Knowledge-Based Systems 57 (2014) 28–40

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A method for extracting rules from spatial data based on rough fuzzy sets

Hexiang Bai [a,b,c], Yong Ge [b,*], Jinfeng Wang [b], Deyu Li [a,c], Yilan Liao [d], Xiaoying Zheng [d]

[a] School of Computer and Information Technology, Shanxi University, Taiyuan 030006, Shanxi, China

[b] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

[c] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, China

[d] Institute of Population Science, Peking University, Beijing, China

ARTICLE INFO

Article history:
Received 2 August 2013
Received in revised form 1 November 2013
Accepted 6 December 2013
Available online 12 December 2013

Keywords:
Rough fuzzy set
Rule extraction
Spatial analysis
Attribute reduct
Fuzzy decision rule

ABSTRACT

With the development of data mining and soft computing techniques, it becomes possible to automatically mine knowledge from spatial data. Spatial rule extraction from spatial data with uncertainty is an important issue in spatial data mining. Rough set theory is an effective tool for rule extraction from data with roughness. In our previous studies, Rough set method has been successfully used in the analysis of social and environmental causes of neural tube birth defects. However, both roughness and fuzziness may co-exist in spatial data because of the complexity of the object and the subjective limitation of human knowledge. The situation of fuzzy decisions, which is often encountered in spatial data, is beyond the capability of classical rough set theory. This paper presents a model based on rough fuzzy sets to extract spatial fuzzy decision rules from spatial data that simultaneously have two types of uncertainties, roughness and fuzziness. Fuzzy entropy and fuzzy cross entropy are used to measure accuracies of the fuzzy decisions on unseen objects using the rules extracted. An example of neural tube birth defects is given in this paper. The identification result from rough fuzzy sets based model was compared with those from two classical rule extraction methods and three commonly used fuzzy set based rule extraction models. The comparison results support that the rule extraction model established is effective in dealing with spatial data which have roughness and fuzziness simultaneously.

1. Introduction

Spatial data mining is the process of discovering the interesting and previously unknown, but potentially useful, patterns and rules (association and classification rules) from spatial datasets [39,50]. Over the last two decades, spatial data mining has been widely used in many applications, such as categorizing and localizing the human action(s) contained in a video [30], evaluating the structural and topological consistency among multiple representations of complex regions with broad boundaries [10], mining the frequent trajectory patterns in a spatial–temporal database [21], extracting the spatial association rules from a remotely sensed image database [22], generating the appropriate polygons from heterogeneous spatial information [38], and analyzing the change of land use [9].

Extraction of spatial decision rules is one of the main targets of spatial data mining [39,40,50] and has been used in many real applications. Ester et al. [13] used ID3 to extract decision rules from spatial databases via the so-called generalized attributes which take the neighborhood relation into consideration through a predefined position path of neighbors. Pontius et al. [35] proposed a model to select the locations of land-use change by the decision rules generated using the nearest neighbors. Daniels [8] introduced domain knowledge base which consists of decision rules to help classification of land cover types of remotely sensed imagery. Frank et al. [15] took the spatial and aggregation literals, such as perimeter, location and area, of spatial objects into account when mining rules from spatial data via a Voronoi-based approach to take non-spatial features into the rule extraction process. Zhu and Hu [54] extracted rules by using support vector machines, which is originally difficult to explain to users or to be understand by users as a black box, via analyzing the consistent regions formed by samples in terms of classification boundary.

In the studies mentioned above, both condition attributes and decision attributes, which are used to describe the objects in a data table, are crisp. The extracted rules are also crisp, i.e., the antecedent and consequent of a rule are expressed in some accurate way. However, because of the complexity of the object world, the subjective limitation of human knowledge and the uncertainty

* Corresponding author. Tel./fax: +86 10 64888053.
E-mail address: gey@lreis.ac.cn (Y. Ge).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.12.008

intrinsic of spatial data, there exists fuzziness in the representations of geographical phenomena [40] such as the classification of transition zones of land cover types, and the detection of the influence of environmental factors on the incidence rate of birth defects [4]. If the traditional models are used in extracting rules from spatial data with fuzzy descriptions by nature, the fuzzy concepts or fuzzy decisions should be degraded into crisp ones. For example, the transition zones between grassland and forest should be assigned to one specific category. The degradation may lead to information loss and decreases prediction accuracy. Meanwhile, the rules extracted using classical models will have crisp decision. These rules are less explainable than the fuzzy decision rules [12]. Some researchers then used fuzzy set theory to handle this issue. For example, Hu et al. [17] proposed the fuzzy-grid-based rule mining algorithm to generate fuzzy association rules. Niu et al. [31] used fuzzy concept lattice to mine the spatial association rules. In fact, there also exists another kind of uncertainty, i.e., roughness, in the geographical data tables.

In general, roughness is interpreted as the uncertainty of a concept while the concept cannot be precisely expressed with other concepts. In data analysis field, roughness is firstly studied by Pawlak in 1982 [34], where a concept is defined as an object subset of the universe with some property. In rough set theory, two operators, upper and lower approximations, are designed to roughly depict a target concept. A concept is called to be rough when its boundary region, i.e., the difference between the upper and lower approximations, is nonempty. The roughness of a concept derives from its boundary region. Especially, in geographical phenomena, roughness means that a target concept cannot be precisely described by the available information granules formed by spatial objects' features. For example, the spectral information of the remotely sensed imagery may not precisely describe a landcover type in the study area. As the roughness concept can be precisely defined when more additional features are collected, it is not a fuzzy concept by nature. Accordingly, it is inappropriate to use fuzzy sets to handling roughness. For example, rock and soil are hard to be discerned using existing band in remotely sensed imagery. But they may be clearly distinguished using other unavailable bands. Rock and soil do not have ill definition of boundary in such situation. Accordingly, fuzzy sets based model is not suitable for such situation. This means that the existing methods for rule extraction need to be extended for well working in spatial data analysis.

Rough set theory [34] is an effective tool in dealing with roughness and it can be used in extracting decision rules in spatial data with roughness [6,7]. However, the classical rough set theory is only suitable to the cases that objects are described by the nominal type of condition attributes and the crisp type of decision attributes in a data table. Nonetheless, in many real applications, the decision value of an object is fuzzy. Taking Heshun Neural Tube Birth Defects (NTD) data as an example, Bai et al. [3] used rough set theory to extract spatial decision rules from Heshun NTD data. However, the decision attribute in that work is "whether there are NTD instances in a village", i.e., it is a Yes–No type decision attribute, which cannot reflect the severity degree of NTD for each village. For example, a village within ten instances suffers more than that within only one instance. It is obvious that treating NTD birth defects as a fuzzy concept on the domain of discourse of all villages in Heshun is a better way than using a Yes–No type decision attribute. It can more intuitively describe the severity degree of NTD birth defects in villages. By the discussion above there are two kinds of uncertainties, i.e., roughness and fuzziness, needed to be handled in this case. Two extensions of rough set theory provide tools for handling data of this type.

Rough fuzzy sets and fuzzy rough sets [11] were proposed to extend the classical rough set theory [34] which allows the existence of fuzziness in decision attributes. They combine the advantages of rough sets and fuzzy sets [52]. The difference between fuzzy rough sets and rough fuzzy sets is that fuzzy rough sets is designed for the cases that both conditional attribute and decision attribute are fuzzy sets of the universe while rough fuzzy sets are specialized in dealing with data tables with crisp conditions and fuzzy decisions. Although rough fuzzy sets are special cases of fuzzy rough sets, the modeling process of fuzzy rough sets needs the fuzzification of conditional attributes. The fuzzification process involves the selection of the fuzzification methods, which will increase the modeling complexity and introduce new source of uncertainty. Therefore, the fuzzy rough sets based model cannot completely replace rough fuzzy sets based model.

Many researches on modeling spatial data by using rough set theory have been reported such as the classification of remotely sensed imagery, modeling spatial topology between spatial objects, uncertainty analysis and rule extraction [2,3,22,33,45,51], while rough fuzzy sets and fuzzy rough sets attract little attention in spatial data mining. Ahlqvist et al. [1] defined rough fuzzy classification and proposed various kinds of accuracy measures on rough fuzzy classification. This model assesses the classification accuracy using some goodness functions that reflect the features of the classification result from different perspectives. However, it did not give the accuracy assessment of prediction result in terms of reference. Furthermore, little research seems to address the detailed process of the reduct which is an essential procedure in spatial data mining using rough fuzzy sets.

In this paper, we focus on the problem of rule extraction from spatial data with crisp condition attributes and fuzzy decisions. A rough-fuzzy set based rule extraction model is used to deal with both fuzziness and roughness in spatial data tables. Unlike other commonly used spatial rule extraction methods, this model can simultaneously consider roughness and fuzziness in data. This model firstly converts the spatial data into a fuzzy decision information system. Rough fuzzy sets are then used to find a reduction of the fuzzy decision information system. Next, some fuzzy decision rules are extracted from the reduced fuzzy decision information system. Using the extracted rules, unseen objects can be classified, and the classification result is assessed by using fuzzy entropy and fuzzy cross entropy. This model is used in the analysis of Heshun NTD data, which is a very critical issue in China and has been studied for years by the authors. The rough fuzzy classification results of NTD data is compared with the results based on two kinds of classical rule extraction methods and three fuzzy decision rule extraction methods. The experimental results show that rough fuzzy set is an appropriate model for spatial analysis with both roughness and fuzziness in data.

The present paper has the following organization. Section 2 outlines the concepts of fuzzy decision information system and rough fuzzy sets. Section 3 proposes a model based on rough fuzzy sets to extract spatial rules from fuzzy decision information system constructed from spatial data. An example of NTD data from Heshun, Shanxi, China is given in Section 4. Section 5 makes detailed discussion on the effectiveness of the new method via entropy based accuracy assessment and performs a comparison of the proposed model with other five commonly used rule extraction methods. The last section concludes this paper.

2. Fuzzy decision information systems and rough fuzzy sets

A decision information system is defined as a pair $(U, AT \cup D)$, where the universe U is a non-empty finite set of objects, AT is a non-empty finite condition attribute set which contains m_c elements, D is a non-empty finite decision attribute set which contains m_d elements ($m = m_c + m_d$), and $AT \cap D$ is empty. Any $a \in AT$

($d \in D$) can be regarded as a mapping from $U \times \{a\}$ ($U \times \{d\}$) to V_a (V_d), where V_a (V_d) is the domain of the attribute a (d).

The equivalent class of $x \in U$ formed by an attribute set $R \subseteq AT$ ($R \subseteq D$) is defined as the set $[x]_R = \{y \in U: a(x) = a(y), \forall a \in R\}$. Obviously, if two objects x and y belong to the same equivalent class $[x]_R$, then $a(x) = a(y)$ for any attribute a in R. So we can use $a([x]_R)$ to represent the value of any object in $[x]_R$ under an attribute $a \in R$.

In the classical rough set theory [34], a target concept $X \subseteq U$ can be approximately depicted by two approximation operators that are associated with a condition attribute set $R \subseteq AT$, i.e., the upper approximation $\overline{P}_R(X) = \{x \in U : [x]_R \cap X \neq \Phi\}$ and the lower approximation $\underline{P}_R(X) = \{x \in U: [x]_R \subseteq X\}$. The boundary of X is defined as $BND(X) = \overline{P}_R(X) - \underline{P}_R(X)$. A set is said to be rough if the boundary region $BND(X)$ is non-empty.

A fuzzy set Z on U can be identified by a membership function $\mu_Z: U \to [0, 1]$. The intersection and union of two fuzzy sets X and Y are defined as $\mu_{X \cap Y}(x) = \min\{\mu_X(x), \mu_Y(x)\}$ and $\mu_{X \cup Y}(x) = \max\{\mu_X(x), \mu_Y(x)\}$ respectively. The cardinality of a fuzzy set X is defined as $|X| = \sum_{x \in U}\mu_X(x)$. A crisp set $X \subseteq U$ can be looked upon as a special case of fuzzy sets. The membership function for the crisp set X can be defined as

$$\mu_X(x \in U) = \begin{cases} 1, & x \in X \\ 0, & x \notin X \end{cases}$$

Let $(U, AT \cup D)$ be a decision information system. If $\forall d \in D$ is a fuzzy set on U, $(U, AT \cup D)$ is known as a fuzzy decision information system (FDIS), and $\forall d \in D$ can be considered as a mapping from $U \times \{d\}$ to $[0, 1]$. The membership function of a fuzzy decision d, which is identified as $\mu_d: U \to [0, 1]$, is assigned by users or transformed through some fuzzification methods. An illustrative example of FDIS is shown in Table 1. It has three crisp condition attributes $A1$, $A2$ and $A3$, and two fuzzy decisions D1 and D2. The three equivalent classes formed by $\{A1, A2, A3\}$ are $C_1 = \{1, 2, 3\}$, $C_2 = \{4, 5\}$ and $C_3 = \{6, 7\}$ respectively.

Based on rough sets and fuzzy sets, Dubois and Prade [11] proposed rough fuzzy set theory in dealing with FDIS. Let $(U, AT \cup D)$ be an FDIS, $R \subseteq AT$ and $F \in D$. In rough fuzzy set theory, two operators called as the upper approximation $\overline{R}(F)$ and lower approximations $\underline{R}(F)$ are designed to approximately represent the fuzzy concept F. $\overline{R}(F)$ and $\underline{R}(F)$ are two fuzzy sets on U with membership functions defined by Formulas (1) and (2). For $\forall x \in U$,

$$\mu_{\overline{R}(F)}(x) = \max\{\mu_F(y), y \in [x]_R\} \tag{1}$$

$$\mu_{\underline{R}(F)}(x) = \min\{\mu_F(y), y \in [x]_R\} \tag{2}$$

It is easy to see that for $\forall x \in U$, $\mu_{\underline{R}(F)}(x) \leqslant \mu_F(x) \leqslant \mu_{\overline{R}(F)}(x)$, i.e., the membership of F lays between the memberships of $\underline{R}(F)$ and $\overline{R}(F)$. The pair $(\overline{R}(F), \underline{R}(F))$ is called a rough fuzzy set [11].

For example, when $R = \{A1, A2, A3\}$, the upper and lower approximations of fuzzy decisions D_1 and D_2 in Table 1 are

$$\overline{R}(D_1) = \{(1, 0.23), (2, 0.23), (3, 0.23), (4, 0.5), (5, 0.5), (6, 0.9), (7, 0.9)\}$$

$$\underline{R}(D_1) = \{(1, 0.1), (2, 0.1), (3, 0.1), (4, 0.3), (5, 0.3), (6, 0.8), (7, 0.8)\}$$

Table 1
An example of fuzzy decision information system.

ID	A1	A2	A3	D1	D2
1	x	Low	High	0.1	0.8
2	x	Low	High	0.1	0.3
3	x	Low	High	0.23	0
4	z	High	Low	0.5	0.9
5	z	High	Low	0.3	0.9
6	y	High	Low	0.8	0.2
7	y	High	Low	0.9	0.1

$$\overline{R}(D_2) = \{(1, 0.8), (2, 0.8), (3, 0.8), (4, 0.9), (5, 0.9), (6, 0.2), (7, 0.2)\}$$

$$\underline{R}(D_2) = \{(1, 0), (2, 0), (3, 0), (4, 0.9), (5, 0.9), (6, 0.1), (7, 0.1)\}$$

Meanwhile, a rough membership function for x for a fuzzy decision $d \in D$ is defined as $l_d(x) = |[x]_R \cap d|/|[x]_R|$. For example, $l_d(1) = |[1]_R \cap D_1|/|[1]_R| = (0.1 + 0.1 + 0.23)/3 \approx 0.143$ for the first object in Table 1.

2.1. Attribute reduction in an FDIS

In an FDIS, attribute reduction is one of the main problems raised in the applications of rough fuzzy sets. The two main motivations of attribute reduction are the removal of redundant attributes and simplification of decision rules. Xu et al. [49] introduced five types of reduction for knowledge discovering in FDISs. The five reduction methods are induced based on the so-called discernibility matrix introduced in [41]. One of them defined as follows is adopted in this paper for it can preserve both decision type and its precision while other four reducts cannot [49].

Let $FI = (U, AT \cup D)$ be an FDIS, and $A \subseteq AT$. For $\forall x \in U$, we define

$$E_A(x) = \left\{(d_i, \alpha) : d_i = \arg \max_{d_k \in D, k \leqslant m_d} \mu_{\underline{A}(d_k)}(x), \quad \alpha = \max_{d_k \in D, k \leqslant m_d} \mu_{\underline{A}(d_k)}(x)\right\} \tag{3}$$

By the definition of the lower approximation given by Formula (2), we have that for $\forall y \in [x]_A$, $E_A(x) = E_A(y)$.

Let $FI = (U, AT \cup D)$ be an FDIS. An attribute subset $A \subseteq AT$ is referred to as an x_E consistent set of FI, if $E_A(x) = E_{AT}(x)$ for $\forall x \in U$. If $A \subseteq AT$ is an x_E consistent set of FI, and no proper subset of A is also an x_E consistent set, then A is referred to as an x_E reduct of FI.

By Formula (3), an important property of an x_E reduct of FI is that it can preserve the largest one among the lower approximation membership values of all decision attributes for every object x in U [49]. Therefore, the decision type of each object given by the x_E reduct is retained, i.e., the decision which has the maximum lower approximation membership value and the membership values of objects under the decision are also unchanged. For example, in the FDIS shown in Table 1,

$$E_{AT}(1) = E_{AT}(2) = E_{AT}(3) = (D_1, 0.1), \quad E_{AT}(4) = E_{AT}(5) = (D_2, 0.9), \quad E_{AT}(6) = E_{AT}(7) = (D_1, 0.8)$$

$$E_{\{A1\}}(1) = E_{\{A1\}}(2) = E_{\{A1\}}(3) = (D_1, 0.1), \quad E_{\{A1\}}(4) = E_{\{A1\}}(5) = (D_2, 0.9), \quad E_{\{A1\}}(6) = E_{\{A1\}}(7) = (D_1, 0.8)$$

It can be seen that $E_{\{A1\}}(x) = E_{AT}(x)$ for every object $x \in U$. Therefore, $\{A1\}$ is a reduct of the FDIS.

Let $U/AT = \{C_1, C_2, \ldots, C_m\}$ be the partition of U formed by AT. The corresponding discernibility matrix can be defined as:

$$D(C_i, C_j) = \begin{cases} \{a_k \in AT : a_k(C_i) \neq a_k(C_j)\}, & E_{AT}(C_i) \neq E_{AT}(C_j) \\ \Phi, & E_{AT}(C_i) = E_{AT}(C_j) \end{cases} \tag{4}$$

According to the discernibility matrix $D(C_i, C_j)$, for example, the discernibility matrix for Table 1 is

	C_1	C_2	C_3
C_1	Φ	$\{A1, A2\}$	$\{A1, A2, A3\}$
C_2	$\{A1, A2\}$	Φ	$\{A1\}$
C_3	$\{A1, A2, A3\}$	$\{A1\}$	Φ

By Formula (4), each non-empty entry in the discernibility matrix is a condition attribute set and any arbitrary attribute in the set can distinguish all pairs of objects from C_i and C_j.

Based on the discernibility matrix, the discernibility function can be defined as

$$\Delta(\mathrm{AT}) = \wedge(\vee D(C_i, C_j))$$

The set of all prime implicants of $\Delta(\mathrm{AT})$ determines the set of all reducts of the FDIS [20].

2.2. Decision rules in FDISs

Once a reduct has been obtained, the fuzzy decision rules are easily constructed through overlaying the reduct on the original decision table and reading off the values. And further, these rules can be used to identify unseen objects. To facilitate the understanding of the synthesis process of decision rules, some related concepts are recalled here.

The decision rules obtained from an FDIS are of the form $A \Rightarrow B$, where A is a crisp set on U with the form $\wedge_{a \in att(A)}(a(x) = *)$ and B is a fuzzy membership value with the form $\vee_{d \in att(B)} \mu_d(x)$, where $att(A)$ consists of the conditional attributes appeared in A, * is a value in the domain of a, and $att(B)$ consists of the decision attributes appeared in B. The support and confidence of a rule are defined as $supp(A \Rightarrow d) = \sum_{x \in U} \mu_{A \cap d}(x)$ and $conf(A \Rightarrow d) = \sum_{x \in U} \mu_{A \cap d}(x)/|A|$ respectively [37].

3. Modeling spatial data using rough fuzzy sets

The rule extraction model based on rough fuzzy sets can be divided into the following four steps (see Fig. 1): (1) Construct an FDIS from the spatial data; (2) find one minimal reduct of the FDIS; (3) generate rules according to the obtained reduct; and (4) apply the rules to unseen cases and perform error analysis. This process is similar to generating decision rules based on the classical rough set theory. However, the classical algorithms and techniques should be improved to meet the requirements of FDISs because the decision attributes in an FDIS are fuzzy concepts.

In the first step, the spatial data, which is commonly represented via a map, should be converted to an FDIS. Each layer of the spatial data should be treated as an attribute of surface objects, such as the land cover type, amount of rainfall. The values of fuzzy decisions of sample data in an FDIS should be assigned by domain experts or be converted from data using some sort of fuzzification methods. Meanwhile, the continuous condition attributes in the FDIS should be discretized. No matter which discretization method is used, the cuts of the discretization of each attribute need to be preserved for later use.

When an FDIS is prepared, the next step is finding a minimal reduct of the FDIS, which will reduce the complexity of the problem at hand through selecting most relevant conditional attributes. The x_E reduct is used in this paper.

Next, all of the fuzzy rules that are hidden in the original system can be extracted through reading off the condition and decision attributes with their values from the reduced FDIS. These fuzzy decision rules extracted can be applied to fuzzy classification of unseen objects through the way of rule matching. The continuous attributes of unseen objects should also be discretized using cuts saved in the first step before rule matching. Standard voting [32] is used to perform the rule matching process. In the voting process, the $votes(r)$ of a rule r [32] is calculated through the support of a rule.

After rules are applied to unseen objects, accuracy assessment of the identified decisions of unseen objects is indispensable. The confusion matrix is commonly used to perform this task. However, the traditional confusion matrix can only be used for hard classification, i.e., each object is associated with only one class. Clearly, such an approach cannot provide a useful evaluation method to a fuzzy classification task [18]. In this paper, the entropy-based methods [14] are used to perform accuracy assessment due to that it cannot only examine the degree of uncertainty of the decision values but also assess the agreement between the identified fuzzy decisions and the real fuzzy decisions of reference data.

Foody [14] uses entropy and cross entropy to assess the accuracy of fuzzy classification. The entropy of a fuzzy set can be defined as:

$$H = -\sum_{s=1}^{n}\sum_{k=1}^{c}\mu_k(x_s)\log_2(\mu_k(x_s)) \tag{5}$$

where $\mu_k(x_s)$ is the fuzzy degree of x_s corresponding to class k; $x_s, s \in \{1, 2, \ldots, n\}$ represents the individual objects; n is the total

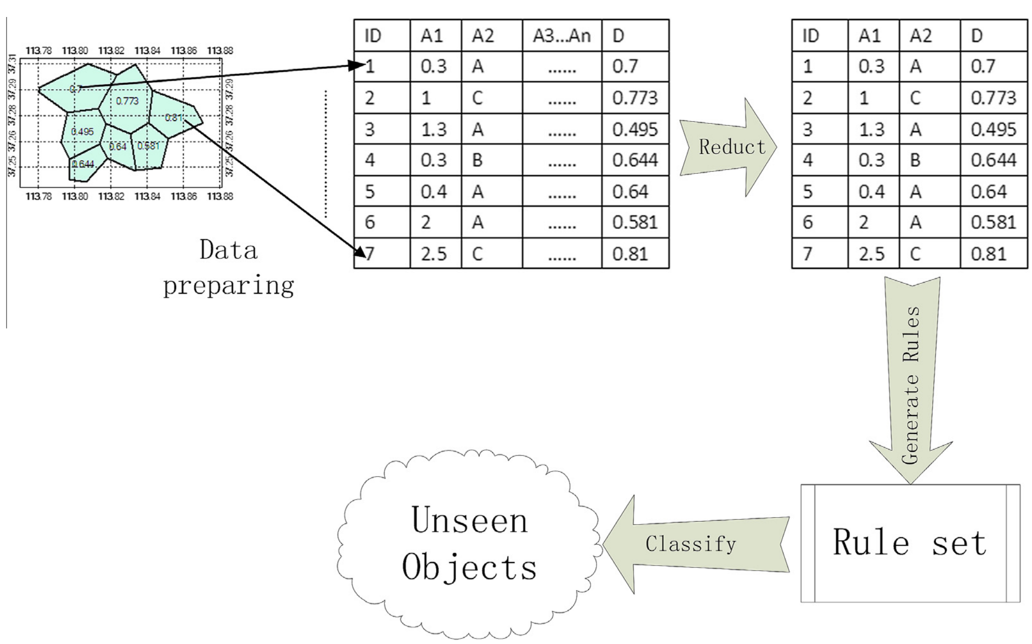

Fig. 1. Illustration of the rough fuzzy set based rule extraction process.

number of objects; $k \in \{1, 2, \ldots, c\}$ denotes the individual categories; c is the total number of categories. Entropy was used to examine the degree of fuzziness in fuzzy classified outputs [53] and to represent the way in which the class memberships (i.e. fuzzy memberships) were partitioned between the classes. A high value of entropy indicates that an object belongs to many classes (and thus is fuzzy) whereas a low value indicates that the object has a high membership of only one class (and thus is crisp) [14]. Cross-entropy can be used to assess the accuracy of fuzzy classification outputs with respect to fuzzy reference data [14]:

$$H_C = -\sum_{s=1}^{n}\sum_{k=1}^{c}\mu_k(x_s)\log_2(\mu'_k(x_s)) + \sum_{s=1}^{n}\sum_{k=1}^{c}\mu_k(x_s)\log_2(\mu_k(x_s)) \tag{6}$$

where $\mu_k(x_s)$ is the fuzzy degree of x_s corresponding to class k in the classification result, and $\mu'_k(x_s)$ is the fuzzy degree of x_s corresponding to class k in the reference data. A small value of cross-entropy indicates accurate classification.

4. Experimental section

In this section, the NTD data set of Heshun, Shanxi, China was used as an example to illustrate how to use the rough fuzzy set based model to extract fuzzy spatial decision rules. In Bai et al. [3], the classical rough set theory was used to model the NTD data, and the decision attribute contained only two decision values that indicate whether there was NTD instances. It was only an overall description of the occurrence of NTD instances. However, the degree of the occurrence of such disease, for example, some villages suffered more than others, was not represented in the classical rough set-based model. This will lead to a fuzzy decision whether a village is vulnerable to NTD while the conditional attributes, which was used to describe the environmental and social factor of each village, is still crisp. Therefore, the rough fuzzy set based model is more appropriate than the classical rough set model.

4.1. Data description

The NTD data has been collected over the years and has been studied in many related works [3,23–26,43,44,48]. In Heshun, most people are farmers whose living environment seldom changes, and no wide-range migration has ever occurred in this district in the past. People here share similar inherited and congenital causes of birth defects. Yet, this explains only a few NTD cases. In the study area, there were 322 villages and one town. The locations of the 322 villages were determined by the Geographical Information System for spatial analysis. All the data was collected by our own field survey. This is a research project approved by the Ministry of Science and Technology of the Peoplés Republic of China. The study used only local statistical data. There are no experimental works or ethical issues.

As there were no boundaries defined for the villages, we drew them for each village using a Voronoi polygon (Fig. 2). In this experiment, the whole study area was divided into two parts randomly. One part with 167 villages was used as sample data, while the other part with 148 villages was used as reference data. The new model extracted rules from sample data. The rules extracted were validated by the reference data.

Both spatial and social attributes of all of the villages in the test area were collected. The social attributes included GDP per capita, number of children born, number of children with NTD, fertilizer used in the area (Fertilizer), access to a doctor (Doctor), production of fruit (Fruit), and production of vegetables (Vegetables). The spatial attributes included elevation, soil type, rivers, roads, lithology type, land cover type, and faulting attributes. All the maps of the attributes can be found in Wang et al. [43] and Bai et al. [3]. According to the data requirements of a decision system, some of the spatial attributes, such as soil type, lithology type and land cover type, were used directly while other spatial attributes, such as rivers, roads, and faulting attributes were transformed to analyze the information they carry. The decision table of the first five

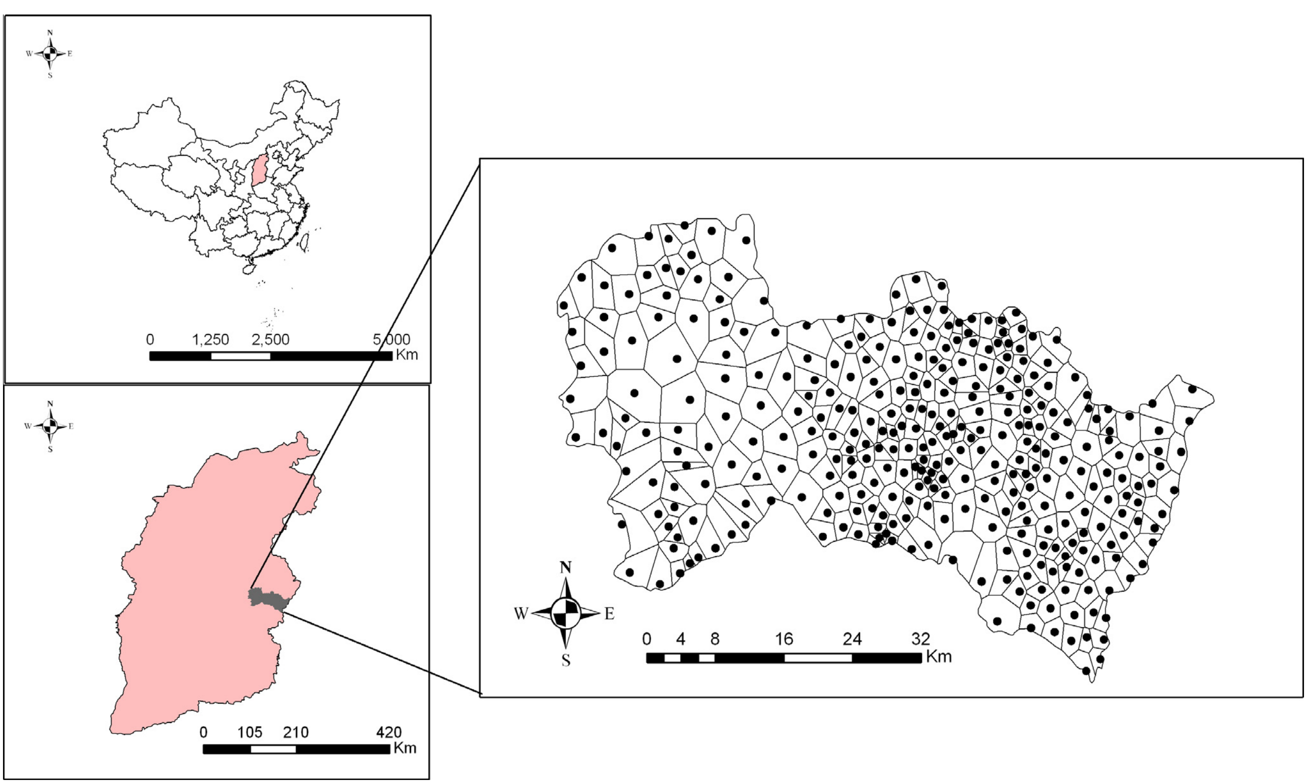

Fig. 2. Position of the study area and villages in the study region with their Voronoi polygons.

Table 2
The first five rows in the original decision table of NTD data.

GDB	Doctor	Fruit	Fertilizer	Vegetable	...	Landcover	Elevation	NEIGHBOR
2.8944	1	0	10	4	...	32	1300	1
3.6411	0	2	18	9	...	21	1300	3
5.0321	1	1	7	18	...	32	1200	2
3.6496	0	1	20	19	...	23	1349.19	4
2.8043	1	2	48	20	...	33	1274.59	5

Table 3
The relationship between the NEIGHBOR with LOW and HIGH.

NEIGHBOR	LOW	HIGH
0	0.723404	0.276596
1	0.736842	0.263158
2	0.693548	0.306452
3	0.615385	0.384615
4	0.291667	0.708333
5	0.307692	0.692308
6	0.333333	0.666667

villages from the original data is shown in Table 2 to illustrate the modeling and pre-processing of NTD data. Some of the attributes are omitted for simplicity. The last column of Table 2 is the decision attribute.

4.2. Fuzzy decision attributes

The occurrence of NTD instances in Heshun was spatially clustered. This can be verified through calculating Moran's I index using ArcGIS. Its value is 0.06 and the Z score is 6.68, which means that there is less than 1% likelihood that the occurrence of NTD is spatially randomly distributed [3]. Therefore, the villages adjacent to the infected ones are more likely to have NTD instances. This may result from the first law of geography [42]. Therefore, in the study area, whether a village has a high or low likelihood of having NTD instances can be determined through the occurrence of NTD instances in nearby villages. These two concepts are fuzzy sets and they can be used as the decision attributes: high likelihood of having NTD instances (denoted HIGH) and low likelihood of having NTD instances (denoted LOW). The memberships of the decision attributes are determined by a voting process. The details are shown below:

i. All the villages are divided into two categories: Have NTD and Have no NTD. In this step, the villages that have few new babies (fewer than five) are removed since they have no statistical significance. The remaining 271 villages participate in the voting process.
ii. A new attribute NEIGHBOR is generated through counting the number of nearby villages which have NTD instances for each village.
iii. For each possible value of NEIGHBOR, the proportion of villages having NTD can be calculated. The proportion is used to represent the corresponding fuzzy membership. For example, there are 26 villages for which NEIGHBOR is 3. Among these 26 villages there are 16 villages that have NTD. Therefore the membership corresponding to HIGH is 0.384615385 and the membership corresponding to LOW is 0.615384615. The relation between the NEIGHBOR and LOW and HIGH is show in Table 3.

The transformed FDIS of the first five villages is shown in Table 4. As the relation between NEIGHBOR and the fuzzy decisions has been calculated, the fuzzification of the decision attribute is replacing NEIGHBOR in the original decision table with corresponding fuzzy membership values.

4.3. Discretization of conditional attributes

In this paper, an FDIS is discretized using the MDLP (Minimum Description Length Principles) [28] algorithm. As the FDIS is transformed from the original decision table, the discretization is performed on the original decision table to eliminate the uncertainty introduced by the fuzzification of the decision attribute. The cuts of the continuous attributes are shown in Table 5.

Unlike other continuous attributes, the GDP per capita attribute of the study area, which was converted into 1970 U.S. dollars, was divided into three groups: villages which have not entered the first stage of industrialization (GDP < \$280), villages which are in the first stage of industrialization (280\$ $\leqslant$ GDP < 560\$) and villages which are in the second stage of industrialization (560\$ $\leqslant$ GDP < 1120\$) [16]. The three stages are denoted as I_0, I_1 and I_2, respectively. The discretization results of the first five villages are shown in Table 6.

4.4. Finding a minimal reduct

After the preprocessing of an FDIS, a genetic based reduction algorithm is performed to find a minimal reduct of the FDIS. The algorithm builds a distinction table according to the discernibility matrix. Let B be a binary matrix $(N^2 - N)/2 \times (m_c + 1)$, which is denoted by DTT. N is the number of objects in U. Each column of the matrix corresponds to one attribute; each row corresponds to one pair (k, l) which denotes a pair of two different surface objects in U. The last column corresponds to the decision (treated as an attribute); $dtt(i, (k, n))$ is an element of DTT, where i means the column number.

$$dtt(m_c + 1, (k, l)) = \begin{cases} 1 & E_{AT}(k) \neq E_{AT}(l) \\ 0 & E_{AT}(k) = E_{AT}(l) \end{cases} \quad (7)$$

Table 4
The first five rows in the FDIS transformed from the original decision table.

GDB	Doctor	Fruit	Fertilizer	Vegetable	...	Landcover	Elevation	HIGH	LOW
2.8944	1	0	10	4	...	32	1300	0.2632	0.7368
3.6411	0	2	18	9	...	21	1300	0.3846	0.6154
5.0321	1	1	7	18	...	32	1200	0.3065	0.6935
3.6496	0	1	20	19	...	23	1349.19	0.7083	0.2917
2.8043	1	2	48	20	...	33	1274.59	0.6923	0.3077

Table 5
Summary of spatial attributes, social attributes and fuzzy decisions.

Attribute Name	Cuts
Doctor	1
Fruit	1
Fertilizer	6, 8, 24, 32, 48
Vegetable	5, 11, 23, 76
Riverbuffer	3, 5, 7
Roadbuffer	3
Faultagebuff	3, 5
Elevation	1384.10

And

$$dtt(i,(k,l)) = \begin{cases} 1 & (a_i(k) \neq a_i(l) \text{ and } dtt(m_c+1,(k,l)) = 0) \\ & \text{or } dtt(m_c+1,(k,l)) = 1 \\ 0 & a_i(k) = a_i(l) \text{ and } dtt(m_c+1,(k,l)) = 0 \end{cases} \tag{8}$$

Taken the first and forth objects in Table 1 as example. $dtt(m_c+1,(1,4)) = 1$ because $E_{AT}(1) \neq E_{AT}(4)$. Meanwhile, $dtt(1,(1,4)) = 1$, $dtt(2,(1,4)) = 1$ and $dtt(3,(1,4)) = 1$ because these two objects are different at all the three attributes and $dtt(m_c+1,(1,4)) = 1$.

The algorithm proceeds as follows. Firstly, the fuzzy decision table is converted to a distinction table according to Eqs. (7) and (8). Secondly, appropriate parameters, including fitness function, crossing-over and mutation, are set for the genetic algorithm. Next, Genetic algorithm [29] is performed to select the best genomes (probably there are several reducts which all have the highest fitness score). Finally, these genomes are translated to attribute sets. These attribute sets are the reducts of FDIS.

In finding minimal reducts, the fitness function is similar to that used by Wróblewski [47], which not only considers the fewest occurrences of "1" in the chromosome, but also distinguishes as many pairs as it can. Chromosomes are candidates of minimal reducts. Formally, a chromosome is a bit string of length m_c. Each bit of a chromosome represents an attribute. If an attribute $a \in AT$ is in a candidate of minimal reducts represented using a chromosome, then the corresponding bit in the chromosome is set to "1" or else the corresponding bit is set to "0". This fitness function is shown below:

$$Fr = \frac{m_c - Lr}{m_c} + \frac{Cr}{K} \tag{9}$$

in which Lr is the number of occurrences of "1" in the chromosome, K is the number of rows in the DTT, and Cr represents the number of object pairs that can be distinguished by the chromosome. If $Cr = K$, then the current chromosome gets a 0.5 bonus.

In the experiment, the crossing-over was set to 0.7 and the mutation was set to 0.05. The attribute set {Vegetable, Landcover} was selected as the reduct of the FDIS. From the reduct result, it can be seen that land cover and production of vegetables related more to the severity of neural tube birth defect in Heshun, Shanxi, China than other attributes. The reduced FDIS of the first five villages is shown in Table 7. Only the attributes in the reduct and fuzzy decisions are pertained.

Table 7
The first five rows in the reducted FDIS using x_E reduct.

Vegetable	Landcover	HIGH	LOW
[*, 5)	32	0.2632	0.7368
[5, 11)	21	0.3846	0.6154
[11, 23)	32	0.3065	0.6935
[11, 23)	23	0.7083	0.2917
[11, 23)	33	0.6923	0.3077

4.5. Rule extraction and identification of unseen objects

In this step, 44 rules with support and confidence measures were generated according to the reduct, using methods by reading off the lines in the reduced FDIS. The rules extracted using the first five villages with their support in the FDIS are shown in Table 8. Taken the rule generated from the first village in Table 7 as an example. There are seven villages in the FDIS which have the same conditional attributes with that of the first rule in Table 8. The seven villages have different HIGH and LOW fuzzy membership values which are calculated using the rough membership functions $l_{HIGH}(1)$ and $l_{LOW}(1)$ respectively. For example

$$\begin{aligned} l_{LOW}(1) &= (0.7368 + 0.6154 + 0.6935 + 0.7368 + 0.7234 + 0.7234 \\ &\quad + 0.7368)/7 \\ &= 0.7095. \end{aligned}$$

To investigate the effectiveness of the fuzzy rough set based model, the reference data were used to verify the accuracy of the rules derived from the sample data set. All the continuous attributes of the reference data were first discretized using same cuts for the sample data prior to labeling reference data. For example, the GDP in the reference data was discretized into three categories. Next, the standard voting [32] was used to complete the rule matching process through the support of the rule. Finally, each village in the reference data was labeled a membership value for each fuzzy concept in terms of the rules selected.

5. Discussion

5.1. Accuracy assessment

To assess the accuracy of the identification result, fuzzy entropy and fuzzy cross-entropy were calculated using Eqs. (5) and (6) respectively. There were 148 villages in the reference data and six of them were predicted as UNDEFINED. The sum of all the fuzzy entropies was 127.58, the average fuzzy entropy was 0.90 and the standard deviation was 0.0501. The sum of all the fuzzy cross-entropies was 10.33, the average fuzzy cross-entropy was 0.073 and the standard deviation was 0.1504. Only 27 villages had cross-entropies larger than 0.1.

Table 6
The first five rows in the discretized FDIS using MDLP methods.

GDB	Doctor	Fruit	Fertilizer	Vegetable	...	Landcover	Elevation	HIGH	LOW
I_0	[1, *)	[*, 1)	[8, 24)	[*, 5)	...	32	[*, 1384.10)	0.2632	0.7368
I_0	[*, 1)	[1, *)	[8, 24)	[5, 11)	...	21	[*, 1384.10)	0.3846	0.6154
I_0	[1, *)	[1, *)	[6, 8)	[11, 23)	...	32	[*, 1384.10)	0.3065	0.6935
I_0	[*, 1)	[1, *)	[8, 24)	[11, 23)	...	23	[*, 1384.10)	0.7083	0.2917
I_0	[1, *)	[1, *)	[48, *)	[11, 23)	...	33	[*, 1384.10)	0.6923	0.3077

Table 8
The first five rules generated of the FDIS.

Rules	Support
If Vegetation is less than 5 and the land cover type is 32, then fuzzy membership to HIGH and LOW are 0.2632 and 0.7095 respectively	4.9661
If Vegetation is between 5 and 11 and the land cover type is 21, then fuzzy membership to HIGH and LOW are 0.3468 and 0.6514 respectively	1.9542
If Vegetation is between 11 and 23 and the land cover type is 32, then fuzzy membership to HIGH and LOW are 0.2740 and 0.7260 respectively	2.9039
If Vegetation is between 11 and 23 and the land cover type is 23, then fuzzy membership to HIGH and LOW are 0.3853 and 0.6147 respectively	2.4588
If Vegetation is between 11 and 23 and the land cover type is 33, then fuzzy membership to HIGH and LOW are 0.3774 and 0.6226 respectively	3.7355

The fuzzy entropies and fuzzy cross-entropies of the villages in the reference data are illustrated in Figs. 3 and 4. There were 19 villages, which correspond to the darkest polygons in Fig. 3, with fuzzy entropy greater than 0.973764. Meanwhile it can be seen that the fuzzy entropy was always greater than 0.831537. Besides the fuzzy entropies, the cross-entropy of each village is shown in Fig. 4. There were only 21 villages with cross-entropy greater than 0.1, and only 14 villages with cross-entropy greater than 0.3. This means that there were only a few villages whose identification results were significantly different from the reference data. At the same time, there were 111 villages whose cross-entropies were less than 0.05. These 111 villages account for 75% of the reference data. From the fuzzy cross-entropies, it can be seen that the identification result coincided with the reference data.

To further illustrate the difference between the identified fuzzy decisions and the real fuzzy decisions, the difference of fuzzy set HIGH on the reference data between the identification result and the reference data is shown in Fig. 5. There were 14 villages with a difference larger than 0.3 and 101 villages with difference less than 0.1. This illustrates that the difference between the reference data and the identification result was slight.

To inspect the reason why the identification results of the 14 villages differed greatly from the real fuzzy decisions of the reference data, some of the villages are analyzed in detail. Gaoqiu was one of the villages wrongly identified; it lies in the right-hand part of the map. It has a landcover value 122, and is the only village that has such landcover type in the east part of the county. The training data has several villages whose landcover attribute is 122 in the middle of the county. The environmental conditions may differ between these two areas, so the rules obtained in the middle area may not be generalizable to the east part. Therefore, Gaoqiu was wrongly labeled. For the same reason, Yichen was also wrongly identified. The accuracy of the identification may be further improved by dedicated sample design or consideration of spatial heterogeneity in the model. From the accuracy assessment, it can be seen that the identification result coincided with the reference data, although the uncertainty of the fuzzy result was high. Meanwhile the comparison between the identification result and the reference data also supports the judgment.

To illustrate the effect of using fuzzy decisions, we also use rough set and rough fuzzy set to handle the crisp decision. The comparison with classical rough set will be introduced in Section 5.2. Here we first use if a village have NTD instances as the severity of the disease, i.e., decision attribute. If a village has NTD instances, then its membership to the decision of high likelihood of having NTD instances is set to 1 and 0 otherwise. Using the rough fuzzy set based model, 37 rules were extracted from the FDIS with the same sample data. Then these rules were applied to the reference data in the same way. The accuracy assessment shows that the average cross entropy of the classification was 5.96 and only a

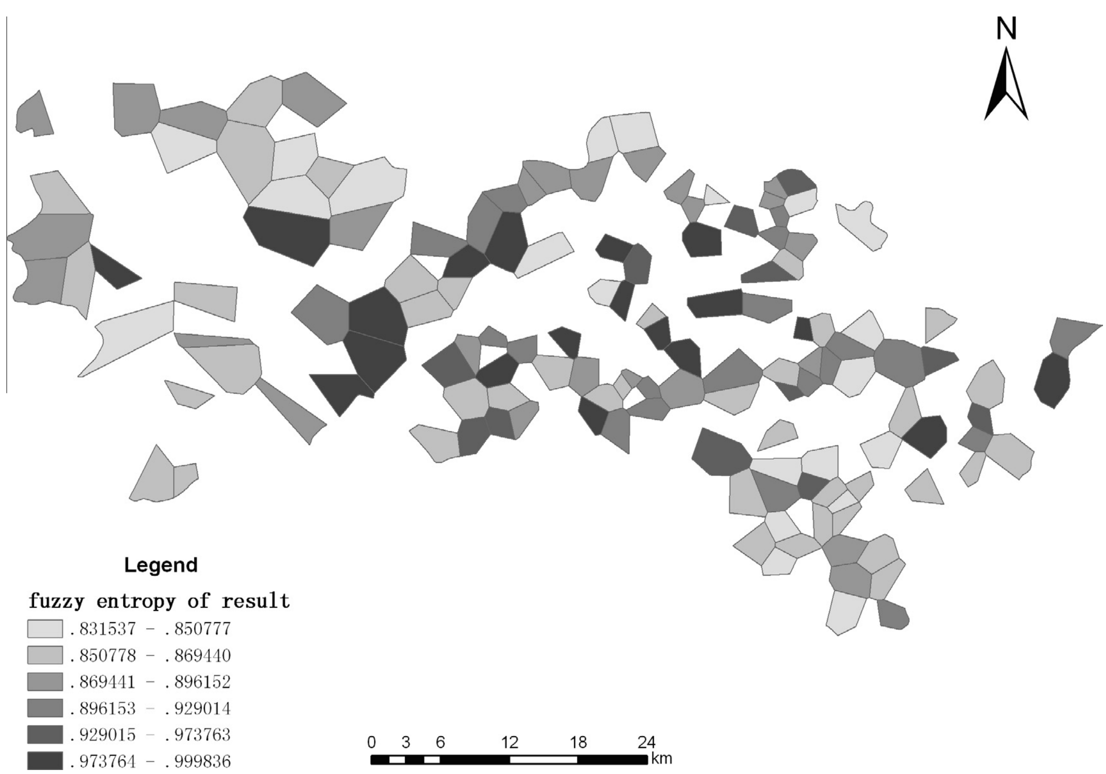

Fig. 3. Fuzzy entropy of the identification results.

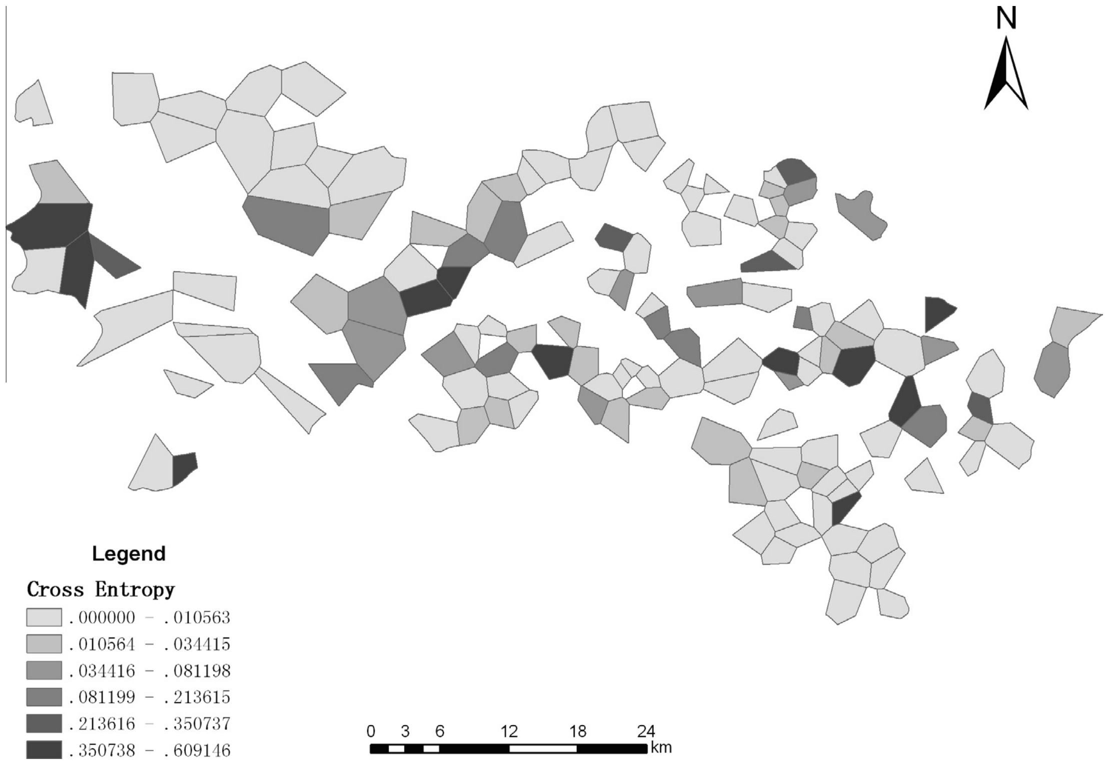

Fig. 4. Fuzzy cross-entropy of the identification results.

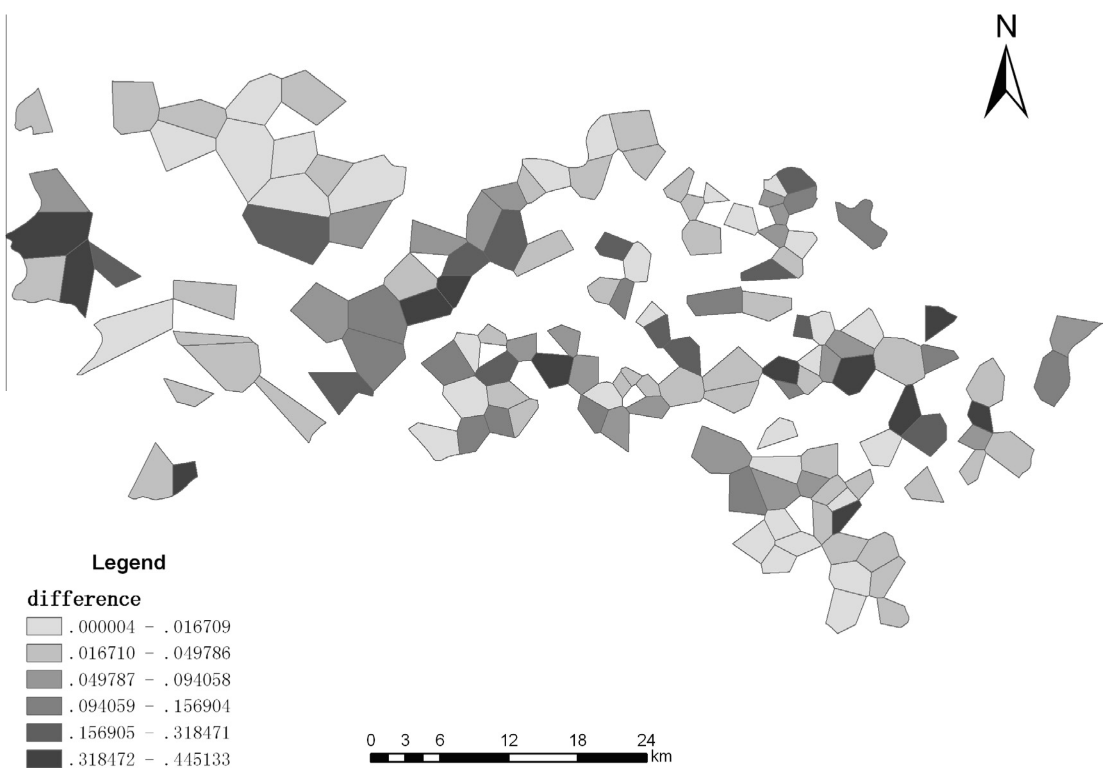

Fig. 5. Difference between the identification results and the reference data.

few villages' cross entropy was less than 1. Meanwhile, the average difference between the classification result and the real value was 0.4. These indices are much larger than the result using fuzzy decisions constructed in Section 4.2. This supports that the use of fuzzy decisions can help improve the identification accuracy.

5.2. Comparison with classical rough set theory

The physical meaning of rough fuzzy sets of spatial data is different from that of the classical rough sets. In the classical rough sets, the upper approximation and the lower approximation of concepts are two crisp sets of spatial objects. For example, a spatial object can either belong to or not belong to the lower approximation of concepts. However, in rough fuzzy set based model, a spatial object can belong to both the lower approximation and upper approximation of concepts to some extent, i.e., the lower approximation and upper approximation of concepts are fuzzy.

Fig. 6(a and b) illustrates the lower approximation and upper approximation of the fuzzy concept LOW respectively and Fig. 6(c and d) represents the lower approximation and upper approximation of fuzzy concept HIGH respectively for the sample data. From Fig. 6, it can be seen that rough fuzzy sets uses two fuzzy sets to describe a fuzzy concept's upper approximation and lower approximation. The membership values of elements in the upper approximation are greater than those of elements in the lower approximation.

Furthermore, although the FDIS can be degraded into a classical decision table, the classical decision table may lose some of the information of the fuzzy sets. For example, the fuzzy sets cannot be degraded too coarsely to hamper appropriate division of the objects, which would finally affect the accuracy of classification result. Meanwhile, if the fuzzy sets are degraded into too many sub crisp sets, the rules obtained from the decision table may be over-fitted. In considering these, it can be seen that the rough fuzzy set based model is more suitable for the FDIS.

The reduct generated using rough fuzzy sets may be different from the reduct from classical rough set theory [3]. In classical rough set theory, the attributes in the reducts included watershed, gradient, neighbor and landcover type. However in rough fuzzy sets, the attributes in the reduct were vegetable and landcover type. From the physical meaning of the classical rough set theory, the four attributes in the reduct are most relevant to the decision whether or not a village has NTD instances in all attributes. The reduct shows that some environmental factors and neighboring villages' influences are the key factors. Nevertheless, the physical meaning of the reduct of rough fuzzy sets is selecting attributes which are closely related to the strength of the membership of a state. In our experiment, the vegetable and landcover types were most relevant to the severity of the NTD occurrences in a village in all attributes. This means that the social factor vegetable can affect the NTD occurrence. The vegetables in the local farm would absorb chemical elements from the soil. Therefore, more poisonous elements will be absorbed and condensed in the people who eat more of these vegetables. And in these villages, when the output of vegetables is sufficient, people mainly eat vegetables produced by themselves. From correlation analysis between vegetable and HIGH, the Pearson value is 0.133 and it is significant at 0.01 level. This indicates that the attribute vegetable positively correlates to the fuzzy decision HIGH.

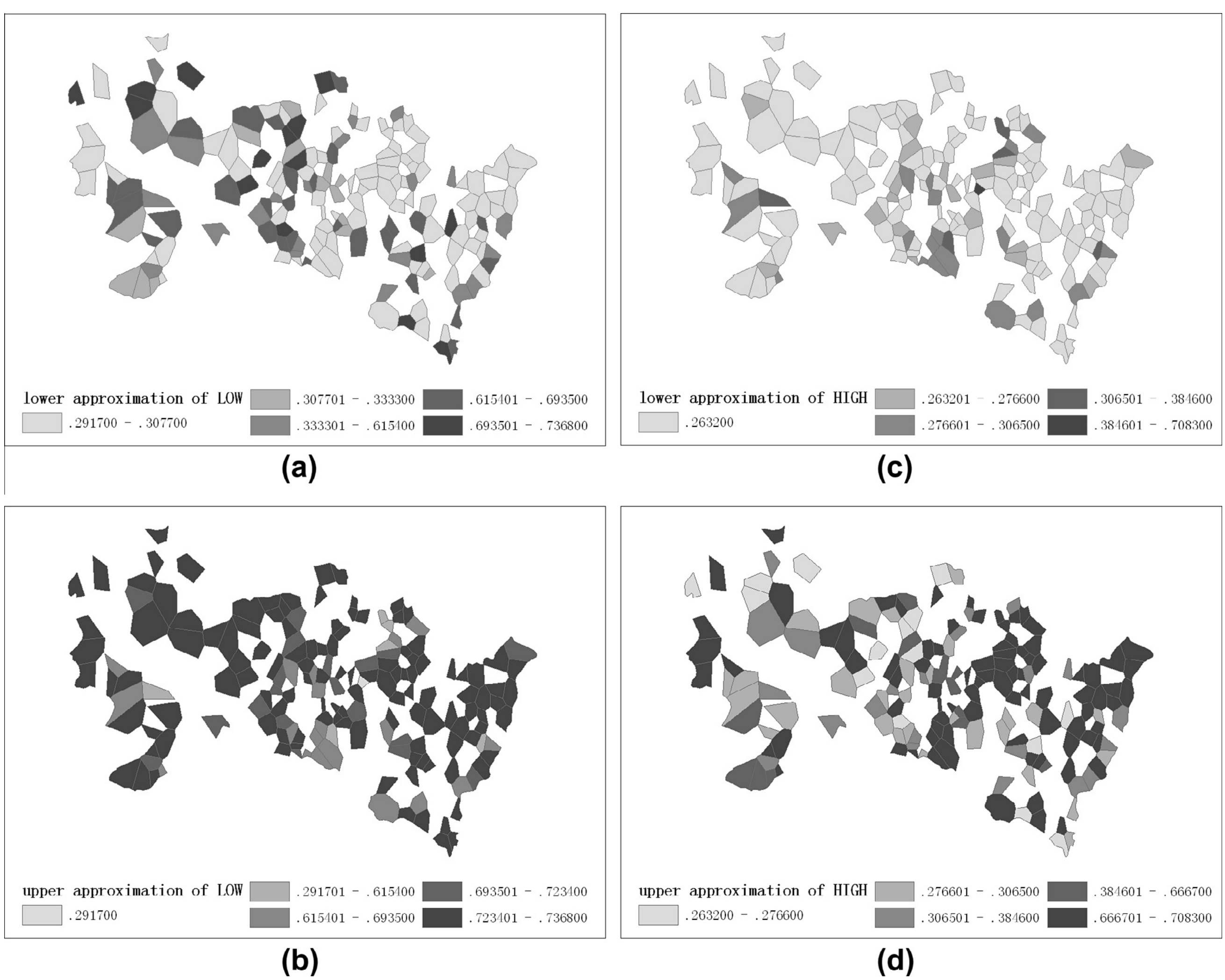

Fig. 6. HIGH and LOW's upper and lower approximations of sample data. (a) The lower approximation of fuzzy set LOW. (b) The upper approximation of fuzzy set LOW. (c) The lower approximation of fuzzy set HIGH. (d) The upper approximation of fuzzy set HIGH.

Rough fuzzy set theory can produce fuzzy decision rules because rough fuzzy sets approximate the fuzzy concepts. Each unseen object can be assigned a fuzzy membership value for each fuzzy decision. For example, a traditional spatial decision rule may look like "if vegetable < 5 and landcover = 32, then the village have no NTD instances", which only gives a crisp decision value. However a fuzzy decision rule is "if vegetable < 5 and landcover = 32, it has a membership value 0.7094 corresponding to fuzzy set LOW, and a membership value 0.2906 corresponding to fuzzy set HIGH" which gives a soft classification. This means that the fuzzy decision rules contain more information than the classical rules. Therefore, the decision maker can obtain more information from rough fuzzy set based decision making. With respect to the NTD data used, besides whether or not a village has NTD instances, the severity or risk of having it is also offered.

5.3. Comparison with classical decision rule extraction methods

To evaluate the effectiveness of the rough fuzzy set based fuzzy decision rule extraction model in NTD data of Heshun, Shanxi, China, two classical rule extraction models which use neither rough sets theory nor fuzzy sets theory are applied to the same data sets for comparison. The first method is the recursive partitioning and regression trees [5], which is denoted as "rpart" for convenient. The second method is C4.5 decision trees and rule-based model [36], which is denoted as "C45" for short. Two ready for use packages, rpart and C50, in R are used to perform the rule extraction and classification tasks.

These two methods use NEIGHBOR as the decision attribute to represent the degree of having NTD instances. First the sample data was used to generate decision rules. Then these rules were applied to the reference data. The decision values of reference data were the number of neighboring villages which have NTD instances. Then the decision values were fuzzified using the method used in Section 4.2. The fuzzified values were used as the HIGH and LOW membership value of each village in the reference data.

The fuzzy entropies and fuzzy cross entropies of all villages are calculated for the results from the two classical methods. The total fuzzy entropies and total fuzzy cross entropies of the two results are shown in Table 9. Both total fuzzy entropies of the two classical methods were larger than that of the rough fuzzy sets based model. This means that the degrees of fuzziness of the results from these two classical methods are larger than that of the result from the rough fuzzy sets model. The total fuzzy cross entropies of the two classical methods were also larger than that of the model proposed, i.e., the decisions made by the two classical methods were less coincide with the real fuzzy decision of reference data compared with the decisions made by the rough fuzzy set based model.

The differences of fuzzy decision HIGH between the identification results from the two classical methods and the real values were also calculated. There were 48 villages which differences were larger than 0.3 and 62 villages which differences were larger than 0.1 for decisions generated using rpart. There were 36 villages which differences were larger than 0.3 and 55 villages which differences were larger than 0.1 for the decisions made by C45. The number of villages, which differences were larger than 0.3, was larger than double the corresponding number of the result from the rough fuzzy sets based model. From these comparisons, it can be seen that the rough fuzzy set based decision rule extraction model is more suitable in dealing with FDIS than classical decision rule extraction methods which ignore roughness and fuzziness in data.

The main difference between traditional methods and rough fuzzy sets based methods is the different strategies in handling inconsistency of decision tables. Traditional methods tend to select the decision which is most common in the sample data. The inconsideration of fuzziness of target concepts will ignore some useful information of the ill-defined borders among fuzzy concepts which then leads to the decreasing of prediction accuracy. The rough fuzzy sets based model considers the fuzziness of target concepts. Accordingly, the prediction accuracy of rough fuzzy sets based model was higher than that of the traditional models in the experiment.

Table 9
Fuzzy entropies, fuzzy cross entropies and the number of villages which difference of real and identified membership value of HIGH is larger than 0.1 (denoted as Difference > 0.1).

Method	Total fuzzy entropy	Total fuzzy cross entropy	Difference > 0.1
rpart	128.39	24.27	62
C45	128.12	19.26	55
Rough fuzzy sets	127.58	10.33	47
WM	141.35	10.91	74
GFS	141.35	10.90	74
HyFIS	138.60	20.03	76

5.4. Comparison with fuzzy set based decision rule extraction methods

Three commonly used fuzzy set based methods were compared with rough fuzzy set based model to provide more comprehensive comparison with existing methods. There were three fuzzy rule extraction methods used in the comparison. The first method was Wang and Mendel model (WM) [46]. WM is a five step procedure based fuzzy rule extraction model. It is a simple one-pass build-up procedure and no time-consuming iterative training is required. The second method was genetic fuzzy systems (GFS) [27]. The genetic algorithm was used to determine the structures of the fuzzy IF-THEN rules and the membership function parameters. The third method was the hybrid neural fuzzy inference system (HyFIS) [19] which introduced the learning power of neural networks to fuzzy logic systems and provides linguistic meaning to the connectionist architectures.

The three methods need different parameters. The numbers of labels for WM, GFS and HyFIS were all set to seven. The type of the membership function of WM was set to trapezoid. The type of the defuzzification method that WM used was weighted average method. The type of t-norm and s-norm of WM were set to minimum and maximum respectively. GFS is a genetic algorithm. Its population size was set to 10. The percentage of crossover was 0.9. The maximal number of iterations was 10. The percentage of mutation was 0.01. The maximum number of iterations and the size of gradient descent of HyFIS were set to 100 and 0.1 respectively. These three methods proceed in the same way. First the conditional attribute and decision attribute are fuzzified using trapezoid based methods. Then the fuzzy decision rules are extracted from sample data. Next, the rules extracted are applied to the reference data. Finally the accuracy assessment is performed. Beside the accuracy assessment step, all other three steps are performed automatically using package "frbs" in R. The meaning of these parameters can be found in the manual of "frbs" package. To facilitate the comparison, Table 9 shows the total fuzzy entropies and total fuzzy cross entropies of the three methods.

From Table 9, it can be seen that the total fuzzy entropies of results from all three fuzzy set based rule extraction model were larger than that of the result from rough fuzzy set based model. This indicates that the result from rough fuzzy set based model was less uncertain than those from the three fuzzy set based methods. The total fuzzy cross entropies of results from WM, GFS and HyFIS were also larger than that of result from rough fuzzy set based model. This indicates that the difference between the real decisions and the decisions made using rule extracted from these three methods are larger, compared with decisions made using rough fuzzy set based model. Meanwhile, the differences of fuzzy decision HIGH between the fuzzy decisions identified and real fuzzy decisions are also calculated for the three fuzzy set based rules extraction methods. The numbers of villages, which differences are larger than 0.1, were 74, 74 and 79 for results from WM, GFS and HyFIS respectively, while the corresponding number of rough fuzzy set based model was 47. From the comparison, it can be seen that the rough fuzzy set based model are more suitable for modeling spatial data with both roughness and fuzziness than the decision rule extraction methods which only take fuzziness into consideration.

When the FDIS is given, the rough fuzzy sets based model are completely data driven and no subjective factors are involved in the modeling process. When there are no expert's knowledge is available, the rough fuzzy sets based model is a better choice than the fuzzy set based model which needs expert knowledge to complete the fuzzification of conditional attributes. Meanwhile, the rough fuzzy sets uses two fuzzy sets upper and lower approximation to approximate fuzzy concepts, i.e., the fuzzy decision membership of an equivalent class have a maximum value and a minimum value. These two sets can approximate a fuzzy concept when there is also roughness in the FDIS. However, the fuzzy sets based models cannot model roughness in the FDIS and have only one membership value for a decision and some methods only pertains the rules which have the maximum degree [46]. This will discard the information carried by the rules with less degrees and lead to the lower prediction accuracy of fuzzy set based model compared with rough fuzzy set based model when there are both roughness and fuzziness in the FDIS.

6. Conclusions

In this paper, a rule extraction model based on rough fuzzy sets is established for extracting spatial fuzzy decision rules in an FDIS. An example of NTD in Heshun, Shanxi, China is presented to illustrate the details of the modeling process. The fuzzy spatial decision rules found by the model can be used to identify degree of likelihood of having NTD for unseen villages through a soft classification. From the accuracy assessment of the identification result through fuzzy entropy and fuzzy cross-entropy, it can be seen that the spatial rules synthesized by the fuzzy rough set model have a reasonably good generalization to the rules obtained by the traditional rough set model. According to the comparison result between the rough fuzzy set based model and other five different methods which include two classical methods and three fuzzy set based methods, the identification results from rough fuzzy sets based model differed less from the reference data than results from other methods. This means that the model as a framework of fuzzy spatial rule extraction is an effective tool for handling spatial objects with crisp conditional attributes and fuzzy decisions.

As an extension of rough set theory, rough fuzzy set theory is also data-driven, and one of its advantages is that it does not require prior assumptions about the data in data mining. Through the computing of reducts, which is a critical step in the decision-making process, it can be used to find a minimal conditional attribute subset on which the decision attributes depend. By using the reduct we can extract more compact decision rules, which can be understood more easily than their forms obtained directly from the original decision information system. Furthermore, the generated decision rule can be used to predict or classify unseen objects. Beyond these advantages, the rough fuzzy set model can make a proper soft classification of unseen objects when there are two kinds of uncertainty in the data, roughness and fuzziness. This paper only shows the situation when there is fuzziness in the decision attributes. When a decision table has both fuzzy conditions and decisions, a fuzzy rough set approach is an alternative. In the future, we will apply the model to other applications and make improvement of the model used. For example, at present, the rough fuzzy sets do not give special attention to spatial association information, a very important issue in spatial analysis. Our future works will give special attention on how to take spatial correlation information into account to improve classification accuracy.

Acknowledgements

This work was supported in part by the National 973 Program (No. 2006CB701305), the Ministry of Science and Technology of the People's Republic of China (No. 2012BAH33B01) and National Natural Science Foundation of China (Nos. 41101440 and 61272095).

References

[1] O. Ahlqvist, J. Keukelaar, K. Oukbir, Rough and fuzzy geographical data integration, Int. J. Geogr. Inf. Sci. 17 (2003) 223–234.
[2] C. Aldridge, A Theory of Empirical Spatial Knowledge Supporting Rough Set based Knowledge Discovery in Geo-Graphic Databases, University of Otago, Dunedin, New Zealand, 1998.
[3] H. Bai, Y. Ge, J. Wang, Y. Liao, Using rough set theory to identify villages affected by birth defects: the example of Heshun, Shanxi, China, Int. J. Geogr. Inf. Sci. 24 (2010) 559–576.
[4] H. Bai, Uncertainty based Spatial Data Mining Using Rough Set Theory and Multiple Point Simulation. Ph.D. Thesis, Chinese Academy of Sciences, Beijing, 2011.
[5] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Chapman and Hall/CRC, Florida, 1984.
[6] Y.S. Chen, C.H. Cheng, Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry, Knowl.-Based Syst. 39 (2013) 224–239.
[7] J. Dai, H. Tian, W. Wang, L. Liu, Decision rule mining using classification consistency rate, Knowl.-Based syst. 42 (2013) 95–102.
[8] A. Daniels, Incorporating domain knowledge and spatial relationships into land cover classifications: a rule-based approach, Int. J. Remote. Sens. 27 (2006) 2949–2975.
[9] Y. Du, F. Liang, Y. Sun, Integrating spatial relations into case-based reasoning to solve geographic problems, Knowl.-Based Syst. 33 (2012) 111–123.
[10] S. Du, Q. Qin, Q. Wang, H. Ma, Evaluating structural and topological consistency of complex regions with broad boundaries in multi-resolution spatial databases, Inform. Sci. 178 (2008) 52–68.
[11] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (1990) 198–208.
[12] D. Dubois, H. Prade, The three semantics of fuzzy sets, Fuzzy Set Syst. 90 (1997) 141–150.
[13] M. Ester, H.M. Kriegel, J. Sander, Spatial data mining: a database approach, in: Scholl, M., Voisard, A. (Eds.) Advances in Spatial Databases, 5th International Symposium, SSD'97. Springer-Verlag, Berlin, Heidelberg, 1997, pp. 47–66.
[14] G.M. Foody, Cross-entropy for the evaluation of the accuracy of a fuzzy landcover classification with fuzzy ground data, ISPRS. J. Photogramm. 50 (1995) 2–12.
[15] R. Frank, M. Ester, A. Knobbe, A multi-relational approach to spatial classification, in: J.F. Elder IV, F. Fogelman-Soulié, P.A. Flach, M.J. Zaki (Eds.), Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, 2009, pp. 309–318.
[16] B. Hollis, R. Sherman, S. Moises, Industrialization and Growth: A Comparative Study, Oxford University, New York, 1986.
[17] Y.C. Hu, R.S. Chen, G.H. Tzen, Discovering fuzzy Association Rules using Fuzzy Partition Methods, Knowl.-Based Syst. 16 (2003) 137–147.
[18] E. Hüllermeier, Implication-based fuzzy association rules, in: L. De Raedt, A. Siebes (Eds.), Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery, Springer Verlag, London, 2001, pp. 241–252.
[19] J. Kim, N. Kasabov, HyFIS: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems, Neural Netw. 12 (1999) 1301–1319.
[20] J. Komorowski, L. Polkowski, A. Skowron, Rough sets: a tutorial, Springer-Verlag, Berlin, Heidelberg, 1999.
[21] A.J.T. Lee, Y.A. Chen, W.C. Ip, Mining frequent trajectory patterns in spatial–temporal databases, Inform. Sci. 179 (2009) 2218–2231.
[22] A.J.T. Lee, R.W. Hong, W.M. Ko, W.K. Tsao, H.H. Lin, Mining spatial association rules in image databases, Inform. Sci. 177 (2007) 1593–1608.
[23] X. Li, J. Wang, Y. Liao, B. Meng, X. Zheng, A geological analysis for the environmental cause of human birth defects based on GIS, Toxicol. Environ. Chem. 88 (2006) 551–559.

[24] Y. Liao, J. Wang, B. Meng, X.H. Li, Integration of GP and GA for mapping population distribution, Int. J. Geogr. Inf. Sci. 24 (2008) 47–67.
[25] Y. Liao, J. Wang, X. Li, Y. Guo, X. Zhen, Identifying environmental risk factors for human neural tube defects before and after folic acid supplementation, BMC Public Health 9 (2009) 391.
[26] Y. Liao, J. Wang, Y. Guo, X. Zhen, Risk assessment of human neural tube defects using a Bayesian belief network, Stoch. Env. Res. Risk. A 24 (2009) 93–100.
[27] J. Liska, S.S. Melsheimer, Complete design of fuzzy logic systems using genetic algorithm, in: B. Sayyarrodsari, A. Homaifar, W.E. Snyder (Eds.), Proceedings of the 3rd IEEE Conference on Fuzzy Systems. IEEE World Congress on Computational Intelligence, IEEE Press, Piscataway, New Jersey, 1994, pp. 1377–1382.
[28] H. Liu, F. Hussain, C.L. Tan, M. Dash, Discretization: an enabling technique, Data. Min. Knowl. Disc. 6 (2002) 393–423.
[29] T.M. Mitchell, Machine Learning, McGraw-Hill Science, Ohio, 1997.
[30] J.C. Niebles, H. Wang, L. Fei-Fei, Unsupervised learning of human action categories using spatial–temporal words, Int. J. Comput. Vis. 79 (2008) 299–318.
[31] J.Q. Niu, Y. Zhang, W.J. Feng, L.L. Ren, Spatial association rules mining for land use based on fuzzy concept lattice, in: X. Li, S. Bao (Eds.), 19th International Conference on Geoinformatics, IEEE, New Jersey, 2011, pp. 1–6.
[32] A. Øhrn, Discernibility and Rough Sets in Medicine: Tools and Applications. PhD Thesis. Department of Computer and Information Science. Norwegian University of Science and Technology, Trondheim, Norway, 1999.
[33] X. Pan, S.Q. Zhang, H.Q. Zhang, X.D. Na, X.F. Li, A variable precision rough set approach to the remote sensing land use/cover classification, Comput. Geosci. 36 (2010) 1466–1473.
[34] Z. Pawlak, Rough sets, Int. J. Inform. Comput. Sci. 11 (1982) 341–356.
[35] R.G. Pontius, J.D. Cornell, C.A.S. Hall, Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica, Agric. Ecosyst. Environ. 85 (2001) 191–203.
[36] R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, Burlington, Massachusetts, 1993.
[37] D.V. Rao, V.V.S. Sarma, A rough–fuzzy approach for retrieval of candidate components for software reuse, Pattern Recogn. Lett. 6 (2003) 875–886.
[38] S. Schockaert, P.D. Smart, F.A. Twaroch, Generating approximate region boundaries from heterogeneous spatial information: an evolutionary approach, Inform. Sci. 181 (2011) 257–283.
[39] S. Shekhar, P. Zhang, Y. Huang, R.R. Vatsavai, Trends in spatial data mining, in: H. Kargupta, A. Joshi (Eds.), Data Mining: Next Generation Challenges and Future Directions, AAAI/MIT, 2003.
[40] W.Z. Shi, Principle of Modelling Uncertainties in Spatial Data and Analysis, Science, Beijing, 2005.
[41] A. Skowron, C. Rauszer, The discernibility matrices and functions in information systems, in: R. Slowinski (Ed.), Intelligent Decision Support: Handbook of Applications and Advances of the Rough Set Theory, Kluwer Academic, Dordrecht, 1992, pp. 331–362.
[42] W. Tobler, A computer movie simulating urban growth in the Detroit Region, Econ. Geol. 46 (1970) 234–240.
[43] J. Wang, X. Liu, G. Christakos, Y. Liao, X. Gu, X. Zheng, Assessing local determinants of neural tube defects in the Heshun Region, Shanxi Province, China, BMC Public Health 10 (2010) 52.
[44] J. Wang, X. Li, G. Christakos, Y. Liao, T. Zhang, X. Gu, X. Zheng, Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China, Int. J. Geogr. Inf. Sci. 24 (2010) 107–127.
[45] S. Wang, D. Li, W.Z. Shi, X.Z. Wang, Geo-rough space, Geo-Spatial Inf. Sci. 6 (2003) 54–61.
[46] L.X. Wang, J.M. Mendel, Generating fuzzy rule by learning from examples, IEEE Trans. Syst., Man, Cybern. 22 (1992) 1414–1427.
[47] J. Wróblewski, Finding Minimal Reducts Using Genetic Algorithms, ICS Research Report, Warsaw University of Technology, Warsaw, Poland, 1995, pp. 16–95.
[48] J. Wu, J. Wang, B. Meng, G. Chen, L. Pang, X. Song, K. Zhang, T. Zhang, X. Zheng, Exploratory spatial data analysis for the identification of risk factors to birth defects, BMC Public Health 4 (2004) 23.
[49] W. Xu, Y. Pang, J. Ma, S. Wang, G. Hao, S. Zeng, Y. Qian, Fraud detection in telecommunication: a rough fuzzy set based approach, in: Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming. IEEE, Beijing, 2008, pp. 1249–1253.
[50] M. Yuan, B. Buttenfield, M. Gahegan, H. Miller, Geospatial data mining and knowledge discovery, in: R.B. McMaster, E.L. Usery (Eds.), A Research Agenda for Geographic Information Science, CRC, Boca Raton, 2004, pp. 365–388.
[51] O. Yun, J. Ma, Land cover classification based on tolerant rough set, Int. J. Remote. Sens. 27 (2006) 3041–3047.
[52] L.A. Zadeh, Fuzzy sets, Inform. Contr. 8 (1965) 338–353.
[53] J. Zhang, M. Goodchild, Uncertainty in Geographical Information, Taylor & Francis, London, 2002.
[54] P. Zhu, Q. Hu, Rule extraction from support vector machines based on consistent region covering reduction, Knowl.-Based Syst. 42 (2013) 1–8.