas in the classification image are converted from pixels to polygons by raster to vector conversion, and then the location and shape features of these polygons are calculated. Finally, the polygons are subdivided into subclasses by deductive reasoning based on the knowledge, e.g. class water is subdivided into subclasses such as river, lake, reservoir and pond. In Figure 2, the final classification results are obtained by post-processing of the initial classification results by deductive reasoning. Except the class label attribute, the attributes for deductive reasoning are the same as that in inductive learning. The knowledge discovered by C5.0 algorithm is a group of classification rules and a default class, and each rule is together with a confidence value between 0 and 1. According to how the rule is activated that the attribute values match the conditions of this rule, the deductive reasoning adopts four strategies: (1) If only one rule is activated, then let the final class be the same as this rule; (2) If several rules are activated, then let the final class be the same as the rule with the maximum confidence; (3) If several rules are activated and the confidence values are the same, then let the final class be the same as the rule with the maximum coverage of learning samples; and (4) If no rule is activated, then let the final class be the default class.

3.2 Cloud model

The cloud model is a model of the uncertainty transition between qualitative and quantitative analysis, i.e. a mathematical model of the uncertainty transition between a linguistic term of a qualitative concept and its numerical representation data. A piece of cloud is made up of lots of cloud drops, visible shape in a whole, but fuzzy in detail, which is similar to the natural cloud in the sky. So the terminology cloud is used to name the uncertainty transition model proposed here. Any one of the cloud drops is a stochastic mapping in the discourse universe from qualitative concept, i.e. a specified realization with uncertain factors. With the cloud model, the mapping from the discourse universe to the interval [0,1] is a one-point to multi-point transition, i.e. a piece of cloud while not a membership curve. As well, the degree that any cloud drop represents the qualitative concept can be specified. Cloud model may mine spatial data with both fuzzy and stochastic uncertainties, and the discovered knowledge is close to human thinking. Now, in geo-spatial science, the cloud model has been further explored to spatial intelligent query, image interpretation, land price discovery, factors selection, mechanism of spatial data mining, and landslide-monitoring (Li, Du, 2005；Wang, 2002) etc.

The cloud model well integrates the fuzziness and randomness in a unified way via three numerical characteristics, Expected value (Ex), Entropy (En), and Hyper-Entropy (He). In the discourse universe, Ex is the position corresponding to the center of the cloud gravity, whose elements are fully compatible with the spatial linguistic concept; En is a measure of the concept coverage, i.e. a measure of the spatial fuzziness, which indicates how many elements could be accepted to the spatial linguistic concept; and He is a measure of the dispersion on the cloud drops, which can also be considered as the entropy of En. In the extreme case, {Ex, 0, 0}, denotes the concept of a deterministic datum where both the entropy and hyper entropy equal to zero. The greater the number of cloud drops, the more deterministic the concept. Figure 2 shows the three numerical characteristics of the linguistic term “displacement is 9 millimeters (mm) around”. Given three numerical characteristics Ex, En and He, the cloud generator can produce as many drops of the cloud as you would like.

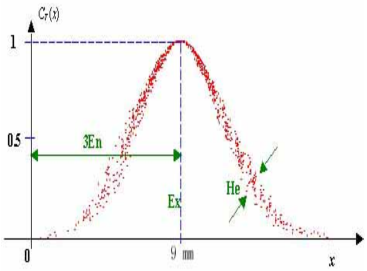


Figure 2.Three Numerical characteristics

The above three visualization methods are all implemented with the forward cloud generator in the context of the given {Ex, En, He}. Despite of the uncertainty in the algorithm, the positions of cloud drops produced each time are deterministic. Each cloud drop produced by the cloud generator is plotted deterministically according to the position. On the other hand, it is an elementary issue in spatial data mining that spatial concept is always constructed from the given spatial data, and spatial data mining aims to discover spatial knowledge represented by a cloud from the database. That is, the backward cloud generator is also necessary. It can be used to perform the transition from data to linguistic terms, and may mine the integrity {Ex, En, He} of cloud drops specified by many precise data points. Under the umbrella of mathematics, the normal cloud model is common, and the functional cloud model is more interesting. Because it is common and useful to represent spatial linguistic atoms (Li et al., 2001), the normal compatibility cloud will be taken as an example to study the forward and backward cloud generators in the following.

The input of the forward normal cloud generator is three numerical characteristics of a linguistic term, (Ex, En, He), and the number of cloud-drops to be generated, N, while the output is the quantitative positions of N cloud drops in the data space and the certain degree that each cloud-drop can represent the linguistic term. The algorithm in details is:

[1] Produce a normally distributed random number En’ with mean En and standard deviation He;

[2] Produce a normally distributed random number x with mean Ex and standard deviation En’;

[3] Calculate $y = e^{-\frac{(x-Ex)^2}{2(En')^2}}$;

[4] Drop (xi, yi) is a cloud-drop in the discourse universe;

[5] Repeat step [1]-[4] until N cloud-drops are generated.

Simultaneously, the input of the backward normal cloud generator is the quantitative positions of N cloud-drops, xi (i=1,…,N), and the certainty degree that each cloud-drop can represent a linguistic term, yi(i=1,…,N), while the output is the three numerical characteristics, Ex, En, He, of the linguistic term represented by the N cloud-drops. The algorithm in details is:

[1] Calculate the mean value of xi(i=1,…,N), $Ex = \frac{1}{N}\sum_{i=1}^{N} x_i$;

[2] For each pair of (xi, yi), calculate $En_i = \sqrt{-\frac{(x_i - Ex)}{2\ln y_i}}$;

[3] Calculate the mean value of Eni (i=1,…, N), $En = \frac{1}{N}\sum_{i=1}^{N} En_i$;

[4] Calculate the standard deviation of Eni, $He = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(En_i - En)^2}$.

With the given algorithms of forward and backward cloud generators, it is easy to build the mapping relationship inseparably and interdependently between qualitative concept and quantitative data. The cloud model improves the weakness of rigid specification and too much certainty, which comes into conflict with the human recognition process, appeared in commonly used transition models. Moreover, it performs the interchangeable transition between qualitative concept and quantitative data through the use of strict mathematic functions, the preservation of the uncertainty in transition makes cloud model well meet the need of real life situation. Obviously, the cloud model is not a simple combination of probability methods and fuzzy methods.

3.3 Data field

The obtained spatial data are comparatively incompleteness. Each datum in the concept space has its own contribution in forming the conception and the concept hierarchy. So it is necessary for the observed data to radiate their data energies from the sample space to their parent space. In order to describe the data radiation, data field is proposed.

Spatial data radiate energies into data field. The power of the data field may be measured by its potential with a field function. This is similar with the electric charges contribute to form the electric field that every electric charge has effect on electric potential everywhere in the electric field. So the function of data field can be derived from the physical fields. The potential of a point in the number universe is the sum of all data potentials.

$$p = k \cdot e^{-\sum_{i=1}^{N}\frac{r_i^2}{\rho_i}}$$

where, k is a constant of radiation gene, r_i is the distance from the point to the position of the ith observed data, ρ_i is the certainty of the ith data, and N is the amount of the data. With a higher certainty, the data may have greater contribution to the potential in concept space. Besides them, space between the neighbor isopotential, computerized grid density of Descartes coordinate, etc. may also make their contributions to the data field.

4. APPLICABILITY AND EXAMPLES

SDMKD may be applied in many spatial data referenced fields. Here are our three studied cases, remote sensing classification, Landslide monitoring data mining, and spatial uncertain reasoning.

4.1 Remote sensing image classification

The land use classification experiment is performed in the Beijing area using SPOT multi-spectral image and 1:100000 land use database. Discover knowledge from GIS database and remote sensing image data can improve land use classification. Two kinds of knowledge on land use and elevation will be discovered from GIS database, which are applied to subdivide water and green patches respectively. Pixel granularity is to subdivide green patches, and polygon granularity is to subdivide water.

The original image is 2412 by 2399 pixels and three bands, which was obtained in 1996. The land use database was built before 1996, which has land use, contour, road, and annotation layers (Figure 3). The original image is stretched and rectified to the GIS data. The image is 2834 by 2824 pixels after rectification, which is used as the source image for classification. We use ArcView 3.0a, ENVI 3.0 and See5.1.10, which is developed based on C5.0 algorithm by Rulequest Cooperation. And also we developed several programs for data processing and format conversion using Microsoft Visual C++6.0.

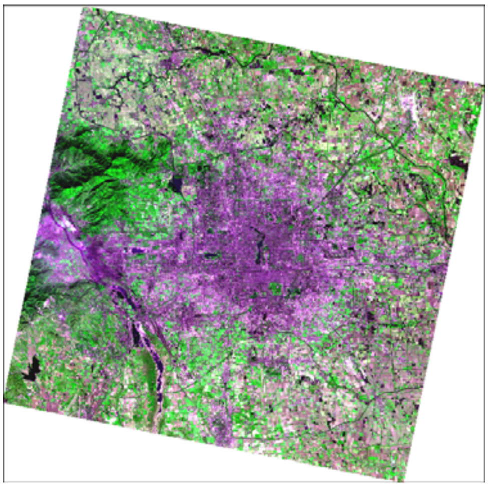


Figure 3. Original SPOT image (resampled)

For the sake of comparison, only the Bayesian method is applied to classify the image at first. The rectified image is overlaid with land use data layers, and the training and test areas are interactively selected. Then the image is classified into 8 classes, such as water, paddy, irrigated field, dry land, vegetable field, garden, forest and residential area. As shown in the confusion matrix (Table 4), the overall accuracy is 77.6199%. Water, paddy, irrigated field, residential area and vegetable field are classified with high accuracy. The vegetable field is easily distinguished from other green patches because it is lighter than the others are. Dry land, garden, and forest are confused seriously and the accuracy is 65.58%, 48.913% and 59.754% respectively. And some forest shadows are misclassified as waters.

Spatial inductive learning is mainly used in two aspects to improve the Bayesian method in land use classification, one is to discover rules to subdivide waters in polygon granularity, the other in to discover rules to reclassify dry land, garden and forest in pixel granularity. The land use layer (polygon) and contour layer (line) are selected for these purposes. Because there are few contours and elevation points, it is difficult to interpolate a DEM accurately, instead, the contours are converted to height zones, such as <50m, 50-100m, 100-200m and >200m, which are represented by polygons.

In the learning for subdividing waters, several attributes of the polygons in land use layer are s calculated as condition attributes, such as area, location of the center, compactness (perimeter^2/(4π•area)), height zone, etc. The classes are river (code 71), lake (72), reservoir (73), pond (74) and forest shadow (99). 604 water polygons are learned, 10 rules are discovered (Table 5).

```
Rule 1: (cover 19)
        compactness > 7.190882
        height = lt50
        ->  class 71  [0.952]

Rule 2: (cover 5)
        Xcoord > 453423.5
        Xcoord <= 455898.7
        Ycoord > 4414676
        Ycoord <= 4428958
        compactness > 2.409397
        compactness <= 7.190882
        ->  class 72  [0.857]

Rule 3: (cover 33)
        Xcoord <= 455898.7
        Ycoord > 4414676
        Ycoord <= 4428958
        compactness <= 7.190882
        height = lt50
        ->  class 72  [0.771]

Rule 4: (cover 4)
        area > 500000
        height = 50 100
        ->  class 73  [0.667]

Rule 5: (cover 144)
        Ycoord <= 4414676
        compactness <= 7.190882
        height = lt50
        ->  class 74  [0.993]

Rule 6: (cover 213)
        Ycoord > 4428958
        compactness <= 7.190882
        height = lt50
        ->  class 74  [0.986]

Rule 7: (cover 281)
        Xcoord > 451894.7
        compactness <= 7.190882
        ->  class 74  [0.975]

Rule 8: (cover 38)
        area <= 500000
        height = 50_100
        ->  class 74  [0.950]

Rule 9: (cover 85)
        height = gt200
        ->  class 99  [0.989]

Rule 10: (cover 7)
        height = 100_200
        ->  class 99  [0.778]

Default class: 74

        ----------------------
        Evaluation (604 cases):

        Errors      7( 1.2%)
```

Table 5. Rules discovered by inductive learning

In Table 5, there are only 1.2% samples misclassified in the learning, thus the learning accuracy is 98.8%. These rules reveal the spatial distribution patterns and general shape features, etc. For example, rule 1 states "If compactness of a water polygon is greater than 7.190882, and locates in the height zone <50m, then it is a river". Here the compactness measure plays a key role to identify river from other waters. Rule 2 identifies lakes by location and compactness, rule 9 and rule 10 distinguish forest shadows from waters by height, and so on.

In the learning for reclassifying dry land, garden and forest, the condition attributes are image coordinates, heights and the probability values to the three classes that produced by Bayesian classification. One percent (2909) samples are selected randomly from the vast amount of pixels. 63 rules are discovered and the learning accuracy is 97.9%. The test accuracy is 94.4%, which was evaluated by another 1% randomly selected samples. These rules are omitted here because of paper size limitation.

After inductive learning, the Bayesian classified image is reclassified by deductive reasoning based on the discovered rules. Because Bayesian method cannot subdivide waters, only the rules to identify forest shadows from waters are used in order to compare the result with Bayesian classification. The final class is determined by the maximum confidence principle (See Figure 4).

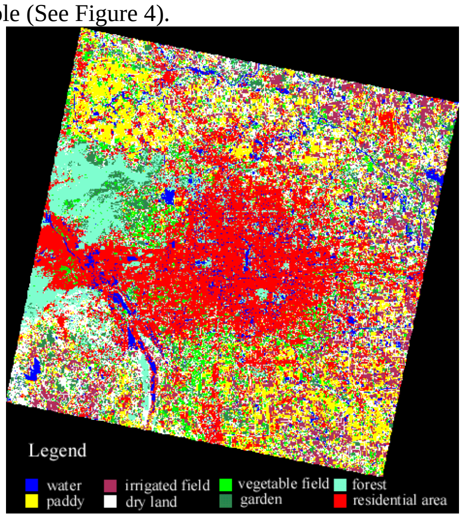


Figure 4. Final image classification (resampled)

Accuracy evaluation is accomplished using the same test areas as that in Bayesian classification. The confusion matrix is shown in Table 6. The overall accuracy of the final result is 88.8751%. The accuracy of dry land, garden and forest is 69.811%, 78.561% and 91.81% respectively. Comparing the final result with the result produced only by Bayesian classification, the overall accuracy increased 11.2552 percent and the accuracy of dry land, garden and forest increased 6.231%, 29.648% and 32.056% respectively.

The result of land use classification shows that the overall accuracy is increased more than 11%, and the accuracy of some classes, such as garden and forest, are further increased about 30%. First this indicates that the proposed SDMKD-based image classification approach of integrating Bayesian classification with inductive learning not only improves classification accuracy greatly, but also extends the classification by subdividing some classes with the discovered knowledge. The approach is implemented feasibly and effectively. Second, spatial inductive learning in GIS databases can resolve the problems of spectral confusion to a great extent. SDMKD is very helpful to improve Bayesian classification, and using probability values generates more accurate learning results than using the pixel values directly. Third, when utilizing the knowledge discovered from GIS databases, it is likely to be more beneficial to improve remote sensing image classification than the conventional way that GIS data are utilized directly in pre- or post- processing of image classification.

Table 4. Confusion matrix of Bayesian classification

Classified	Real class								
	Water	Paddy	Irrigated field	Dry land	Vegetable field	Garden	Forest	Residential area	Sum
Water	3.900	0.003	0.020	0.013	0.002	0.021	2.303	0.535	6.797
Paddy	0.004	8.496	0.087	0.151	0.141	0.140	0.103	0.712	9.835
Irrigated field	0.003	0.016	10.423	0.026	0.012	0.076	0.013	0.623	11.192
Dry land	0.063	0.48	0.172	1.709	0.361	2.226	2.292	1.080	8.384
Vegetable field	0.001	0.087	0.002	0.114	3.974	0.634	0.435	0.219	5.465
Garden	0.010	0.009	0.002	0.325	0.263	4.422	4.571	0.065	9.666
Forest	0.214	0.006	0.000	0.271	0.045	1.354	15.671	0.642	18.202
Residential area	0.132	0.039	0.127	0.080	0.049	0.168	0.839	29.024	30.459
Sum	4.328	9.135	10.834	2.689	4.846	9.041	26.227	32.901	100
Accuracy (%)	90.113	93.010	96.204	63.580	81.994	48.913	59.754	88.217	
Overall accuracy = 77.6199% Kappa coefficient = 0.7474									

Table 6. Confusion matrix of Bayesian classification combined with inductive learning

Classified	Real class								
	Water	Paddy	Irrigated field	Dry land	Vegetable field	Garden	Forest	Residential area	Sum
Water	3.900	0.003	0.020	0.012	0.002	0.019	0.139	0.535	4.631
Paddy	0.004	8.496	0.087	0.151	0.141	0.14	0.103	0.712	9.835
Irrigated field	0.003	0.016	10.423	0.026	0.012	0.076	0.013	0.623	11.192
Dry land	0.063	0.480	0.172	1.877	0.361	0.205	0.149	1.080	4.386
Vegetable field	0.001	0.087	0.002	0.114	3.974	0.634	0.435	0.219	5.465
Garden	0.009	0.009	0.002	0.210	0.263	7.102	0.470	0.065	8.131
Forest	0.215	0.006	0.000	0.218	0.045	0.696	24.079	0.642	25.899
Residential area	0.132	0.039	0.127	0.080	0.049	0.168	0.839	29.024	30.46
Sum	4.328	9.135	10.834	2.689	4.846	9.041	26.227	32.901	100
Accuracy (%)	90.113	93.01	96.204	69.811	81.994	78.561	91.81	88.217	
Overall accuracy = 88.8751% Kappa coefficient = 0.8719									

4.2 Landslide monitoring

Baota landslide locates in Yunyang, Chongqing, China. And the landslide monitoring started from June 1997. Up to now, this database on the displacements has amounted to 1Gigabytes, and all the attributes are numerical displacements, i.e. dx, dy, and dh. Respectively, the properties of dx, dy, and dh, are the measurements of displacements in X direction, Y direction and H direction of the landslide-monitoring points, and |dx|, |dy| an |dh| are their absolute values. In the following, it is noted that all spatial knowledge is discovered from the databases with the properties of dx, dy, and dh, while |dx|, |dy| an |dh| are only used to visualize the results of spatial data mining. And the properties of dx are the major examples. The linguistic terms of different displacements on dx, dy and dh may be depicted by the pan-concept hierarchy tree (Figure 6) in the conceptual space, which are formed by cloud models (Figure 11). It can be seen from Figure 6 and Figure 10 that the nodes “very small” and “small” both have the son node “9 millimeters around”, which indicates that the pan-concept hierarchy tree is a pan-tree structure.

From the observed landslide-monitoring values, the backward cloud generator can mine Ex, En and He of the linguistic term indicating the level of that landslide displacement, i.e. gain the concept with the forward cloud generator. Then, with the three gained characteristics, the forward cloud generator can reproduce as many deterministic cloud-drops as you would like, i.e. produce synthetic values with the backward cloud generator.

According to the landslide-monitoring characteristics, let the linguistic concepts of “smaller (0~9mm), small (9~18mm), big(18~27mm), bigger(27~36mm), very big(36~50mm), extremely big(＞50mm)” with Ex, “lower (0~9), low(9~18), high(18~27), higher(27~36), very high(36~50), extremely big(＞ 50)”with En, “more stable (0~9), stable (9~18), instable(18~27), more instable (27~36), very instable (36~50), extremely instable (＞50)” with He respectively depict the movements, scattering levels and stabilities of the displacements. Further, let the |dx|-axis, |dy|-axis respectively depict the absolute displacement values of the landslide- -monitoring points. The certainty of the cloud drop (dxi, CT(dxi)), CT(dxi) is also defined as,

$$C_T(dx_i) = \frac{dx_i - \min(dx)}{\max(dx) - \min(dx)}$$

where, max(dx) and min(dx) are the maximum and minimum of dx ={dx1 , dx2 , …, dxi, …, dxn }. Then the rules on Baota landslide-monitoring in X direction can be discovered from the databases in the conceptual space.

BT11: the displacements are big south, high scattered and instable;
BT12: the displacements are big south, high scattered and very instable;
BT13: the displacements are small south, lower scattered and more stable;
BT14: the displacements are smaller south, lower scattered and more stable;
BT21: the displacements are extremely big south, extremely high scattered and extremely instable;
BT22: the displacements are bigger south, high scattered and instable;
BT23: the displacements are big south, high scattered and extremely instable;
BT24: the displacements are big south, high scattered and more instable;
BT31: the displacements are very big south, higher scattered and very instable;
BT32: the displacements are big south, low scattered and more instable;
BT33: the displacements are big south, high scattered and very instable; and
BT34: the displacements are big south, high scattered and more instable.

Figure 5 visualizes the above rules, where different rules represented by ellipses with different colors, and the number in ellipses denotes the number of rules. The generalized result at a higher hierarchy than that of Figure 4 in the feature space is the displacement rule of the whole landslide, i.e. “the whole displacement of Baota landslide are bigger south (to Yangtze River), higher scattered and extremely instable”. Because large amounts of consecutive data are replaced by discrete linguistic terms, the efficiency of spatial data mining can be improved. Meanwhile, the final result mined will be stable due to the randomness and fuzziness of concept indicated by the cloud model.

All the above landslide-monitoring points form the potential field and the isopotential lines spontaneously in the feature space. Intuitively, these points can be grouped naturally into clusters. These clusters represent different kinds of spatial objects recorded in the database, and naturally form the cluster graph. Figure 5(a) further shows all points’ potential.

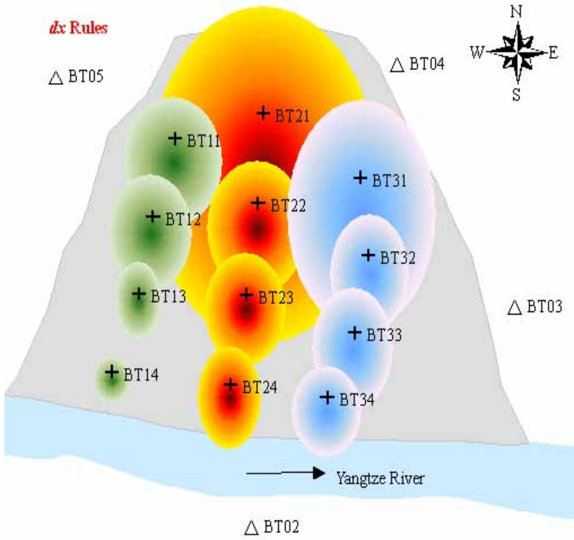


Figure 5. Spatial rules on Baota landslide-monitoring points

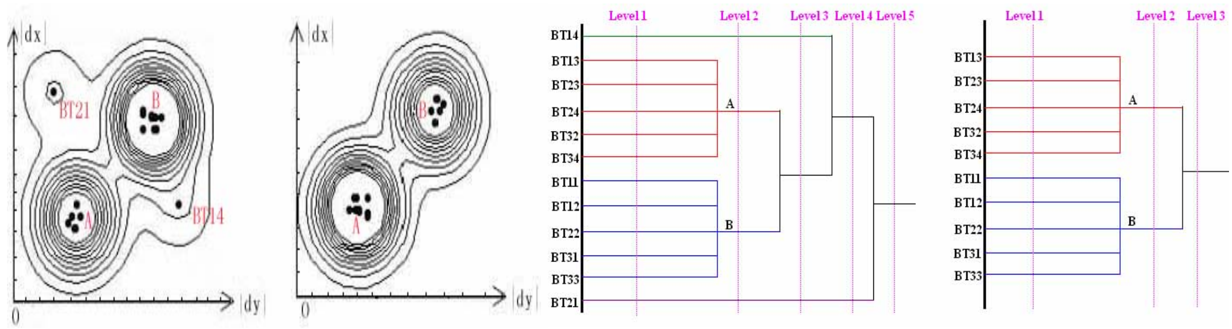
In Figure 6(a), all points’ potentials form the potential field and the isopotential lines spontaneously. Seen from this figure, when the hierarchy jumps up from Level 1 to Level 5, i.e. from the fine granularity world to the coarse granularity world, these landslide-monitoring points can be intuitively grouped naturally into different clusters at different hierarchies of variant levels. That is,

[1] No clusters at the hierarchy of Level 1;
[2] Four clusters at the hierarchy of Level 2 that are cluster BT14, cluster A (BT13, BT23, BT24, BT32, BT34), cluster B (BT11, BT12, BT22, BT31, BT33) and cluster BT21;
[3] Three clusters at the hierarchy of Level 3 that are cluster BT14, cluster (A, B) and cluster BT21;
[4] Two clusters at the hierarchy of Level 4 that are cluster (BT14, (A, B)) and cluster BT21; and
[5] One cluster at the hierarchy of Level 5 that is cluster ((BT14, (A, B)), BT21).

Respectively, they denote, [1] the displacements of landslide-monitoring points are separate at the lowest hierarchy; [2] at the lower hierarchy, the displacements of landslide-monitoring points (BT13, BT23, BT24, BT32, BT34) have the same trend of "the displacements are small", and the same with (BT11, BT12, BT22, BT31, BT33) of "the displacements are big", while BT14, BT21 show the different trend with both of them, and each other i.e. the exceptions, "the displacement of BT14 is smaller", "the displacement of BT21 is extremely big"; [3] when the hierarchy becomes high, the displacements of landslide-monitoring points (BT13, BT23, BT24, BT32, BT34) and (BT11, BT12, BT22, BT31, BT33) have the same trend of "the displacements are small", however, BT14, BT21 are still unable to be grouped into this trend; [4] When the hierarchy gets higher, the displacements of landslide-monitoring point BT14 can be grouped into the same trend of (BT13, BT23, BT24, BT32, BT34) and (BT11, BT12, BT22, BT31, BT33) that is "the displacements are small", however, BT21 is still an outlier; [5] the displacements of landslide-monitoring points are unified at the highest hierarchy, that is, the landslide is moving.

These show different "rules plus exceptions" at different changes from the fine granularity world to the coarse granularity world. That is to say, the clustering or associations between attributes at different cognitive levels make many combinations, "rules plus exceptions", showing the discovered knowledge with different information granularities. When the exceptions BT14, BT21 are granted to eliminate, the rules and the clustering process will be more obvious (Figure 6(b)). Simultaneously, these clusters represent different kinds of landslide-monitoring points recorded in the database. And they can naturally form the cluster spectrum figures as Figure 6(c) and Figure 6(d). Seen from these two figures, the displacements of landslide-monitoring points (BT13, BT23, BT24, BT32, BT34) and (BT13, BT23, BT24, BT32, BT34) firstly compose two new classes, cluster A and cluster B, then the two new classes compose a larger class with cluster BT14, and they finally compose the largest class with cluster BT21, during the process of which the mechanism of spatial data mining is still "rules plus exceptions". In other words, the so-called spatial data mining is particular views for a viewer to look at the spatial database on the displacements of Baota landslide-monitoring by different distances only, and a longer distance leads a piece of more meta-knowledge to be discovered.

When the Committee of Yangtze Riverinvestigated in the region of Yunyang Baota landslide, they found out that the landslide had moved to Yangtze River. By the landslide-monitoring point BT21, a small size landslide had taken place. Now there are still two pieces of big rift. Especially, the wall rift of the farmer G. Q. Zhang's house is nearly 15 millimeters. These results from the facts match the discovered spatial knowledge very much, which indicates that the techniques of cloud model-based spatial data mining are practical and creditable.



(a) All points' potential (b) Potential without exceptions (c) All points' cluster graph (d) Cluster graph without exceptions
Figure 6. Landslide-monitoring points' potential form the clusters and cluster graph

4.3 Uncertainty reasoning

Uncertainty reasoning may include one-rule reasoning and multi-rule reasoning.

4.3.1 One-rule Reasoning

If there is only one factor in rule antecedent, we call the rule one-factor rule. Figure 6 is a one-factor one-rule generator for the rule "If A, then B". CG_A is the X-conditional cloud generator for linguistic term A, and CG_B is the Y conditional cloud generator for linguistic term B. Given a certain input x, CG_A generates random values μ_i . These values are considered as the activation degree of the rule and input to CG_B. The final outputs are cloud drops, which forms a new cloud.

Combining the algorithm of X and Y conditional cloud generators (Li, 2005), we present the following algorithm for one-factor one-rule reasoning.

[1] $En'A = G(En_A, He_A)$ //Produce random values that satisfy with the normal distribution probability of mean En_A and standard deviation He_A.

[2] Calculate
$$\mu = \exp\left[-\frac{(x - Ex_A)^2}{2{En'_A}^2}\right]$$

[3] $En'_B = G(En_B, He_B)$ //Produce random values that satisfy with the normal distribution probability of mean En_B and standard deviation He_B.

[4] Calculate $y = Ex_B \pm \sqrt{-2\ln(\mu)}En'_B$, let (y, μ) be cloud drops. If $x<Ex_A$, "-"is adopted, while id $x>Ex_A$, "+" is adopted.

[5] Repeat step [1] to [4], generate cloud drops as many as you want.

Figure 7 is the output cloud of a one-factor one-rule generator with one input. We can see that the cloud model based reasoning generate uncertain result. The uncertainty of the linguistic terms in the rule is propagated during the reasoning process. Since the rule output is a cloud, we can give the final result in several forms. 1) One random value; 2) Several random values as sample results; 2) Expected value, which is the mean of many sample results; 4) Linguistic term, which is represented by a cloud model, and the parameters of the model are obtained by inverse cloud generator method.

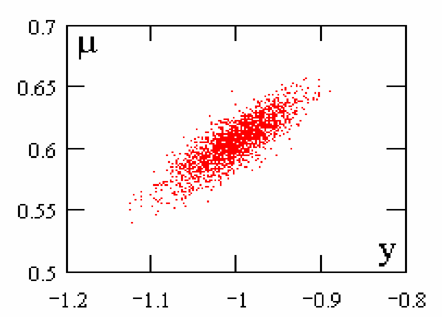


Figure 7. Output cloud of one-rule reasoning

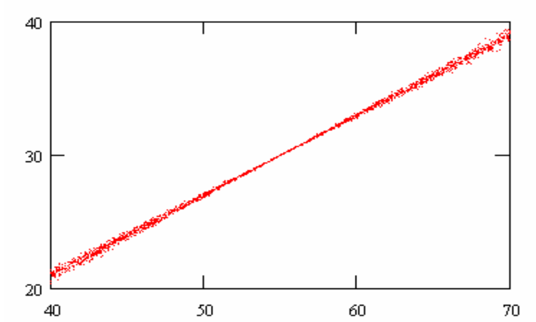


Figure 8. Input-output response of one-rule reasoning

If we input a number of values to the one-factor rule and draw the inputs and outputs in a scatter plot, we can get the input-output response graph of the one-factor one-rule reasoning (Figure 8). The graph looks like a cloud band, not a line. Closer to the expected values, the band is more focusing, while farther to the expected value, the band is more dispersed. This is consistent with human being's intuition. The above two figures and the discussion show that the cloud model based uncertain reasoning is more flexible and powerful than the conventional fuzzy reasoning method.

If the rule antecedent has two or more factors, such as "If A_1, A_2,…, A_n, then B", we call the rule multi-factor rule. In this case, a multi-dimensional cloud model represents the rule antecedent. Figure 9 is a two-factor one-rule generator, which combines a 2-dimensional X-conditional cloud generator and a 1-dimensional Y-conditional cloud generator. It is easy to give the reasoning algorithm on the basis of the cloud generator algorithms stated in section 3.2. And consequently, multi-factor one-rule reasoning is conducted in a similar way.

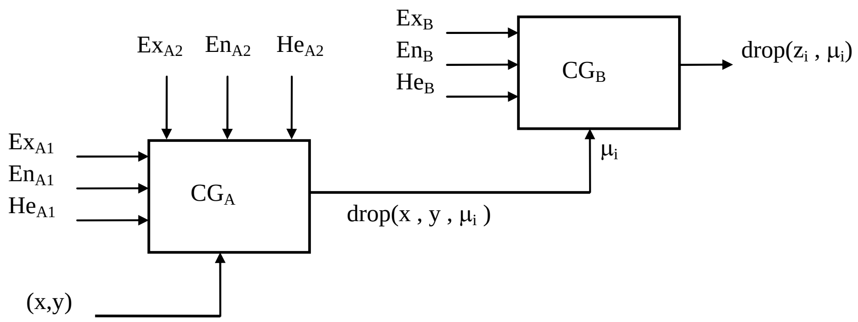


Figure 9. A two-factor one-rule generator

4.3.2 Multi-rule Reasoning

Usually, there are many rules in a real knowledge base. Multi-rule reasoning is frequently used in an intelligent GIS or a spatial decision support system. Figure 10 is a one-factor multi-rule generator, and the algorithm is as follows.

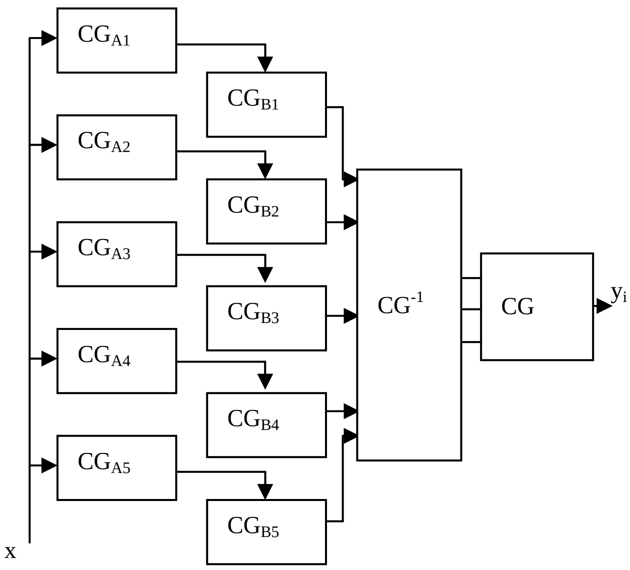


Figure 10. One-factor 5-rule generator

The algorithms of the one-factor multi-rule generator (Figure 10) are as follows.

[1] Given input x, determine how many rules are activated. If $Ex_{Ai} - 3En_{Ai} < x < Ex_{Ai} + 3En_{Ai}$, then rule i is activated by x.

[2] If only one rule activated, output a random y by the one-factor one-rule reasoning algorithm. Go to step [4]

[3] If two or more rules are activated, firstly each rule outputs a random value by the one-factor one-rule reasoning algorithm, and a virtual cloud is constructed by the geometric cloud generation method in section 3.2. A cloud generator algorithm is conducted to output a final result y with the three parameters of the geometric cloud. Because the expected value of the geometric cloud is also a random value, we can take the expected value as the final result for simplicity. Go to step [4]

[4] Repeat step [1] to [3], generate outputs as many as you want.

The main idea of the multi-rule reasoning algorithm is that when several rules are activated simultaneously, a virtual cloud is constructed by the geometric cloud method. Because the property of least square fitting, the final output is more likely to close to the rule of high activated degree. This is consistent with the human being's intuition. Figure 11 is a situation of two rules activated, and Figure 12 is the situation of three rules activated. Only the mathematical expected curves are drawn for clearness, and the dash curves are the virtual cloud. The one-factor multi-rule reasoning method can be easily extended to multi-factor multi-rule reasoning on the basis of multi-dimensional cloud models. We omitted the algorithm here.

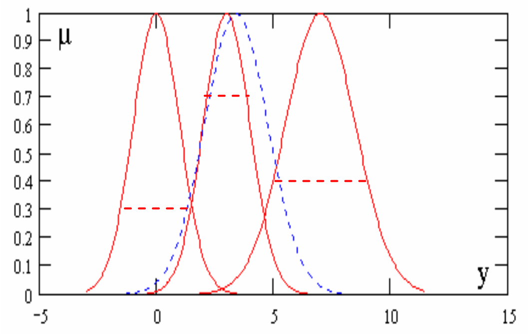


Figure 11.A situation of two activated rules

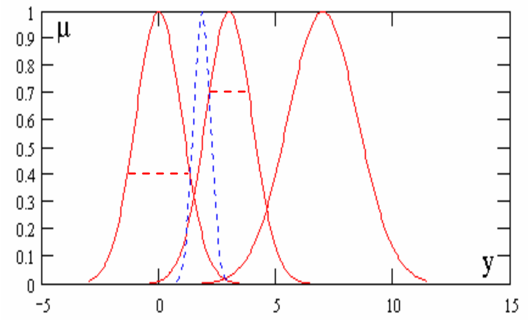


Figure 12. A situation of three activated rules

4.3.3 An illustrative example

The following is an illustrative example of multi-factor multi-rule reasoning. Suppose we have the following five rules to describe the terrain features qualitatively in a GIS.

Rule 1: If location is southeast, then elevation is low.

Rule 2: If location is northeast, then elevation is low to medium.

Rule 3: If location is central, then elevation is medium.

Rule 4: If location is southwest, then elevation is medium to high.

Rule 5: If location is northwest, then elevation is high.

Rule input is the location. Because the location is determined by x and y coordinates, the linguistic terms for location are represented by 2-dimensional clouds (Figure 13). Rule output is the elevation, which is represented by 1-dimensional clouds (Figure 14).

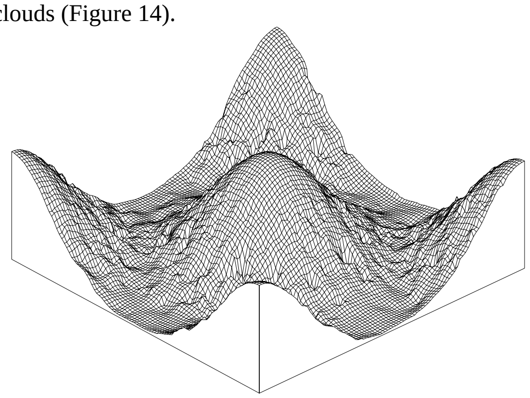


Figure 13. 2-dimensional clouds to represent linguistic term of location

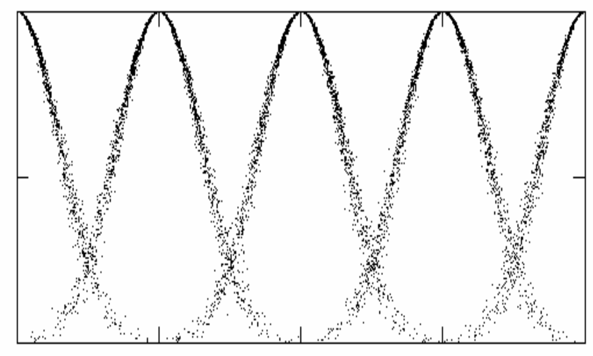


Figure 14. 1-dimensional clouds to represent linguistic term of elevation

Figure 15 is the input-output response surface of the rules. The surface is an uncertain surface and the roughness is uneven. Closer to the center of the rules (the expected values of the clouds), the surface is more smooth showing the small uncertainty; while at the overlapped areas of the rules, the surface is more rough showing large uncertainty. This proves that the multi-factor multi-rule reasoning also represents and propagates the uncertainty of the rules as the one-factor one-rule reasoning does.

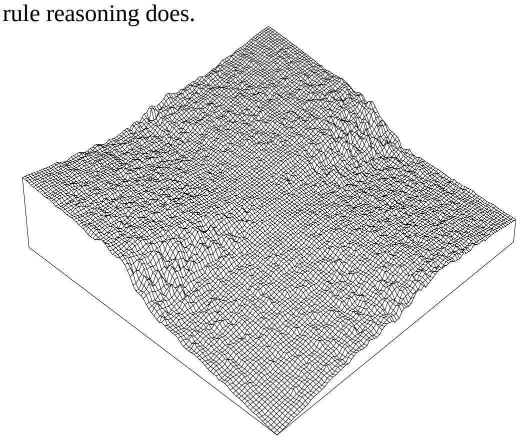


Figure 15. The input-output response surface of the rules

5 CONCLUSIONS AND DISCUSSIONS

After the concept and principle of spatial data mining and knowledge discovery (SDMKD) were briefed, this paper proposed three mining techniques, SDMKD-based image classification, cloud model and data field, together with applicable examples, i.e. remote sensing image classification, Baota landslide-monitoring data mining and spatial uncertain reasoning.

The technical progress in computerized spatial data acquisition and storage resulted in the growth of vast databases, which made a branch of data mining, SDMKD, developed in geo-spatial science. This paper took the mechanism of SDMKD as a process of discovering a form of rules plus exceptions at hierarchal view-angles with various thresholds.

SDMKD-based image classification integrated spatial inductive learning from GIS database and Bayesian classification. In the experimental results of remote sensing image classification, the overall accuracy was increased more than 11%, and the accuracy of some classes, such as garden and forest, was further increased about 30%. For the intelligent integration of GIS and remote sensing, it is encouraged discover knowledge from spatial databases and apply the knowledge in image interpretation for spatial data updating. The applications of inductive learning in other image data sources, e.g. TM, SAR, and the applications of other SDMKD techniques in remote sensing image classification, may be the valuable future directions of further study.

Cloud model integrated the fuzziness and randomness in a unified way via the algorithms of forward and backward cloud generators in the contexts of three numerical characteristics, {Ex, En, He}. It took advantage of human natural language, and might search for the qualitative concept described by natural language to generalize a given set of quantitative datum with the same feature category. Moreover, the cloud model could act as an uncertainty transition between a qualitative concept and its quantitative data. With this method, it was easy to build the mapping relationship inseparably and interdependently between qualitative concept and quantitative data, and the final discovered knowledge with hierarchy could match different demands from different level users. Data field radiated the energy of observed data to the universe discourse, considering the function of each data in SDMKD. In the context of data field, the conceptual space represented different concepts in the same characteristic category, while the feature space depicts were complicated spatial objects with multi-properties. The isopotential of all data automatically took shape a concept hierarchy. At the same time, the isopotential of all objects automatically took shape clusters and clustering hierarchy. The clustering or associations between attributes at different cognitive levels might make many combinations.

The experimental results on Baota landslide monitoring show the cloud model-based spatial data mining can reduce the task complexity, improve the implementation efficiency, and enhance the comprehension of the discovered spatial knowledge. On the basis of the cloud model and data field, the experimental results on Baota landslide-monitoring indicated cloud model- and data field- based SDMKD could reduce the task complexity, improve the implementation efficiency, and enhance the comprehension of the discovered spatial knowledge. Cloud model was a model for concept representation and uncertainty handling. A series of uncertainty reasoning algorithms were presented based on the algorithms of cloud generators and conditional cloud generators. The example case of spatial uncertainty reasoning showed that knowledge representation and uncertainty reasoning based on cloud model were more flexible, effective and powerful than the conventional fuzzy theory based methods.

ACKNOWLEDGEMENTS

This study is supported by National Basic Research Program of China (973 Program) “Transform mechanism among observed data, spatial information, and geo-knowledge”, National Natural Science Foundation of China (70231010), China Postdoctoral Science Foundation (2004035360), and Wuhan University (216-276081).

REFERENCES

1. ARTHURS, A. M., 1965, *Probability theory* (London:Dover Publications)
2. BUCKLESS,B.P., PETRY,F.E., 1994, *Genetic Algorithms* (Los Alamitos, California: IEEE Computer Press)
3. CRESSIE,N., 1991, *Statistics for Spatial Data*. (*New York*: John Wiley and Sons)
4. DASU, T., 2003, *Exploratory Data Mining And Data Cleaning* (New York: John Wiley & Sons, Inc.)
5. DI, K.C., 2001, Spatial Data Mining and Knowledge Discovery (Wuhan: Wuhan University Press)
6. ESTER, M. et al., 2000, Spatial data mining: databases primitives, algorithms and efficient DBMS support. Data Mining and Knowledge Discovery, 4, 193-216
7. GALLANT, S. I., 1993, Neural Network Learning And Expert Systems (Cambridge: MIT Press)
8. HAN, J., KAMBER, M., 2001, Data Mining: Concepts and Techniques (San Francisco: Academic Press)
9. HAN, J., 1998, Towards on-line analytical mining in large databases. ACM SIGMOD-Record
10. HAINING, R., 2003, Spatial Data Analysis: Theory and Practice (Cambridge: Cambridge University Press)
11. HAWKINS,D., 1980, Identifications of outliers(London: Chapman and Hall)
12. HONG J.R., 1997, Inductive Learning – algorithm, theory and application (Beijing: Science Press)
13. LI D.R., GUAN Z.Q., 2002, Integration and realization of spatial information system (Wuhan: Wuhan University Press)
14. LI, D.R., CHENG, T., 1994, KDG: Knowledge Discovery from GIS - Propositions on the Use of KDD in an Intelligent GIS. In Proc. ACTES, The Canadian Conf. on GIS,
15. LI D.R., et al., 2001, On spatial data mining and knowledge discovery (SDMKD), Geomatics and Information Science of Wuhan University, 26(6):491-499
16. LI D.R., et al., 2002, Theories and technologies of spatial data mining and knowledge discovery. Geomatics and Information Science of Wuhan University, 27(3): 221-233
17. LI D.R., WANG S.L., LI D.Y.,2005, Theories and Applications of Spatial Data Mining (Beijing: Science Press)
18. LI D.Y., DU Y., 2005, Artificial Intelligence with Uncertainty (Beijing: National Defense Industry Press)
19. MILLER H J, HAN J, 2001, Geographic Data Mining and Knowledge Discovery (London and New York: Taylor and Francis)
20. MUGGLETON, S., 1990, *Inductive Acquisition Of Expert Knowledge* (Wokingham, England: Turing Institute Press in association with Addison-Wesley)
21. PAWLAK,Z., 1991, Rough Sets: Theoretical Aspects Of Reasoning About Data (London: Kluwer Academic Publishers)
22. QUINLAN, J., 1986, Introduction of decision trees. Machine Learning, (5), 239-266
23. SHAFER, G., 1976, *A Mathematical Theory of Evidence* (Princeton: Princeton University Press)
24. SHEKHAR S., LU C.T., Zhang P., 2003, A Unified Approach to Detecting Spatial Outliers. GeoInformatica,7(2):139-166
25. SOUKUP T., DAVIDSON I., 2002, Visual Data Mining: Techniques and Tools for Data Visualization and Mining (New York: Wiley Publishing, Inc.)
26. WANG, S.L., 2002, Data Field and Cloud Model - Based Spatial Data Mining and Knowledge Discovery. Ph.D. Thesis (Wuhan: Wuhan University)
27. ZADEH, L.A., 1975, The concept of linguistic variable ant its application to approximate reasoning. Information Science, 8, 199-249.