RESEARCH ON KNOWLEDGE-BASED PREDICTIVE NATURAL RESOURCE MAPPING SYSTEM

Fei Du[a, b,] *, A-Xing Zhu[a, c]

[a]State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, NO.11A, Datun Road, AnWai, Beijing100101, China - duf@lreis.ac.cn
[b]Graduate University of Chinese Academy of Sciences, NO. 19A, Yu Quan Road, Beijing 100049, China
[c]Department of Geography, University of Wisconsin Madison, 550 North Park Street, Madison, WI53706, USA - axing@lreis.ac.cn

KEY WORDS: Digital Mapping, Thematic Maps, Image Understanding, Spatial Modeling, Thematic Mapping, Inventory

ABSTRACT:

Knowledge-based predictive mapping is a kind of mapping approach which links the existing geo-spatial data and knowledge to predict the status of unknown geographical phenomena. The knowledge reflects the relationships between the known environmental conditions and the phenomenon to be predicted. In the field of natural resource inventory, such a mapping approach is very valuable. However, standard GIS Software can not effectively support knowledge management, which is crucial to knowledge-based predictive natural resource mapping, so new tools and methods should be developed. In this paper, we present our approach to manage knowledge and how the knowledge base is integrated with environmental database to produce natural resource maps. We first introduce basic concepts of knowledge-based natural resource mapping. Then, we investigate the needs and limitations of current systems. Our own approach that utilizes fuzzy logic, object-oriented method and hierarchical structure to represent knowledge and organize knowledge base is then put forward. This paper also presents our prototype knowledge-based predictive natural resource mapping system and its application in soil mapping which is conducted in a small watershed in Northeast China. Finally, the paper also discusses the existing problems and research challenges.

1. INTRODUCTION

Natural resource mapping is important to environmental protection and resource management. Though the survey and mapping technology progressed fast in the past few decades, there are still many problems. Firstly, producing natural resources maps still mainly relies on field investigation and human interpretation. Take soil mapping as an example, experts use their field experiences to delineate map unit polygons to produce soil maps, which is not only very labour-intensive but also very costly (Hudson, 1992). Secondly, for places that are hard to reach, such as mountainous areas, it is not easy or even impossible to conduct field survey. Thirdly, some spatial phenomena can not be investigated directly, for example, natural disasters need to be predicted before they really happen. New methods are imperative to solve these problems and to conduct natural resource mapping in a fast and intelligent way.

We can often find some relationships between different geographical phenomena. For example, ecological differences are largely controlled by changes in topography that produce gradients in moisture, energy and nutrients across the landscape (Moore et al., 1993). Soil type differences are controlled by factors such as terrain, climate, bed rock (McBratney et al., 2003; Bockheim et al., 2005). These concepts support the theoretical basis of predictive natural resource mapping: the spatial distribution of natural resource types can be predicted economically and accurately by apply relationships to existing data sets to predict the status of unknown phenomenon. Predictive natural resource mapping is an automatic mapping method with which predicts the natural resource type at given locations by assessing existing digital environmental data sets using a set of rules. The rules reflect relationship between the environmental data sets and the unknown natural resource types. This predictive mapping approach could solve many problems existing in current natural resource mapping. It can greatly reduce our workload. Besides, for it uses a predictive approach, we do not need to investigate every inch of land and besides, some phenomena can also be forecasted before they finally happen.

Predictive mapping has been studied for many years. In fact, many of existing GIS applications are in the form of predictive mapping, such as predictive ecosystem mapping (Jones et al., 1999), Fire Regimes mapping (Parsons, 2003) and soil survey (Zhu et al., 2001). With the recent development of photogrammetry, Remote Sensing and digital terrain analysis, large amount of digital environmental data sets can be accessed and data availability is not a very big issue any more, so the most important part in predictive mapping becomes the relationship between the existing data and the unknown phenomenon. There are mainly three ways to get this relationship: physical process-based approach, statistical approach and knowledge-based approach (Zhu, 2005). Physical process-based way needs a good understanding of geographic systems. Statistical method needs large number of samples. Knowledge-based approach is situated between them because it requires less understanding about geographic systems but can make use of existing geographical knowledge. We mainly focus on this approach in this paper.

* duf@lreis.ac.cn; phone 010-64889461

The knowledge reflects the relationship, so the quality of knowledge makes a great difference to the reliability of the mapping result. Currently, standard geographical information systems mainly provide some basic spatial data management and spatial analysis functionalities. They have no or very weak capability to mange knowledge, so they are hard to be used or directly used in predictive mapping. Some existing predictive mapping system has very simple method and limited number of forms to represent knowledge. At the same time, little research has been devoted to predictive natural resource mapping from a systematic view.

In this paper, we mainly discuss the architecture of a knowledge-based predictive natural resource mapping system with a focus on the knowledge management. Our method to represent knowledge and organize knowledge base is discussed in detail. We also developed a prototype software system and conducted a case study in soil mapping which is conducted in a small watershed in Northeast China.

2. SYSTEM OVERVIEW AND ARCHITECTRUE

Knowledge-based predictive natural resource mapping is based on the concept that there are relationships between the natural resource types to be mapped and a set of predictive variables (environmental data sets) as expressed in the following equation (Zhu et al., 2001):

$$S_{ij} = f(E_{ij}) \tag{1}$$

where S_{ij} is the natural resource type at location (i,j), E_{ij} is a set of predictive environmental conditions at the location, and *f* is the relationships between the property to be predicted and the set of predictive conditions for that location.

The required environmental conditions can be characterized using GIS and Remote Sensing techniques. The relationship between the feature to be mapped and the related environmental conditions is approximated with knowledge (Zhu, 1999). The knowledge and the characterized environmental conditions are then combined to predict the spatial distribution of the phenomenon through a set of inference techniques such as spatial overlay and raster map algebra. A knowledge-based predictive natural resource mapping system mainly contains three parts: knowledge base, environmental database and inference engine (Figure 1).

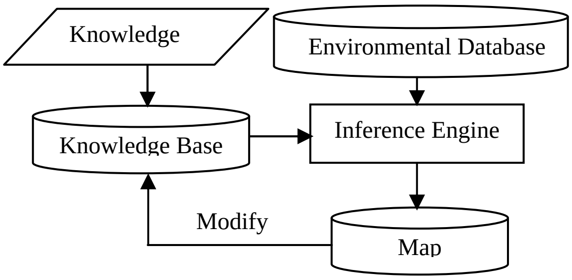

Figure 1. System Architecture

3. KNOWLEDGE MANAGEMENT AND INFERENCE

3.1 Different Types of Knowledge

There are mainly three ways to extract knowledge that reflects the relationship: extracting knowledge from existing documents, from human experts or from existing spatial database (Zhu, 2005). No matter what method is used, when using knowledge in computer environment, knowledge need to be encoded into a numeric form. It should be noted that knowledge is often expressed as a set of rules, so in this paper, rule is equivalent to knowledge.

In order to encode knowledge (rules), we should know what kinds of knowledge we may get. Traditionally, Boolean logic is used to describe knowledge. Under Boolean logic each rule is encoded as a step function. The function produces a value of 1 (or true) if the stated condition in the rule is met, 0 (false) otherwise.

Boolean rule is suitable for nominal data. For example, rock type. Still take soil as an example, if we know one kind of soil only appears on certain bed rock, e.g. Jordan Sandstone, we can express such knowledge in the following way:

If Bed Rock = “Jordan Sandstone” *f* (Bed Rock) = 1 (2)
Otherwise *f* (Bed Rock) = 0

Other types of data can be change to nominal data by pre-processing. Take elevation as an example. If we know that one kind of soil type occurs when the elevation is greater than 950 feet, we can use the following expression:

If Elevation > 950 feet *f* (Elevation) = 1 (3)
Otherwise *f* (Elevation) = 0

However, Boolean expression is not always an ideal way because of the fuzziness of geographical phenomena. Many geographical phenomena change gradually, that is we can not say one things belongs to set A, but not belongs to set B. Actually, it may have similarity with both typical A and B. In order to express such geographical knowledge, fuzzy logic is introduced (Zhu et al., 2001).

Under fuzzy logic, fuzzy membership curve is used to portray knowledge. Fuzzy membership indicates how much one thing is in a set .In a fuzzy membership curve, the optimal value is the value where fuzzy membership is 1. In another word, when the predictive condition is equal to this value, the natural resource type to be predicted is “pure”. The most frequently used curves are bell-shaped curve, Z-shaped curve, and S-shaped curve (Figure 2). Besides optimal value, the cross point where the fuzzy membership is equal to 0.5 should also be defined.

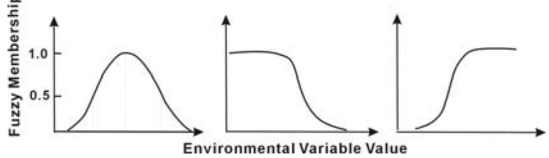

Figure 2. Bell-shaped, Z-shaped , and S-shaped curves

For bell-shaped membership curve, the optimal value can be a single point or can be a range. For one-sided curves, the optimal value is a point from which the membership decreases (Z-shaped) from or increases (S-shaped) to 1.0 as the attribute value increases.

Once the type of the curve and the optimal and crossover points on the curve are determined, the membership function can be formulated quite easily. There are many mathematical functions that can be used (Robinson, 2003). Burrough (1989) put forward the following function to depict a symmetrical bell-shaped membership function which is adopted by our system:

$$f(x) = \frac{1}{1 + (\frac{x-b}{d})^2} \quad (4)$$

where x is the value of a predictive variable, b is the attribute value representing the optimal value, and d is the width of the bell-shaped curve between one of the cross-over points and b. The S-shaped curve can be created by setting the membership value to 1 for x greater than b while the Z-shaped curve can be created by setting the membership value to 1 for x less than b.

For a fuzzy membership curve mentioned above, optimal value and cross point must be defined explicitly. In fact, many relationships can not be easily defined using this way.

A possible situation is that neither bell-shaped, Z-shaped nor S-shaped can effectively express the relationship. For example, there might be more than one optimal condition, so the fuzzy membership curve may have more than one peak. To support this kind of knowledge, we need more complex curves.

Sometimes, domain experts can only provide qualitative description, that is, they may use linguistic terms as values (such as curvature variables with “convex”, “linear”, “concave” terms). The determination of optimal and crossover points are difficult. Little attention is paid on how to express such kind of knowledge.

Another case is domain experts sometimes can not define a curve directly, that is they are not able to give optimal value, but they know where the typical location is (a typical location is the place where certain natural resource type is the most typical). The value of environmental variable in the typical location can be viewed as optimal value. By providing typical locations, experts describe their knowledge implicitly. This knowledge can be used in case-based reasoning (Shi et al, 2004).

3.2 Encoding Different Kinds of Knowledge

In order to meet various needs, we designed different forms to express different kinds of knowledge (rules). Five different types of rules are defined: enumerated rule, range rule, word rule, curve rule, and point rule, each one is corresponding to one typical need.

Enumerated rules are used to express Boolean knowledge. All the possible conditions when certain natural resource type occurs are listed. If the actual environmental condition matches any item in this list, the fuzzy membership value yielded by this rule will be 1; otherwise, the value will be 0.

Range rules are designed to express commonly-used membership curves (bell-shaped, Z-Shaped and S-Shaped curves). The optimal value and cross point of the curve need to be defined.

Curve rules, word rules and point rules are designed to supplement traditional fuzzy membership curves. For Curve rules, some key points need be defined, we adopted spine curve technique to interpolate the points into a curve. Word rule contains a linguistic term and its corresponding word library. The term is a descriptive word, such as “convex”, “linear”, “concave” etc., while the word library is an index, where every linguistic item has a curve associated with it. So when choose a linguistic term, we can get a fuzzy membership curve. Point rule records the x-y coordinates (spatial location) of the typical point, and the cross point of the fuzzy membership curve. It is very much like the range rule except its optimal value is determined by the x-y coordinates rather than the direct input.

3.3 Knowledge Organization and Management

A rule is a basic unit of a knowledge base. One rule defines how one environmental variable (e.g. elevation) determines the fuzzy membership value to certain natural resource type. A set of rules make up an instance which is a situation when a certain natural resource type occurs.

Besides instance, we introduce another concept: exception. An instance takes effect globally, while exceptions are local contributing, which means that exception are very like instances except that they only affect certain area. So exceptions not only contain rules, but they also have spatial information associated with them, e.g. the central point and impact radius. Besides, we need to define how the effect will decrease when the location derivate from the central point.

A knowledge base contains knowledge about the occurrence of many natural resource types. One natural resource type consists of instances, exceptions. One instance or exception is comprised of many rules. The rule can be any type that was mentioned in section 3.2. As we can see, the structure of a knowledge base in such a predictive natural resource mapping system is hierarchical (Figure 3).

Knowledge representation deals with how to store the knowledge into computer. There are representation techniques such as frames, rules and semantic networks which have originated from theories of human information processing (Guo and Ren, 2003). Since knowledge is used to achieve intelligent behaviour, the fundamental goal of knowledge representation is to represent knowledge in a manner as to facilitate inference from knowledge.

A newly-developed method to represent knowledge is object-oriented approach which learns from the object-oriented method in computer programming language. The key features of object-oriented way to represent knowledge are encapsulation, inheritance and polymorphism. We find it’s convenient to organize knowledge in knowledge-based predictive mapping using object-oriented approach.

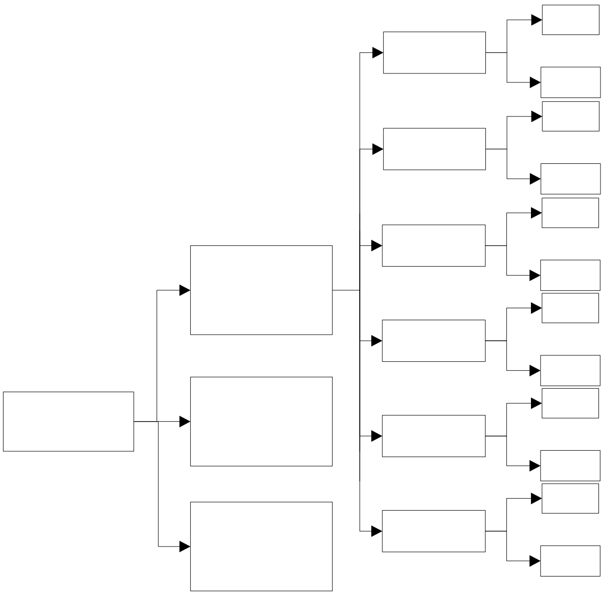

Figure 3. Hierarchical structure of knowledge base

Different kinds of rules have some common attributes and methods. For example, every rule must have a name and a tag to mark whether or not to be used in inference engine. They also have some common methods, such as the method *Evaluate* which compute the inference result. We can define a base class to be the parent class of different kinds of rule classes.

The method *Evaluate* must be implemented by all subclasses, say all kinds of rules, so it is designed as a virtual function in parent class. Besides the attributes and methods inherited from the base class, different rule classes have their own attributes and methods. Take “curve rule” class as an example, it contains a series of x-y coordinates of points that define the curve. Its *Evaluate* method is also different from other kinds of rules. With object-oriented approach, every rule is responsible for its own validation, inference and other behaviors, so rules have kind of “intelligence”.

We use Extensible Markup Language (XML) to store knowledge in hard disk. XML is a general-purpose specification for creating custom markup languages. It is classified as an extensible language because it allows its users to define their own elements. The primary purpose of developing XML is to facilitate the sharing of structured data across different information systems (Bray et al., 2006).

Using XML to store knowledge has many advantages, which has been discussed by much literature(De Vries, 2004; Zhou et al, 2005). For predictive mapping system, as stated in the previous section, the structure of knowledge base is hierarchical, and XML file is also hierarchical, so it’s rather simple to store knowledge in a XML file. XML is designed to be self-descriptive, and we can define our own tags, so it’s very easy to integrated different types of knowledge into one file. At the same time, it is convenient to add new types of knowledge if we want to do so in the future. This is important for natural resource mapping because we may have various kinds of knowledge from different sources, so we should keep the knowledge base is extensible. Thirdly, XML uses text to store information, so is easy to understand and distribute.

3.4 Inference

Once the knowledge is encoded and stored in the knowledge base and the environmental database is prepared, we can do inference. Inference is done by inference engine which is a set of programs. The work of inference engine is to integrating knowledge to produce the final predictive result.

Since we adopt an object-oriented approach to organize knowledge, it’s rather easy to do inference. Every instance and exclusion is responsible for computing a fuzzy membership value by integrating the set of rule under it. Then every natural resource type is responsible for computing a fuzzy membership to its own by integrating different instances and exceptions under it. These two steps are all carried out pixel by pixel. There are a number of operators that can be used to accomplish the integration: The commonly used implementation is minimum operator, and maximum operator (Zhu, 2005).

If there are totally N natural resource types in one study area, after those two steps, N fuzzy membership files should be created, each of which describes to what extent every location (pixel) belongs to one natural resource type. So for every location, there are N fuzzy membership values. Then the ID of the natural resource type which has the maximum fuzzy membership value at each location will be assigned to that location. Of course, the ID should be unique for every natural resource type. If all the above operations are done, every location gets one natural resource type (one ID), and the inference is finished.

4. SYSTEM IMPLEMENTAION

We used Visual C++ 7.0 (in Visual Studio.NET 2003) to build a prototype software system.

The prototype system has the capability to prepare data, mange raster database and facilitate natural resource experts to explore the data conveniently. The source data of natural resource mapping mainly come from remote sensing, digital terrain analysis. In the software, we provide a toolbox that can derive new data sets from existing data by digital terrain analysis and remote sensing data analysis. Using the system, one can construct and modify raster database conveniently. Interactive visualizing and exploring raster data in both graphical and numerical mode are also provided, which is proved to be an effective way to make users better understand the data (Figure 4).

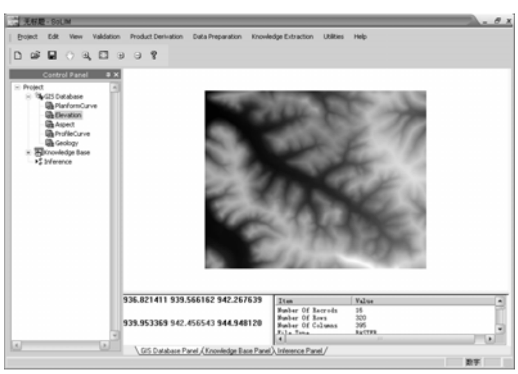

Figure 4. Interface to manage and explore database

Domain experts in the filed of natural resource inventory are usually not GIS or cartography experts, so a good human-computer interaction interface is necessary in predictive resource mapping system. The prototype system provides easy-to-use, intuitive tools to acquire knowledge interactively. One can add, delete, and rename rules, instances, exceptions, or natural resource types. Fuzzy membership curve can be defined by using graphical tools (Figure 5).

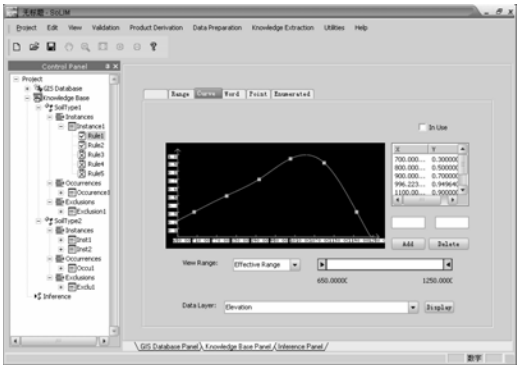

Figure 5. Interface to manage knowledge base

Once the environmental database and knowledge base are all prepared with the software, one can do inference to get the final predictive map by running inference engine. Our prototype software also provides many other utilities which can facilitate the mapping process.

5. CASE STUDY

The prototype knowledge-based predictive natural resource mapping system was used to conduct a test research in the field of soil mapping in a watershed located in Heshan farm of Nenjiang county Heilongjiang province in Northeast of China.

Firstly, five environmental data layers were chosen to setup the environmental database. They are: gradient layer, planform curvature layer, profile curvature layer, relative terrain position layer, and wetness layer. They are thought to be the predictive variables that determine the formation of soil types in that area. Then soil experts used the prototype system to input their knowledge interactively. According to soil experts, there are totally 6 soil subgroups (types) in this area. Expert used the system to define instances for each type (there are no exception).Finally, the inference engine was run to get the final soil map. Figure 6 shows the fuzzy membership maps of two soil types. Different grey scales represent different fuzzy membership values. The lighter the grey scale, the higher the membership value.

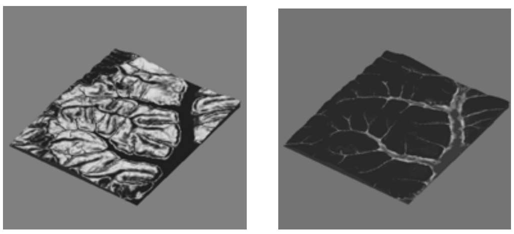

Figure 6. Fuzzy membership maps of two soil types

The final hardened soil map is shown in Figure 7. Different grey scales represent different types of soils.

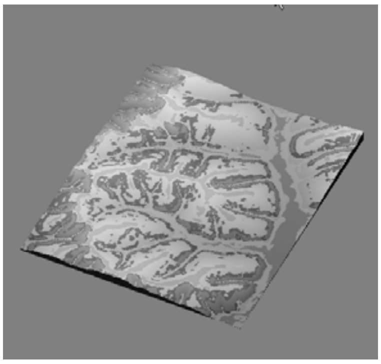

Figure 7. Soil map produced by our prototype system

Comparison was made with the existing soil map, which shows that the map produced with our system shows more spatial details. The evaluation from soil experts prove that the prototype software system has considerable appeal for it can effectively assist natural resource experts in natural resource mapping.

6. CONCLUSIONS AND DISCUSSIONS

This paper mainly discusses knowledge-based natural resource mapping from a systematic view. Our method that utilizes fuzzy logic, object-oriented method and hierarchical structure to represent knowledge and organize knowledge base is presented. The prototype software, which is an implementation of our thought, proves to be very useful and efficient in practical natural resource mapping.

Nevertheless, knowledge-based predictive natural resource mapping still faces many challenges and has many research issues. We use fuzzy membership curve to represent knowledge, to what extent the fuzzy membership can reflect the true situation needs to be carefully researched. Geographical knowledge may exist in many forms, so how to integrate knowledge from different sources is also an important research issue.

From technical point of view, current GIS software provides powerful data management and spatial analysis functionalities, but lacks capability to mange knowledge base. How to add knowledge management functionality to existing GIS software to support knowledge-based natural resource mapping needs to be further examined.

REFERENCES

Bockheim, J.G., Gennadiyev, A.N., Hammer, R.D. and Tandarich J.P., 2005. Historical development of key concepts in pedology. Geoderma 124 (1-2), pp. 23-36.

Bray, T., Paoli, J., Sperberg-McQueen, C.M., Maler, E., Yergeau, F., 2006. Extensible Markup Language (XML) 1.0. Fourth Edition-Origin and Goals. World Wide Web Consortium. http://www.w3.org/TR/2006/REC-xml-20060816/#sec-origin-goals.

Burrough, P.A., 1989. Fuzzy mathematic methods for soil survey and land evaluation. Journal of Soil Science, 40, pp. 447-92.

De Vries, A., 2004. XML framework for concept description and knowledge representation. eprint arXiv:cs/0404030. http://arxiv.org/abs/cs/0404030.

Guo, Q. and Ren, X., 2003. Intelligent Processing Technology of Geographical Information（In Chinese）. Wuhan University Press, Wuhan, pp. 33-70.

Hudson, B.D., 1992. The soil survey as paradigm-based science. Soil Science Society of America Journal, 56(3), pp. 836-841.

Jones, K., Meidinger, D. Clark D. and Schultz F, 1999. Towards the establishment of predictive ecosystem mapping standards: A white paper, 1st approximation, Prepared for Terrestrial Ecosystem Mapping Alternatives Task Force, Resource Inventory Committee (RIC).

McBratney, A.B., Mendoca Santos M.L. and Minasny, B., On digital soil mapping, Geoderma, 117(1-2), pp. 3-52.

Moore, I.D., Norton T.W. and Williams J.E., 1993. Modelling environmental heterogeneity in forested landscapes. Journal of Hydrology, 150(2/4), pp. 717–747.

Robinson, V.B., 2003. A perspective on the fundamentals of fuzzy sets and their use in geographic information systems. Transactions in GIS 7(1), pp. 3-30.

Shi, X., Zhu A.X., Burt, J., Qi, F. and Simonson, D., 2004. A case-base reasoning approach to fuzzy soil mapping. Soil Science Society of American Journal, 68, pp. 885-894

Zhou, K., Jin B., Zhou J., Feng, S., 2005. An object oriented case knowledge representation using XML. Journal of Wuhan University of Technology (Information & Management Engineering), 29(3), pp. 86-89.

Zhu, A.X., 1999. A personal construct-based knowledge acquisition process for natural resource mapping. International Journal of Geographical Information Science 13(2), pp. 119-141.

Zhu, A. X., Hudson, B., Burt, J., Lubich, K., and Simonson, D., 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of America Journal, 65, pp. 1463-72.

Zhu, A.X., 2005. Rule-based mapping. In: Wilson, J.P. and Fotheringham, A.S. eds. Handbook of GIS, pp. 273-293.

ACKNOWLEDGEMENTS

The author is grateful to the support from the “One-Hundred Talents” Program of Chinese Academy of Sciences, National Basic Research Program of China（2007CB407207）and Chinese Academy of Sciences International Partnership Project "Human Activities and Ecosystem Changes"(CXTD-Z2005-1）.The assistance provided by Yanjun Lu in system implementation is greatly appreciated.