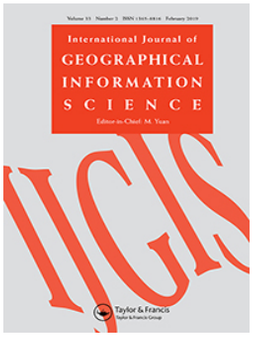


International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: http://www.tandfonline.com/loi/tgis20

A spatial heterogeneity-based rough set extension for spatial data

Hexiang Bai, Deyu Li, Yong Ge & Jinfeng Wang

To cite this article: Hexiang Bai, Deyu Li, Yong Ge & Jinfeng Wang (2019) A spatial heterogeneity-based rough set extension for spatial data, International Journal of Geographical Information Science, 33:2, 240-268, DOI: 10.1080/13658816.2018.1524148

To link to this article: https://doi.org/10.1080/13658816.2018.1524148

Published online: 15 Oct 2018.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2019, VOL. 33, NOS. 1–2, 240–268
https://doi.org/10.1080/13658816.2018.1524148

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

A spatial heterogeneity-based rough set extension for spatial data

Hexiang Bai[a], Deyu Li[a,b], Yong Ge[c] and Jinfeng Wang[c]

[a]School of Computer and Information Technology, Shanxi University, Taiyuan, China; [b]Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, China; [c]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

ABSTRACT

When classical rough set (CRS) theory is used to analyze spatial data, there is an underlying assumption that objects in the universe are completely randomly distributed over space. However, this assumption conflicts with the actual situation of spatial data. Generally, spatial heterogeneity and spatial autocorrelation are two important characteristics of spatial data. These two characteristics are important information sources for improving the modeling accuracy of spatial data. This paper extends CRS theory by introducing spatial heterogeneity and spatial autocorrelation. This new extension adds spatial adjacency information into the information table. Many fundamental concepts in CRS theory, such as the indiscernibility relation, equivalent classes, and lower and upper approximations, are improved by adding spatial adjacency information into these concepts. Based on these fundamental concepts, a new reduct and an improved rule matching method are proposed. The new reduct incorporates spatial heterogeneity in selecting the feature subset which can preserve the local discriminant power of all features, and the new rule matching method uses spatial autocorrelation to improve the classification ability of rough set-based classifiers. Experimental results show that the proposed extension significantly increased classification or segmentation accuracy, and the spatial reduct required much less time than classical reduct.

ARTICLE HISTORY
Received 3 May 2018
Accepted 11 September 2018

KEYWORDS
Rough set model; spatial information table; spatial heterogeneity; spatial autocorrelation; feature selection

1. Introduction

Spatial data play an important role in scientific research and real-world applications. It is one of the most important types of data analyzed in many research fields, such as geography, ecology, environmentology and meteorology. During the processing of spatial data, it is almost inevitable to encounter the issue in which the target geographical process or phenomenon cannot be completely described or explained using all available features or attributes of geographical objects. Pawlak (1982) found that this issue arises from a new type of uncertainty, that is, roughness, and proposed using rough set theory to manage datasets with roughness. Since then, many extensions of rough sets have been proposed to address different challenges encountered while

CONTACT Hexiang Bai baihx@sxu.edu.cn

processing real-world data with roughness, such as rough fuzzy sets and fuzzy rough sets (Dubois and Prade 1990), variable precision rough sets (Ziarko 1993), probability rough sets (Ziarko 2005) and neighborhood system rough sets (Wu and Zhang 2002).

Recently, many researchers have applied rough set theory and its extensions to model roughness in spatial data. These works can be roughly divided into two categories. One category is applying rough set theory to define rough objects and express topological relations in spatial data (Beaubouef *et al*. 2004, Murgante *et al*. 2008), or provide a geospatial representation of spatial data (Petry and Elmore 2015). The other category is using rough set theory and the concept of multi-granulation to perform machine learning tasks, such as feature extraction (Lei *et al*. 2008, Fiedukowicz 2015, Patra *et al*. 2015), rule extraction (Liu *et al*. 2011, Yan *et al*. 2016), clustering (Pal and Mitra 2002), classification (Yun and Ma 2006, Leung *et al*. 2007, Jindal 2017), prediction (Bai and Ge 2009, Bai *et al*. 2010, 2014), hierarchical inference (Sheikhian *et al*. 2015), outlier detection (Albanese *et al*. 2014), similarity measurement (Sharmila Banu and Tripathy 2016) and accuracy assessment(Ahlqvist, 2005, Ahlqvist *et al*. 2000, 2003, Ge *et al*. 2012, Banu and Tripathy 2018). This is an incomplete list of studies on the application of rough set theory to the analysis of spatial data, and the number of related studies is still increasing.

Although there have been a number of applications of rough set theory to spatial data, the rough set model or extensions used, such as the dominance-based rough set model, variable precision rough set model and fuzzy rough set model, have not taken into account the spatial pattern of spatial data. Clearly, this does not correctly reflect the actual situation of spatial data. Location gives rise to at least two classes of spatial patterns: spatial autocorrelation and spatial heterogeneity (Anselin 1992). Spatial autocorrelation refers to the concept in which objects with similar attribute values tend to aggregate in space (Anselin 1992), and spatial heterogeneity refers to the concept in which objects at a location have some degree of uniqueness relative to objects at distant locations (Wikipedia 2016).

Currently, the measurement and exploitation of spatial autocorrelation and heterogeneity has attracted the attention of many researchers. In the field of measuring spatial association, many methods, such as Ripley's *K*-function (Ripley 1976), Moran's *I* (Moran 1948), Getis' *G* (Getis and Ord 1992), join count statistics (JCS) (Cliff and Ord 1970) and normalized conditional probability Bai2016, have been developed for point-based data, continuous data on lattice data and nominal data on lattice data. Many methods are also available for analyzing heterogeneity in space. For example, the local indicators of spatial association (Anselin 1995) and local indicators of categorical data (Boots 2003) are used to measure the spatial heterogeneity of geographical phenomena. The *Q*-statistic can be used to measure spatial stratified heterogeneity (Wang *et al*. 2016). Simultaneously, many studies have exploited spatial association and heterogeneity information in the analysis of spatial data, such as geo-statistics (Goovaerts 1997), spatial filtering (Ursin 1983), spatial regression (Griffith 1989), spatial sampling (Haining 2003) and the geographical detector (Wang *et al*. 2010a).

The ignorance of spatial heterogeneity and spatial autocorrelation may lead to incorrect conclusions and solutions (Haining 2003, De Smith *et al*. 2007, Fisher and Wang 2011). We consider the boundary region (the boundary region contains equivalent classes that belong to at least two categories) in rough set theory as an example. In

Figure 1, the square box in the gray area represents an equivalent class in the boundary region. This equivalent class is mainly distributed over two parts of the study area. Objects with a 'red' label are in the west of the study area and objects with a 'blue' label are in the east of the study area.

Assuming that some unseen instances of this equivalent class are in the west of the study area, the question arises of how to label these unseen objects. All unseen instances are in the vicinity of the objects labeled 'red' in the west. As objects with the same decision in spatial data tend to aggregate rather than randomly distribute over space, it is more reasonable to label these objects 'red' than 'blue'. However, using classical rule matching methods, for example, standard voting (SV) (∅Hrn 1999), these objects are more likely to be labeled as 'blue' because there are more objects with the label 'blue' in the equivalent class. Clearly, the classical rough set (CRS) model will probably incorrectly label most of these objects because it does not consider the spatial pattern of objects.

From the perspective of rough set theory, if the conditional attributes can completely define the target concept, that is, there is no boundary region, then the ignorance of the spatial pattern barely affects the classification accuracy. However, it is difficult to collect all the necessary attributes to completely define the target concept in real-world applications. In most cases, there exist some unknown or unobservable attributes (UKO-attributes) that are necessary for defining the target concept. Because classical rule matching methods cannot use the information of UKO-attributes, most of the unseen objects in the previous example are incorrectly labeled 'blue'. However, an interesting aspect is that the spatial pattern of objects can uncover some of the information of UKO-attributes because many UKO-attributes that are used to describe the target geographical phenomenon have some form of spatial pattern.

In fact, the spatial pattern of objects is an important source of information rather than causing a problem. Many classical statistical methods have been extended to use spatial distribution information (Anselin 1988, Haining 1990, Goovaerts 1997, Wang *et al*. 2002, Leslie and Kronenfeld 2011, Páez *et al*. 2012, Guo *et al*. 2013, Bai *et al*. 2016). Consider linear regression as an example (Cressie 1993, Anselin 2002). Because traditional linear regression cannot take into account spatial patterns, the linear regression of spatial data may lead to non-Gaussian residuals. Spatial scientists have proposed using spatial linear regression to analyze spatial data. Spatial linear regression uses a novel strategy to take advantage of the spatial patterns of spatial data. This strategy uses a spatial weighted matrix or simply an adjacency matrix to consider the spatial autocorrelation of dependent variables or residuals. As spatial regression adjusts the dependent variable using neighboring objects' values and simultaneously ignores the influence of distant objects, that is, takes into account both spatial autocorrelation and spatial heterogeneity, it greatly reduces the prediction error for spatial data. In addition to spatial linear regression, this strategy has been adopted by many spatial data-oriented analysis tools and methods.

Inspired by the previous studies in the area of spatial statistics, this paper attempts to transfer the strategy used by spatial linear regression to CRS theory to take advantage of the spatial heterogeneity and spatial autocorrelation of spatial data. The new model proposed uses a spatial information system (SIS), which is a combination of a classical

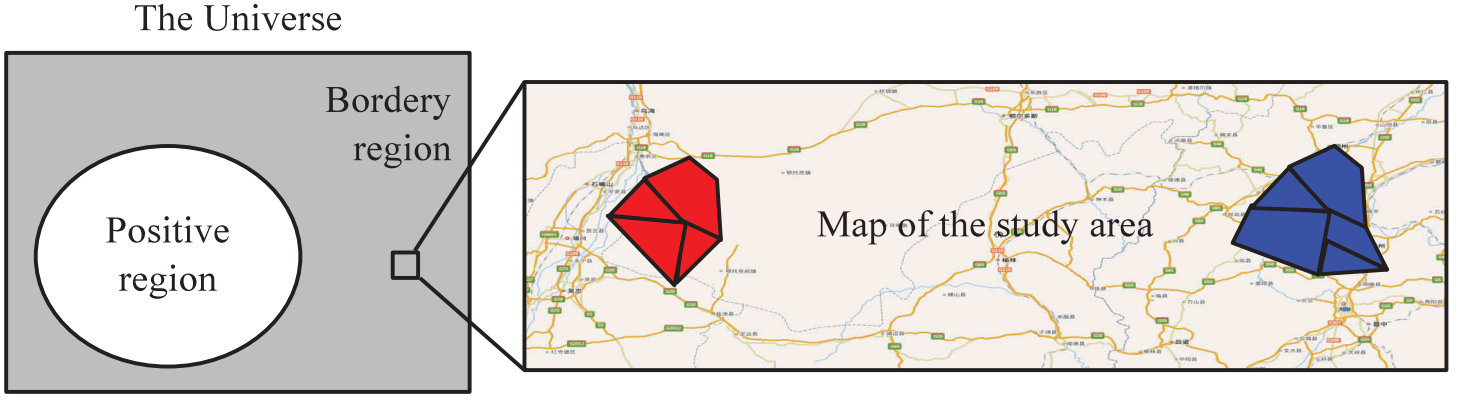


Figure 1. Illustration of the spatial distribution for an equivalent class on the boundary region for the classical rough set model.

information system and an adjacency matrix, to represent spatial data. The target concept is roughly approximated in the local spatial region of each object, and all local rough approximations are aggregated to form an overall rough approximation of the target concept with the help of the adjacency matrix. Based on the principle of preserving the local discriminatory power of all attributes, two new types of reduct are proposed. Finally, we improve the rule matching method by considering spatial autocorrelation. Two experiments were performed to validate the effectiveness of the proposed method. The first experiment predicted the occurrence of neural tube birth defect (NTD) instances in Heshun, Shanxi, China. The results showed that the proposed model was superior to the CRS. The second experiment used the spatial absolute reduct (SAR) to select an appropriate set of textural features for high-resolution remotely sensed imagery. The results showed that the spatial reduct was superior to the classical reduct in terms of both effectiveness and efficiency.

There are two main advantages of this extension. First, the consideration of spatial heterogeneity and spatial autocorrelation better reflects the characteristics of spatial data and can improve the effectiveness of the rough set model. Second, the spatial reduct requires substantially fewer computing resources than those of the CRS because the spatial reduct only considers neighboring objects and ignores distant objects.

The remainder of the paper is organized as follows: In Section 2.1, the concept of an SIS is proposed and a series of new concepts and their properties, such as the local indiscernible relation, local lower approximation and local reduct, are introduced. Then, the concept of a spatial reduct and corresponding spatial reduct finding algorithm are proposed. In Section 2.3, a new rule matching method that considers spatial autocorrelation is proposed based on the traditional SV method. In Section 3.1, the prediction of the occurrence of NTD instances and the segmentation of high-resolution remotely sensed imagery are used as two examples to validate the effectiveness and efficiency of the spatial rough set model. In Section 4, the paper is concluded.

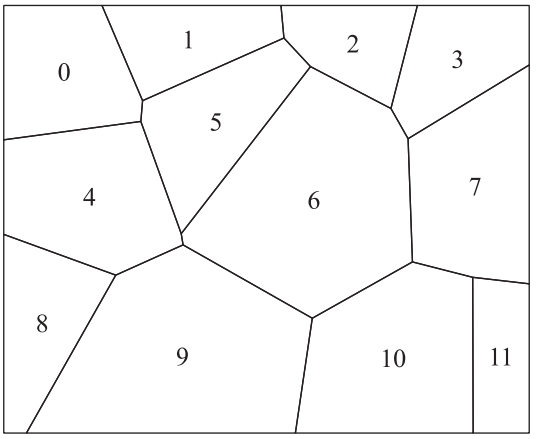
2. Spatial rough set-based prediction model

The CRS uses an information system to organize data collected for the target problem. Each row of the information system can represent an event or object. Each column of the information system represents a feature, property or attribute of objects, for example, 'Whether a patient got a fever'. This is formally defined as follows:

Definition 1: An information system is a pair $S = (U, A \cup \{d\})$, where universe U is a non-empty finite set of objects, A is a non-empty finite condition attribute set and $d \notin A$ is the decision attribute. Any $a \in A$ can be regarded as mapping $U \times \{a\} \rightarrow V_a$, where V_a is the domain of attribute a. Decision attribute d can be regarded as mapping $U \times \{d\} \rightarrow V_d$, where V_d is the domain of attribute$d$(Pawlak 1982).

Figure 2(b) is an example of an information system. It has three attributes $a1, a2, a3$ and one decision attribute *dec*. '*id*' is used to identify different objects, and the object with '$id = i$' is denoted by x_i.

However, the information system is not sufficient for representing spatial phenomena and processes because it does not take into account spatial relations among objects. For example, it is impossible to determine whether objects x_1 and x_5 are neighbors in space only from Figure 2(b). To take advantage of the spatial relations among objects, a simple



(a)

ID	a1	a2	a3	d
0	1	2	1	1
1	1	2	1	1
2	2	1	1	1
3	2	1	1	1
4	2	3	1	2
5	2	3	1	1
6	3	2	3	1
7	3	3	2	2
8	4	3	2	2
9	3	2	3	2
10	1	2	1	2
11	1	2	1	2

(b)

$$\mathrm{M}^1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

(c)

Figure 2. Example of an SIS: (a) map of the geographical objects of the study area; (b) information system corresponding to the geographical objects in (a); (c) first-order adjacency matrix of geographical objects in (a).

solution is to add the adjacency information of objects to the information system using an adjacency matrix. This is formally defined as follows:

Definition 2: The first-order adjacency matrix of the study area is defined as $M^1 = [m^1_{xy}]_{N\times N}$, where $m^1_{xy} = m^1_{yx} = 1$ if the two objects x and y in the study area are 1-adjacent; otherwise $m^1_{xy} = m^1_{yx} = 0$ (Bai *et al*. 2016).

In Definition 2, the 1-adjacency of two surface objects can be established using any connectivity algorithm (Fortin and Dale 2005). Consider Figure 2(a) as an example. Twelve objects in total are distributed over the study area. The 1-adjacency is calculated by determining whether two polygons touch each other. The corresponding first-order adjacency matrix M^1 is shown in Figure 2(c).

k-th order adjacency matrices M^k can be established using the concept of relation composition, that is, $M^k = M^{(k)} - \cup_{i=1}^{k-1} M^{(i)} - E$, where $M^{(i)}$ is the relation composition of i Ms. The k-th order adjacency matrix is defined as $M^k = [m^k_{xy}]_{N\times N}$, where $m^k_{xy} = 1$ only if

two objects x and y are adjacent to each other via other $k-1$ surface objects and not adjacent to each other via any $k' < k-1$ surface objects; otherwise, $m^k_{xy} = 0$.

By adding spatial adjacency information into the information system, we propose using a SIS to represent spatial data. An SIS consists of two parts: the traditional information system and union of different order adjacency matrices. This is formally defined as follows:

Definition 3: An SIS is a pair $SIS = (S, M(k))$, where $S = (U, A \cup \{d\})$ is an information system and $M(k) = \cup_{i=0}^{k} M^i = [m(k)_{xy}]_{N\times N}$ is the union of all adjacency matrices from the first order to the k-th order of the study area, k is a parameter that represents the maximum order of adjacency that SIS takes into account, and $m(k)_{xy}$ is the element of $M(k)$ that represents whether x and y are the α-th order neighbors of each other, where $1 \le \alpha \le k$.

For convenience, an SIS can also be denoted by a triple $SIS = (U, A \cup \{d\}, M(k))$. Figure 2(a) is an example study area from which a first-order adjacency matrix (see Figure 2(c)) can be established. Figure 2(b) is the corresponding information system. If k is set to one, then the corresponding information system and first-order adjacency matrix form an SIS of the study area. Clearly, it is easy to determine that objects x_1 and x_2 are mutually first-order neighbors in terms of Figure 2(c).

As an SIS takes into account spatial adjacency and the CRS cannot take advantage of this information, it is natural to improve the CRS for the SIS accordingly.

2.1. *Rough set model for an SIS*

The CRS uses an indiscernible relation to form a partition of the universe. Two objects are indiscernible if they share the same attribute values for all conditional attributes. An equivalent class of object x contains all the objects in the universe that are indiscernible from x, and is defined as $[x]_{A'} = \{y \in U : a(x) = a(y)\ \forall a \in A'\}$.

In spatial analysis, it is not sufficient to conclude that two objects are indiscernible only in terms of attribute values when they are distant from each other in space. Spatial heterogeneity may lead to the difference of some UKO features for the two distant objects. Accordingly, the indiscernible relation must be adapted to accommodate spatial heterogeneity. In the following, we propose using local indiscernible relation to replace a traditional indiscernible relation to improve the CRS for the SIS.

Definition 4: For three objects x, y, z in an SIS $SIS = (U, A \cup \{d\}, M(k))$, two objects y and z are x-local indiscernible if $a(y) = a(z)\ \forall a \in A$ and $m(k)_{xz} = 1$ and $m(k)_{xy} = 1$, or $a(y) = a(z)\ \forall a \in A$, where $y = x$ and $m(k)_{yz} = 1$, or $z = x$ and $m(k)_{yz} = 1$.

To summarize, two objects are x-local indiscernible if they share the same attribute values and belong to the set constituted by object x and its adjacent objects. For convenience, we call this set the x-centered region, which is denoted by $C(x) = \left\{y \in U : m(k)_{xy} = 1\right\} \cup \{x\}$. Based on the concept of an x-local indiscernible relation, the following definition shows how to construct x-local indiscernible sets, which can be considered as an x-local version of the equivalent classes in the CRS. For convenience, all the following definitions and discussions share the same SIS, $SIS = (U, A \cup \{d\}, M(k))$, unless otherwise specified.

Definition 5: The local indiscernible set of y using attribute set $A' \subseteq A$ with respect to x, or simply the x-local indiscernible set, is a set

$$LInd(y,x,A') = \{z \in U : a(y) = a(z)\, \forall a \in A' \ \wedge \ y \in C(x) \ \wedge \ z \in C(x)\}.$$

It is easy to show that $LInd(y,x,A') = [y]_{A'} \cap C(x)$. Consider the SIS in Figure 2 as an example: x_0 and x_1 are x_5-local indiscernible. x_5 centered region $C(x_5) = \{x_0, x_1, x_2, x_4, x_5, x_6\}$ and $LInd(x_1, x_5, \{a1, a2, a3\}) = \{x_0, x_1\}$. Simultaneously, we found that $[x_1]_{\{a1,a2,a3\}} = \{x_0, x_1, x_{10}, x_{11}\}$. Clearly, $LInd(x_1,\ x_5, \{a1, a2, a3\}) = [x_1]_{\{a1,a2,a3\}} \cap C(x_5)$.

Using the local indiscernible relation, objects that are spatially distant from each other can be discerned regardless of whether they have different attribute values. If two SISs of a study area share the same information system, then the corresponding $LInd(y,x,A')$s satisfies the following property.

Proposition 1. *For two SISs* $SIS1 = \{S, M(k)\}$ *and* $SIS2 = \{S, M(k')\}$, *where* $S = \{U, A \cup \{d\}\}$ *and* $k' > k$, $LInd(y,x,A')_{SIS1} \subseteq LInd_{SIS2}(y,x,A')$ *for any x,y, and* $A' \subseteq A$.

Proof: For any x, y, and A, $k < k' \Rightarrow C_{SIS1}(x) \subseteq C_{SIS2}(x) \Rightarrow [y]_{A'} \cap^{C_{SIS1}}(x)\}$, $\subseteq [y]_{A'} \cap C_{SIS2}(x) \Rightarrow LInd(y,x,A')_{SIS1} \subseteq LInd_{SIS2}(y,x,A')$. The proof is complete. □

The CRS uses two sets, lower approximation and upper approximation, to approximate a target concept. The lower approximation contains all the equivalent classes that belong to the target concept, and the upper approximation contains all the equivalent classes that have a non-empty intersection with the target concept. For an SIS, the target concept $X \subseteq U$ can be approximated at each x-centered region using all x-local indiscernible sets.

Definition 6: The x-local lower approximation of concept X using $A' \subseteq A$ consists of all the $LInd(y,x,A')$ that are subsets of X, that is,

$$\underline{appr}^{x}_{A'}(X) = \{y : LInd(y,x,A') \subseteq X\},$$

where $A' \subseteq A$. The x-local upper approximation of concept X using $A' \subseteq A$ consists of all the $LInd(y,x,A')$s that have a non-empty intersection with X, that is,

$$\overline{appr}^{x}_{A'}(X) = \{y : LInd(y,x,A') \cap X \neq \emptyset\}.$$

The x-local rough set of target concept X using $A' \subseteq A$ is a pair

$$\left\{\underline{appr}^{x}_{A'}(X), \overline{appr}^{x}_{A'}(X)\right\}.$$

Similar to the CRS, the x-local positive region and boundary region can also be constructed in terms of x-local lower approximations.

Definition 7: The x-local positive region and x-local boundary region of d using $A' \subseteq A$ is defined as $POS^{x}_{A'}(d) = \underset{d_i \in V_d}{\cup} \underline{appr}^{x}_{A'}(\{x \in U : d(x) = d_i\})$ and $BN^{x}_{A'}(d) = C(x) - POS^{x}_{A'}(d)$, respectively.

For example, in Figure 2,

$$\underline{appr}^{x_5}_{\{a1,a2,a3\}}([x_0]_d) = \{x_0, x_1, x_2, x_6\},$$

$$\overline{appr}^{x_5}_{\{a1,a2,a3\}}([x_0]_d) = \{x_0, x_1, x_2, x_4, x_5, x_6\},$$

$$POS^{x_5}_{\{a1,a2,a3\}}(d = 1) = \{x_0, x_1, x_2, x_6\}.$$

The *x*-local rough sets are a local description of the target concept. To obtain an overall description of the target concept in the entire study area, all *x*-local rough sets of target concept X are aggregated to form an overall approximation of target concept X. This is formally defined as follows:

Definition 8: The spatial rough set of target concept X using $A' \subseteq A$ is set $SRS_{A'}(X) = \left\{ \left\{ \underline{appr}^x_{A'}(X), \overline{appr}^x_{A'}(X) \right\} | x \in U \right\}$.

2.2. *Spatial reduct*

The reduct in the CRS uses a minimal attribute set that can form the same approximation of target concepts as all attributes to remove redundant attributes, such as the absolute reduct (AR) and positive region reduct (PR). However, the classical reduct only uses the difference between the collected attributes between objects. In spatial analysis, for objects that are distant from each other, the UKO features may have changed because of spatial heterogeneity. This leads to two distant objects being discernible in the SIS, even if the two objects have the same attribute values. Although it is difficult to collect UKO features, the classical reduct concept can be improved for spatial data from the perspective of spatial heterogeneity.

For convenience, a reduct for an *x*-centered region is proposed in advance. For each *x*-centered region, the minimal attribute set required for approximating the target concept can be defined from two perspectives. The *x*-local AR can preserve the *x*-local indiscernible sets, and the *x*-local PR can preserve the *x*-local positive region of the target concept. The *x*-local AR and *x*-local PR are collectively called the *x*-local reduct. Definitions 9 and 10 are the definitions of the *x*-local AR and PR, respectively. The difference between these two definitions is to preserve $LInd(y,x,B) = LInd(y,x,A)$ or $POS^x_B(d) = POS^x_A(d)$.

Definition 9: Suppose $SIS = \{U, A \cup \{d\}, M(k)\}$ is an SIS. If $LInd(y,x,B) = LInd(y,x,A)$ for $B \subseteq A$ and $LInd(y,x,B') \neq LInd(y,x,A)$ for any $B' \subset B$ and any $y \in C(x)$, then B is an *x*-local AR of *SIS*.

Definition 10: Suppose $SIS = \{U, A \cup \{d\}, M(k)\}$ is an SIS. If $POS^x_B(d) = POS^x_A(d)$ for $B \subseteq A$ and $POS^x_{B'}(d) \neq POS^x_A(d)$ for any $B' \subset B$, then B is an *x*-local PR of *SIS*.

The *x*-local reduct only preserves the discriminant power for a local area. Although it considers spatial heterogeneity, the *x*-local reduct is not sufficient for the analysis of the entire study area. Accordingly, it is necessary to calculate a minimal attribute set that can preserve the discriminant power for each *x*-centered region. Definitions 11 and 12 are the definitions of spatial AR and PR, respectively. Similarly, the difference between these two definitions is to preserve $LInd(y,x,B) = LInd(y,x,A)$ or $POS^x_B(d) = POS^x_A(d)$.

Definition 11: Suppose $SIS = \{U, A \cup \{d\}, M(k)\}$ is an SIS. If $LInd(y,x,B) = LInd(y,x,A)$ for $B \subseteq A$ and any object $x \in U$, and there exists $LInd(y,x,B') \neq LInd(y,x,A)$ for some $x \in U$ and any $B' \subseteq A$, then B is called a spatial AR (SAR) of *SIS*.

Definition 12: Suppose $SIS = \{U, A \cup \{d\}, M(k)\}$ is an SIS. If $POS^x_B(d) = POS^x_A(d)$ for $B \subseteq A$ and any object $x \in U$, and there exists $POS^x_{B'}(d) \neq POS^x_A(d)$ for some $x \in U$ and any $B' \subseteq A$, then B is called a spatial PR (SPR) of *SIS*.

For simplicity, the SAR and SPR are collectively called the spatial reduct. Consider the SAR of the SIS in Figure 2 as an example. The x_0-local AR is $\{a1\}$ or $\{a2\}$. The x_1-local AR

is $\{a2\}$. The x_2-local AR is $\{a1, a2\}$ or $\{a2, a3\}$. The x_3-local AR is $\{a2\}$ or $\{a3\}$. The x_4-local AR is $\{a1\}$ or $\{a2, a3\}$. The x_5-local AR is $\{a1, a2\}$ or $\{a2, a3\}$. The x_6-local AR is $\{a1, a2\}$ or $\{a2, a3\}$. The x_7-local AR is $\{a1, a2\}$ or $\{a2, a3\}$. The x_8-local AR is $\{a1\}$ or $\{a3\}$. The x_9-local AR is $\{a1\}$ or $\{a2, a3\}$. The x_{10}-local AR is $\{a1, a2\}$ or $\{a3\}$. The x_{11}-local AR is $\{a1\}$, $\{a2\}$ or $\{a3\}$. Accordingly, the SAR is $\{a1, a2\}$ or $\{a2, a3\}$. It is easy to show that the AR of the information system is $\{a1, a2\}$.

Because calculating the x-local reduct concerns only a small number of objects, the x-local reduct can be determined via limited computing resources. The difficult part of calculating the spatial reduct is how to calculate a minimal set that can preserve the discriminant power for all x-centered regions. For example, suppose that each x-centered region has T x-local reducts and there are N objects in the SIS, then there will be $T \times N$ possible compositions of the x-local reduct that should be taken into account when calculating the spatial reduct. Clearly, this is a typical combination explosion problem. Accordingly, we propose using a greedy algorithm (see Algorithm 1) to calculate an approximate spatial reduct for the SIS.

The greedy algorithm starts with an empty candidate spatial reduct $red = \emptyset$. For each x-centered region, candidate reduct red is tested for whether it contains an x-local reduct, that is, whether $LInd(y, x, red) = LInd(y, x, A)$ or $POS^x_{red}(d) = POS^x_A(d)$. If red does not contain an x-local reduct, then a minimal subset of $A - red$ is added to red to ensure that $LInd(y, x, red) = LInd(y, x, A)$ or $POS^x_{red}(d) = POS^x_A(d)$. When all objects are visited, the algorithm outputs an approximate spatial reduct of the target SIS. Because the entire calculating process ensures that the final attribute set contains an x-local reduct for each object x, it is easy to show that the attribute set can ensure that $LInd(y, x, red) = LInd(y, x, A)$ $(POS^x_{red}(d) = POS^x_A(d))$ for $red \subseteq A$ and any object $x \in U$. Accordingly, the attribute set found is a superset of a spatial reduct.

Because each x-centered region only contains a small fraction of objects, the checking of $LInd(y, x, red) = LInd(y, x, A)$ or $POS^x_{red}(d) = POS^x_A(d)$ is not time-consuming and can be accomplished by checking each attribute for every pair of objects in the x-centered region. During the checking process, new attributes can be added to red to ensure $LInd(y, x, red) = LInd(y, x, A)$ or $POS^x_{red}(d) = POS^x_A(d)$. Assuming that each x-centered region averagely contains m objects, the time complexity for the checking process is $O(m^2 \times |A|)$. Accordingly, the time complexity of Algorithm 1 is $O(N \times m^2 \times |A|)$.

In the CRS model, reduct finding is an NP-hard problem, and many approximate reduct finding algorithms exist. Most of these algorithms' time complexity is $O(N^2 \times |A|)$ when the computing time of entropy and the positive region are taken into account (Qian *et al*. 2010, Chen *et al*. 2012, Liang *et al*. 2012). Clearly, the computing time of spatial reduct is closely related to m. When m is far less than N, the time complexity of Algorithm 1 is approximately $O(N \times |A|)$, which is much smaller than that of the CRS model. When m is very large, the calculation of spatial reduct approaches of the CRS model and is time-consuming.

However, from the experiment in Section 3.1.3, it can be found that continuously increasing parameter k, which determines m, first increases the classification accuracy, but decreases the classification accuracy gradually when k is so large that there exists spatial heterogeneity in the neighbors. In real-life applications, this parameter can be set using spatial association measures to calculate from which order there are no statistically

significant spatial associations of the decision attribute. An example is provided in Section 3.1.

Generally, m is small compared with N in most real-life applications. Even if m is too large to compute the spatial reduct in a reasonable amount of time, to balance the computation time and accuracy, a small m can be used, and a small m can also increase the classification accuracy according to our experimental results.

Spatial information system $SIS = \{U, A \cup \{d\}, M(k)\}$ Spatial reduct *red* FindSpatialReduct$U, A \cup \{d\}, M(k)$ *red* $= \emptyset$ each $x \in U$ *red* does not contain an x-local reduct of *SIS*. Add attributes of $A - red$ to ensure *red* contains an x-local reduct of *SIS red*

Algorithm 1 Calculating the spatial reduct for an SIS

Input: Spatial information system *SIS* = {*U, A*∪{*d*}, *M*(*K*)}
Output: Spatial reduct *red*
function FINDSPATIAL REDUCT (*U, A*∪{*d*}, *M*(*K*))
red ⊘
for each *x* ∈ U **do**
if *red* does not contain an *x*-local reduct of *SIS* **then**
Add attributes of *A*–*red* to ensure *red* contains an *x*-local reduct of *SIS*
end if
end for
return *red*
end function

2.3. *Rule matching for an SIS*

In the CRS, rules can be extracted by reading off the information system using the reduct *red*. The rules extracted are of the form *Pre* ⇒ *Succ*, where *Pre* is called the predecessor and is of the form $\wedge_{a \in red} a(x) = *$, and *Succ* is called the successor and is a value in the domain of d, where $*$ is a value in the domain of a. The confidence of a rule is defined as

$$conf(red(x) \Rightarrow d(x)) = \frac{|\{x \in U | \wedge_{a \in red} a(x) = * \wedge d(x) = Succ\}|}{|\{x \in U | \wedge_{a \in red} a(x) = *\}|}.$$

Considering x_0 in the SIS in Figure 2 as an example, the rules extracted from x_0 using the reduct $a1, a2$ is $a1(x) = 1 \wedge a2(x) = 2 \Rightarrow 1$. The confidence of the rule is 0.5 because only half of the objects that match the predecessor in U have decision 1.

These rules form a rule set and can be used to predict or classify future unseen objects. This process does not take spatial autocorrelation into account. However, spatial autocorrelation has been proven to be effective in improving classification accuracy (Gong 1992, Allard 1998, Ge and Bai 2011).

Based on the SV (∅Hrn 1999) method, we propose a spatial rule matching method, that is, spatial SV(SSV), for the SIS. To take into account spatial autocorrelation, when there are unsolvable conflicts or no rules can match the unseen object, the rule matching process uses the neighbors of the unseen object to determine its decision

value. The rule set constructed using the reduct is denoted by *RUL*. For unseen object x, the voting process is as follows:

(1) Set *RUL* is scanned for rules that match the conditional attributes' values of x. These rules constitute a set that is denoted by $RUL(x) \subseteq RUL$.
(2) If $RUL(x) = \emptyset$, then $RUL(x)$ can be constructed using the rule of which antecedents are most similar to those of x (Slowiński 1993). If $RUL(x)$ is still empty, then the decision value that dominates $C(x)$ is used as the decision value of x.
(3) An election is performed for $RUL(x)$ to resolve conflicts and rank all decisions for x. The election process proceeds as follows:
(a) Each rule $r \in RUL(x)$ casts one vote for its decision $\beta \in V_d$ and the vote is denoted by $votes(r)$.
(b) Compute normalization factor $norm(x) = \sum_{r \in RUL(x)} votes(r)$.
(c) Calculate certainty coefficient $certainty(x, \beta)$ for each possible decision $\beta \in V_d$ of x using the following three equations:

$$R_\beta = \{r \in RUL(x) | r predicts \beta\} \tag{1}$$

$$votes(\beta) = \sum_{r \in R_\beta} votes(r) \tag{2}$$

$$certainty(x, \beta) = votes(\beta)/norm(x) \tag{3}$$

(4) If $|\arg\max_\beta (certainty(x, \beta))| = 1$, then the decision value for x is set to $\arg\max_\beta (certainty(x, \beta))$. Otherwise, the decision value that dominates $C(x)$ is used as the decision value of x.
(5) If the decision value of x cannot be determined using $C(x)$, then its decision value is set to the decision value that dominates the entire training set. If there is no dominant decision value in the training set, then x is labeled as 'Undefined'.

For example, to classify x_1 in Figure 2, $Rul(x_1)$ includes $F1 : a1(x) = 1 \wedge a2(x) = 2 \Rightarrow 1$ and $F2 : a1(x) = 1 \wedge a2(x) = 2 \Rightarrow 2$. If the confidences of the rules are used as *votes* of each rule, then $norm(x_1) = conf(F1) + conf(F2) = 1$. Consequently, $R_1 = \{F1\}$ and $R_2 = \{F2\}$, and $votes(1) = votes(2) = 0.5$. Accordingly, $certainty(x_1, 1) = certainty(x_1, 2) = 0.5$, which means that there is a conflict between rules. Because $k = 1$, $C(x_1)$ contains all the first-order neighbors of x_1. The dominant category in $C(x_1)$ is 1. Therefore, x_1 is labeled as 1. If SV is used, then x_1 is assigned to 'Undefined' because there is no single category that dominates the entire study area. Clearly, SSV can correctly classify x_2, whereas SV cannot.

SSV has two major parts: selecting candidate rules and electing the best rule. To select candidate rules, the rule set is scanned once, and the time complexity is $O(|RUL|)$. To elect the best rule, the candidate rule set needs to be scanned once, and its time complexity is $O(|RUL(x)|)$. Finally, to find the dominant category in $C(x)$, $|C(x)|$ needs to be scanned once, and its time complexity is $O(|C(x)|)$. To summarize, the time complexity of SSV is $O(|RUL| + |RUL(x)| + |C(x)|)$. The time complexity of SV is $O(|RUL| + |RUL(x)|$. In many real-life applications, $|C(x)|$ is very small, in which case the time complexities of SV and SSV are almost the same.

2.4. *Comparison with classical sets and some of its extensions*

Similar to the CRS model and some of its extensions, the spatial rough set model proposed a new binary relation between objects according to the needs of real-life applications. Based on the new binary relation, the knowledge granule, approximations of target concepts and positive regions are formally defined. Finally, the reduct calculated can be used to perform feature selection, and the rule extraction and matching can be used to perform classification.

However, there are also some improvements that make the spatial rough set model more suitable for analyzing spatial datasets. For example, because the binary relation, that is, the x-local indiscernible relation, between objects not only relates to the attributes of the two objects but also is influenced by which x's neighboring area is currently under investigation, two objects may have different binary relations with respect to different x. Consequently, the subsequent concepts $LInd(y, x, A)$, and the lower and upper approximations are closely related to the geographical object x under inspection. Therefore, spatial rough sets approximate the target concept at each local region, which can preserve the local variability information or the spatial autocorrelation and heterogeneity information of geographical phenomena over space.

By contrast, CRS theory and its extensions calculate the binary relation for two objects without considering their locations. For example, CRSs and probability rough sets (Yang *et al*. 2017) calculate the relations between objects using the equivalence of attributes. Rough sets for incomplete information systems (Leung and Li 2003) calculate the relations between objects using the tolerance relation between objects. Although composite rough sets (Zhang *et al*. 2014a, 2016) can consider several binary relations in approximating target concepts, they also do not take into account the locations of objects. Moreover, the locations of objects are also neglected in the concept of the basic knowledge granule, and lower and upper approximations. Clearly, local variability information or spatial autocorrelation and heterogeneity information, which are important for modeling geographical phenomena, are neglected. Therefore, a spatial rough set model can better reflect the reality of geographical phenomena than other rough sets extensions.

3. Experiments and discussion

To further validate the effectiveness of the proposed spatial reduct and spatial rule matching method, two real-world examples are provided. The SPR and SSV methods were used in the analysis of NTDs in Heshun, Shanxi, China. The SAR was used to select relevant features for the segmentation of high-resolution remotely sensed imagery. The experiments showed that the method proposed could be successfully applied in the analysis of spatial data. All the related algorithms were implemented using Python. The experiments were run using a computer with 8GB RAM and an Intel Core i7 4790 CPU. The operating system was Ubuntu 16.04.

3.1. *Prediction of the occurrence of NTDs*

NTD data have been collected over several years and analyzed in many previous studies (Wu *et al*. 2004, Liao *et al*. 2009a, 2009b, Bai *et al*. 2010, 2014, Wang *et al*. 2010b). In the

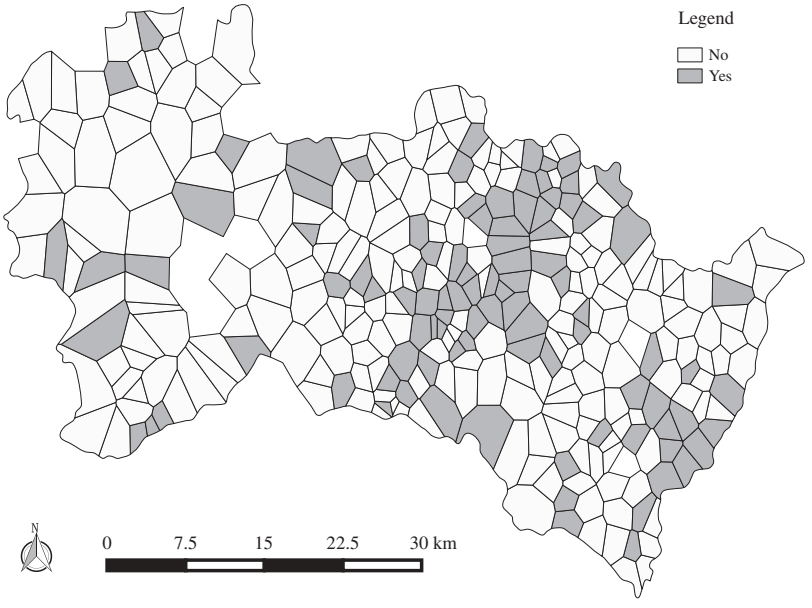


Figure 3. Map of the occurrence of NTDs in Heshun, Shanxi, China.

study area, there were 322 villages and 1 town. The locations of the 322 villages were determined by the Geographical Information System for spatial analysis. The data were collected using our field survey. The research project was approved by the Ministry of Science and Technology of the People's Republic of China. The study used only local statistical data. There were no experimental works or ethical issues. Because there were no boundaries defined for the villages, we drew them for each village using a Voronoi polygon (see Figure 3). All the villages that did not have new babies from 1998 to 2003 are not included in the figure and the following experiment.

Each village had 14 conditional attributes and 1 decision attribute. The conditional attributes were gross domestic product per capita, number of children born, number of children with NTDs, fertilizer used in the area (Fertilizer), access to a doctor (Doctor), production of fruit (Fruit), production of vegetables (Vegetables), 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*. (2010b) and Bai *et al*. (2010). The data were transformed to an information system and discretized in the same manner as that of Bai *et al*. (2010). A detailed description of the NTD data can be found in Wang *et al*. (2010b). The decision attribute was whether there were NTD instances in a village. If there were NTD instances, then the village was labeled 'Yes'; otherwise, the village was labeled 'No'. The villages and all the attributes used constituted the information system part of the SIS for the study area. $M(k)$ of the SIS was constructed using the first-order to k-th order adjacency matrices. In our experiment, two villages were mutually first-order neighbors if their boundaries touched.

To determine the maximum order of adjacency (parameter k for the SIS) that the SIS should take into account, the spatial autocorrelation of all villages in the study area was

Table 1. First- to fifth-order JCSs for Heshun, Shanxi, China.

k	$O_{01} - E_{01}$	$O_{00} - E_{00}$	$O_{11} - E_{11}$
1	− 49	477[a]	139
2	− 77	833[a]	241
3	− 57[a]	1240[a]	355
4	− 2[a]	1664[a]	484[a]
5	35[a]	2209[a]	589[a]

0 represents 'No' and 1 represents 'Yes'; 00 represents joins between two villages both labeled 'No'; 11 represents joins between two villages both labeled 'Yes'; 01 represents joins between two villages labeled 'No' and 'Yes'; O denotes the observed value; and E denotes the expected value under the assumption of completely spatial random distribution.

[a]Failed to pass the permutation test.

analyzed using JCS (Cliff and Ord 1970). Table 1 shows the overall spatial autocorrelation and spatial autocorrelation for each category. It can be found that there was at least one spatial autocorrelation index that was statistically significant when the order of adjacency was smaller than 4 and the significance level was set to 0.01. When the order of adjacency was 4, the *p*-value for the spatial autocorrelation of 'Yes' was still very small, that is, 0.016. Moreover, the number of fifth-order *rs* joins was larger than that from a completely spatially random distribution. This means that there may exist spatial heterogeneity at this order of adjacency, although it is not statistically significant. Accordingly, the order of adjacency was set to 4.

3.1.1. *Comparison with the CRS model*

Three models were compared to show the effectiveness of the spatial rough set model. The first was the CRS (Bai *et al*. 2010) model. In this model, the PR was used to select features. Then, rules were extracted from the information system using selected features. Finally, SV was used to classify unseen objects. The second model replaced the PR with the SPR in the CRS. Additionally, the other steps followed the same approach as those of the CRS model. The third model used the SPR to perform the feature selection task and SSV method to classify unseen objects. The rule extraction step of the third model simply followed the same approach as those of the CRS model. For simplicity, these three models are denoted by 'PR + SV', 'SPR + SV' and 'SPR + SSV', respectively.

To avoid the influence of the selection of training data, a comparison experiment was performed on the NTD dataset 1000 times, and the average accuracy assessment indices were used to compare the effectiveness of different models. In each round, half of the villages were randomly drawn as training data. All other villages were used as validation data. All three models used the training data to extract rules. Then, the extracted rules were used to classify the villages in the validation data.

During the classification process, SV and SSV used different approaches to manage conflicts between rules and the situation in which there were no rules matching the unseen objects. For example, if two rules 'watershedid = 0 and gradient $< 8^\circ \Rightarrow$ Yes' and 'watershedid = 0 and gradient $< 8^\circ \Rightarrow$ No' match the predecessor of an unseen object and have the same confidence 0.5, then there is a conflict between these two rules. In SV, the unseen object was classified as the category that dominated the entire study area. In SSV, the unseen object was classified as the category that dominated the

Table 2. Accuracy assessment results of the three models.

Models	Average overall accuracy	Average precision of 'Yes'	Average recall of 'Yes'	Average precision of 'No'	Average recall of 'No'
1. PR + SV	0.5989	0.4523	0.2936	0.7319	0.7388
2. SPR + SV	0.6149	0.4816	0.2836	0.7364	0.7658
3. SPR + SSV	0.6814	0.4938	0.2892	0.7278	0.8609

The first line is the average accuracy of results from the model using the PR and SV, that is, the CRS model; the second line is the average accuracy of results from the model using the SPR and SV methods; and the third line is the average accuracy of results from the model using the SPR and SSV methods.

neighbors of the unseen object to take advantage of the spatial autocorrelation and heterogeneity information.

A confusion matrix (Wikipedia 2017) was used to access the classification accuracy. The average overall accuracy, average precision of 'Yes', average recall of 'Yes', average precision of 'No' and average recall of 'No' for all 1000 experiments are shown in Table 2. The third model had the highest average overall accuracy, average precision of 'Yes' and average recall of 'No' among all three models. Although the average recall of 'Yes' and average precision of 'No' decreased by less than 0.005, the other indices increased by at least 0.03 compared with the CRS. The average overall accuracy and average recall of 'No' increased by 0.0825 and 0.1221, respectively. This means that the combination of SPR and SSV greatly increased the classification accuracy. Meanwhile, even if only SPR was used, that is, the second model, the spatial rough set-based model increased the classification accuracy. Compared with the CRS, only the average recall of 'Yes' decreased by 0.01. However, all the other indices increased. This means that the attribute subset selected by the SPR outperformed the PR in the classification of unseen objects. Accordingly, the consideration of spatial heterogeneity and spatial autocorrelation increased the classification accuracy for spatial data.

3.1.2. *Comparison with commonly used feature selection methods and classification methods*

To show the effectiveness of the SPR, we empirically compared it with five representative filter-based feature selection methods: Las Vegas filter (LVF) (Dash and Liu 2003), combination of weakest components (CWC) (Shin *et al*. 2011), INTERACT (Zhao and Liu 2007), correlation-based feature selection (CFS) (Hall 2000) and RelieF (Liu *et al*. 2004) The search time for LVF was set to 5000. The moderate denoising algorithm was used for the CWC algorithm. The δ parameter of the INTERACT algorithm was set to zero. The sample size of the RelieF algorithm was set to 10% of the number of objects.

As different classifiers have different induction biases, different classifiers, including a support vector machine (SVM), classification and regression tree (CART) and *k*-nearest neighbor (*k*NN) ($k = 3$), were used to maintain fairness between the classifiers. We used the implementations in the 'scikit-learn' package (0.10.1 version) in Python for the four classifiers. All algorithms used the default parameters except *k*NN, which used the three nearest neighbors. The average accuracy assessment results of 1000 runs of the model that combined different classifiers and different feature selection methods are shown in Table 3.

Table 3. Accuracy assessment of the classification results using different feature selection methods and different classifiers.

Classifier	Feature selection method	Average overall accuracy	Average precision of 'Yes'	Average recall of 'Yes'	Average precision of 'No'	Average recall of 'No'
SVM	PR	0.6876	**0.0939**	**0.0832**	0.6339	0.9195
	SPR	0.6906	**0.1569**	**0.0950**	0.6452	0.9168
	LVF	0.6925	**0.1469**	**0.1069**	0.6443	0.9159
	RelieF	0.6907	**0.1560**	**0.0818**	0.6567	0.7708
	CWC	0.6883	**0.1052**	**0.0771**	0.6434	0.9267
	INTERACT	0.6867	**0.0803**	**0.0542**	0.6547	0.9486
CART	PR	0.6242	0.4239	0.4095	0.6989	0.7093
	SPR	0.6441	0.4515	0.3913	0.7038	0.7469
	LVF	0.6879	0.3372	**0.1383**	0.6666	0.9001
	RelieF	0.6509	0.4437	0.3307	0.7031	0.7872
	CWC	0.6266	0.4305	0.4316	0.7070	0.7048
	INTERACT	0.6286	0.4274	0.4270	0.7151	0.7136
3NN	PR	0.6214	0.4209	0.4113	0.6975	0.7027
	SPR	0.6318	0.4329	0.4061	0.7014	0.7189
	LVF	0.5811	0.3599	0.3925	0.6645	0.6521
	RelieF	0.6257	0.4186	0.3963	0.7051	0.7203
	CWC	0.6305	0.4313	0.4086	0.7029	0.7191
	INTERACT	0.6309	0.4257	0.4003	0.7097	0.7278
SPR + SSV		0.6814	0.4938	0.2892	0.7278	0.8609

The bold font is only used for emphasize that these values are too small for practical use.

From Table 3, we found that, except for the averages of the recalls of 'Yes' for the CART and 3NN classifiers, and the average recall of 'No' for the SVM classifier that corresponded to the PR that decreased by less than 0.02 compared with that of the SPR, all the other 12 indices that corresponded to the SPR were larger than those of the PR. Accordingly, it is reasonable to state that the SPR outperformed the PR in the experiment.

Table 4 shows the comparison of the average overall classification accuracy between SPR and other feature selection methods. '>' ('<') indicates that the average overall classification accuracy of the feature selection method is greater (less) than that of the SPR. The Student's two-tailed *t*-test was used to test if significant difference existed between the overall classification accuracies. Obviously, among all 15 comparisons, there were 11 in which the overall classification accuracy of the SPR was significantly larger than that for other feature selection methods, whereas there were only two comparisons in which the overall classification accuracy of the SPR was significantly less than that for other feature selection methods. The remaining two comparisons were not statistically significant. From the comparison, it can be found that SPR outperformed the other feature selection methods in most cases.

Compared with other feature selection methods, SPR took spatial heterogeneity into account. It did not neglect the detail of the spatial variations, which is important for distinguishing different classes. Accordingly, SPR was superior to other feature selection methods regarding analyzing spatial data in our experiment. However, SPR did not take noise in data into account. In some cases, it was not as robust as expected. The reason that the LVF and RelieF, which were less sensitive to noise in data, outperformed SPR may be because they used random samples of objects to perform feature selection when the CART classifier was used.

Compared with classical classifiers, SSV took advantage of spatial association information. When there were contradictions in the classification process, it classified the current object into the category that dominated the neighboring area, while other classifiers labeled that object as the category that dominated the study area. Table 5 shows the comparison of the average overall classification accuracy between SSV and other

Table 4. Comparison of the average overall accuracies between SPR and the other feature selection methods.

Classifier	Feature selection method	SPR
SVM	PR	< (*p*-Value = 0.002)
	LVF	> (*p*-Value = **0.048**)
	RelieF	> (*p*-Value = **0.892**)
	CWC	< (*p*-Value = 0.010)
	INTERACT	< (*p*-Value = 0.000)
CART	PR	< (*p*-Value = 0.000)
	LVF	> (*p*-Value = 0.000)
	RelieF	> (*p*-Value = 0.000)
	CWC	< (*p*-Value = 0.000)
	INTERACT	< (*p*-Value = 0.000)
3NN	PR	< (*p*-Value = 0.000)
	LVF	< (*p*-Value = 0.000)
	RelieF	< (*p*-Value = 0.000)
	CWC	< (*p*-Value = 0.000)
	INTERACT	< (*p*-Value = 0.000)

p-Value reports the probability associated with the Student's paired two-tailed *t*-test.

Table 5. Comparison of the average overall accuracies between SPR + SSV and the other classifiers when SPR was used to perform feature selection.

	SV	SVM	CART	*k*NN
SPR + SSV	> (*p*-Value = 0.000)	< (*p*-Value = 0.000)	> (*p*-Value = 0.000)	> (*p*-Value = 0.000)

p-Value reports the probability associated with the Student's paired two-tailed *t*-test.

classifiers when SPR was used as the feature selection method. '>' ('<') indicates the average overall classification accuracy of SSV is greater (less) than that of the classifier used. The Student's two-tailed *t*-test was used to test if a significant difference existed between the overall classification accuracies. The accuracy of SSV was significantly larger than that of SV, CART and *k*NN. The average overall accuracy of SSV increased by at least 0.03 compared with SV, CART and *k*NN. With respect to the SVM, although the accuracy of SSV was significantly smaller, the average precision and recall of 'Yes' were approximately the half of those of SSV. The experiment results showed that SSV outperformed other classifiers with the help of spatial association information.

3.1.3. *Sensitivity analysis of the order of adjacency*

To analyze the relation between the order of adjacency and prediction accuracy of the third model, a comparison experiment for the prediction accuracy that corresponds to the first- to fifth-order adjacency SIS was performed. The third model was performed on different order adjacency SISs 1000 times. In each round, all villages were randomly divided into two equal parts: the training dataset and validation dataset. The average accuracy indices are summarized in Figure 4.

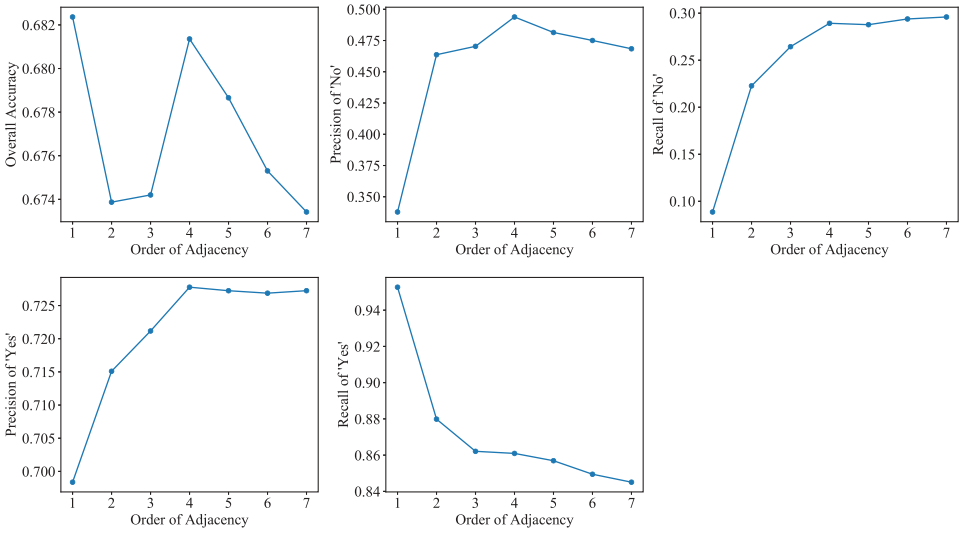


Figure 4. Accuracy assessment of the classification results for the SIS with parameter *k* from 1 to 7. All the accuracy indices are the average values of 1000 experiments. The *y*-axis is the value of the corresponding accuracy index and the *x*-axis is the order of neighbors (parameter *k*) used in classification.

When the order adjacency was one, the average overall accuracy was the highest compared with other orders of adjacency. However, the average precision of both 'Yes' and 'No', and the average recall of 'No' were the smallest compared with other orders of adjacency. Particularly, the average recall of 'No' was only 0.0886. Accordingly, the result that corresponded to first-order adjacency could not match the application needs. When the number of neighbors was too small, there were only a few neighboring villages that could be used to calculate the SPR and perform spatial rule matching. Accordingly, prediction accuracy might suffer from insufficient local training data. Therefore, the prediction accuracy might be too low to match practical needs.

After the first-order of adjacency, all other indices increased until the fourth-order of adjacency, although the average recall of 'Yes' decreased by less than 0.1. Particularly, the recall of 'No' increased by almost 0.2. When the order of adjacency was four, the average overall accuracy reached a local maximum. When the order of adjacency was larger than four, all indices decreased as the order of adjacency increased. From the experiment, it was clear that the prediction accuracy increased as an increasing amount of spatial autocorrelation information was considered when the order of adjacency increased. However, when the order of adjacency was too large, spatial heterogeneity had an effect. Accordingly, the consideration of too many neighboring villages may deteriorate the prediction accuracy.

3.2. *Segmentation of high-resolution remotely sensed imagery*

The automatic selection of textural features for the region merging segmentation of high-resolution remotely sensed imagery was used as an example to evaluate the effectiveness of SAR. A QuickBird remotely sensed image was used in the experiment. The upper-left latitude and longitude coordinates of this image were $116^\circ 3' 36.72'' E$ and $40^\circ 2' 40.90'' N$, and its lower-right latitude and longitude coordinates were $116^\circ 4' 3.83'' E$ and $40^\circ 2' 18.33'' N$, respectively. For simplicity, only the Panchromatic band of the image was used for segmentation. The spatial resolution of the panchromatic band was 0.63 m. The size of the image was 1000 pixels $\times$ 1000 pixels, and its panchromatic band is shown in Figure 5.

First, the Gabor filter (Dunn *et al*. 1994) and gray-level co-occurrence matrix (GLCM) (Haralick *et al*. 1973) were used to extract textural features from the Panchromatic band. The scale parameters of the Gabor method were set to $0.1, 0.2, \cdots, 0.9$, respectively. Additionally, the angle parameters of the Gabor method were set to $0, \pi/4, \cdots, 6\pi/4$, respectively. A total of 72 Gabor filter features were generated using different parameter combinations. When using the GLCM method, the window parameters were set to squares of length 5, 7 and 9. The numbers of directions were set to 4, 6 and 8, respectively. For each combination of the previous two parameters, the contrast, dissimilarity, homogeneity, angular second moment, energy and correlation GLCM features were calculated. A total of 108 GLCM textural features were generated using different combinations of parameters.

Next, an over-segmented initial segmentation of the high-resolution remotely sensed imagery was generated using the Quickshift algorithm (Vedaldi and Soatto 2008) based on the Panchromatic band of the image. A total of 67,854 regions were in the initial segmentation. For each region in the initial segmentation, the algorithmic mean of each spectral or textural feature was used as its feature value. For example, the average gray

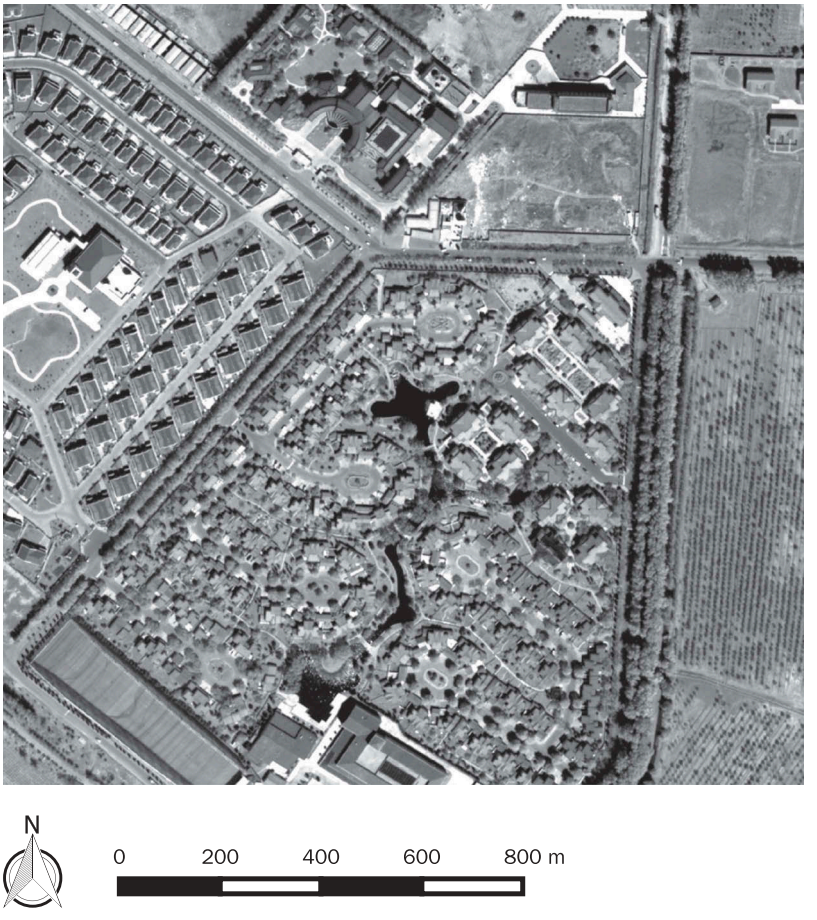


Figure 5. QuickBird remotely sensed imagery used in the segmentation experiment.

value of the Panchromatic band of all pixels in a region was used as the spectral feature value for the region. When all features were generated, we used the MIN-MAX normalization method to map all feature values to the interval $[0, 1]$, and all features were sorted in descending order by their variance.

Finally, an SIS was constructed for the initial segmentation of the image. Each region in the initial segmentation was an object in the SIS, and the normalized features were used as attributes for the SIS. Meanwhile, the parameter the maximum order of adjacency of the SIS was set to one because most region merging segmentation methods merge the current region with one or several its first-order neighbors in each merging step.

Before the target image was segmented, the SAR and AR were used to remove redundant textural features. All features were discretized into 100 intervals using the equal-width discretization method. The features selected by the SAR were the 78th and 80th textural features, and the calculating process took only 1.19 s. The features selected by the AR were the 74th, 78th, 79th and 80th textural features, and the calculating process took 5766.05 s (approximately 1.5 h). Because the AR took into account all

object pairs and the SAR only considered the pairs whose tails and heads were mutually first-order neighbors, the SAR took much less time than the AR.

To show the effectiveness of the SAR, the selected normalized textural features and normalized Panchromatic band were used to segment the target image. A typical region merging segmentation method, hierarchical step-wise optimization (HSWO) (Beaulieu and Goldberg 1989), was used to generate the final segmentation results based on the initial segmentation. To generate compact segments (Zhang *et al*. 2014b), the similarity between regions was calculated using the Euclidean distance and area of the two regions, that is, $sim(x,y) = D(x,y) \times area(x) \times area(y)$, where $D(x,y)$ is the Euclidean distance between objects x and y, and $area(x)$ is the area of object x. To avoid the influence of the difference between the attribute subset used in segmentation, HSWO stopped when the given number of regions remained. To validate the effectiveness of the SAR for different scales, we stopped the segmentation process when the numbers of remaining objects were 500, 1000, 1500, 2000, 2500, 3000, 3500 and 4000. The larger the number of remaining objects, the finer the scale of the segmentation result. The segmentation results of the target image using all features, features selected using the AR and features selected using the SAR when 500 regions remained, are shown in Figure 6(b–d).

Four indices (Liu *et al*. 2012) were adopted to measure the accuracy of the segmentation results: quality rate (QR), under-segmentation rate (UR), over-segmentation rate (OR) and composite index $ED1 = \sqrt{(OR^2 + UR^2)/2}$. QR is a goodness measure that evaluates the number of both 'miss' and 'hit' (Clinton *et al*. 2010). Zero values of QR and ED1 both indicate that the segmentation perfectly matches the reference polygons, that is, there are no over-segments or under-segments of the target image. A zero value for UR or OR indicates that there is no under-segmentation or over-segmentation, respectively. The maximum values of the four indices are all one, and indicate that the segmentation results are undesirable.

The manually segmented 80 reference polygons used in the accuracy assessment are shown in Figure 6(a). The four accuracy measures for all segmentation results that corresponded to different scales and different feature selection methods are shown in Figure 7. The *x*-axes of all subfigures in Figure 7 represent different scales. The *y*-axes of all subfigures in Figure 7 represent the value of the accuracy index. The blue, green and red lines in the figure represent the accuracy measures that correspond to the segmentation results using all textural features, textural features selected by the AR and textural features selected by the SAR, respectively.

From Figure 7(s), for almost all accuracy indices at all scales, the position of the red line is at the bottom of all three lines. This means that the segmentation accuracy of the results using textural features selected by the SAR was higher than that using the textural features selected by the AR and all textural features. The green line is between the blue line and red line in most cases. This means that the AR also improved the segmentation accuracy compared with results using all features. However, the AR took much more time (1.5 h versus almost 1 s) compared with the SAR in selecting features, and the corresponding segmentation accuracy was lower than that of the SAR at different scales. Accordingly, the SAR outperformed the AR in segmenting high-resolution remotely sensed imagery from the perspective of both effectiveness and efficiency.

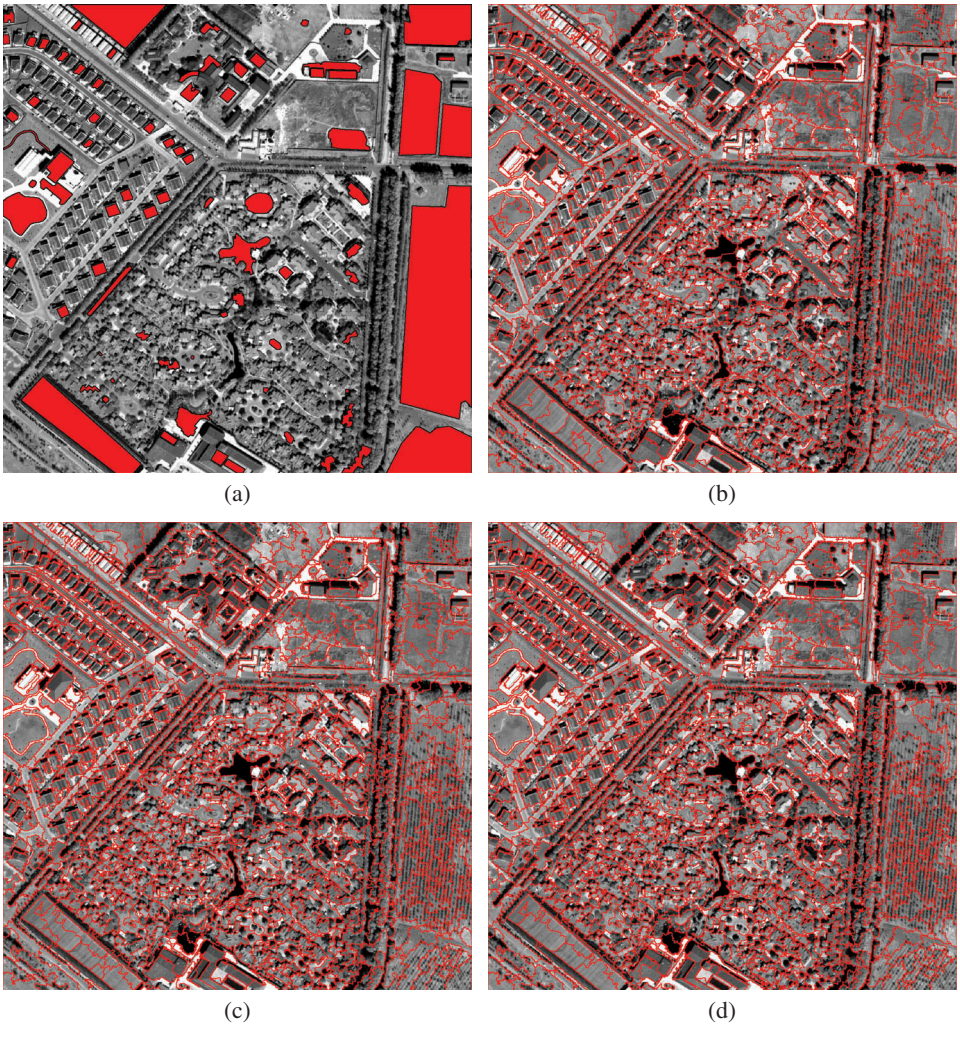


Figure 6. Segmentation results of the QuickBird images: (a) reference objects manually segmented from the remotely sensed imagery, (b) HSWO segmentation results using all features, (c) HSWO segmentation results using features selected by the AR and (d) HSWO segmentation results using features selected by the SAR.

4. Conclusion

This paper proposed a new rough set-based model for analyzing spatial data by taking into account spatial heterogeneity and autocorrelation. We proposed using a variant of equivalent classes that considered spatial heterogeneity, that is, local indiscernible sets, as basic knowledge granules to approximate target concepts. Based on this new knowledge granule, the target concept could be approximated using local upper approximation and local lower approximation at every *x*-centered region. We proposed using a spatial reduct to determine the minimal attribute set that could preserve the

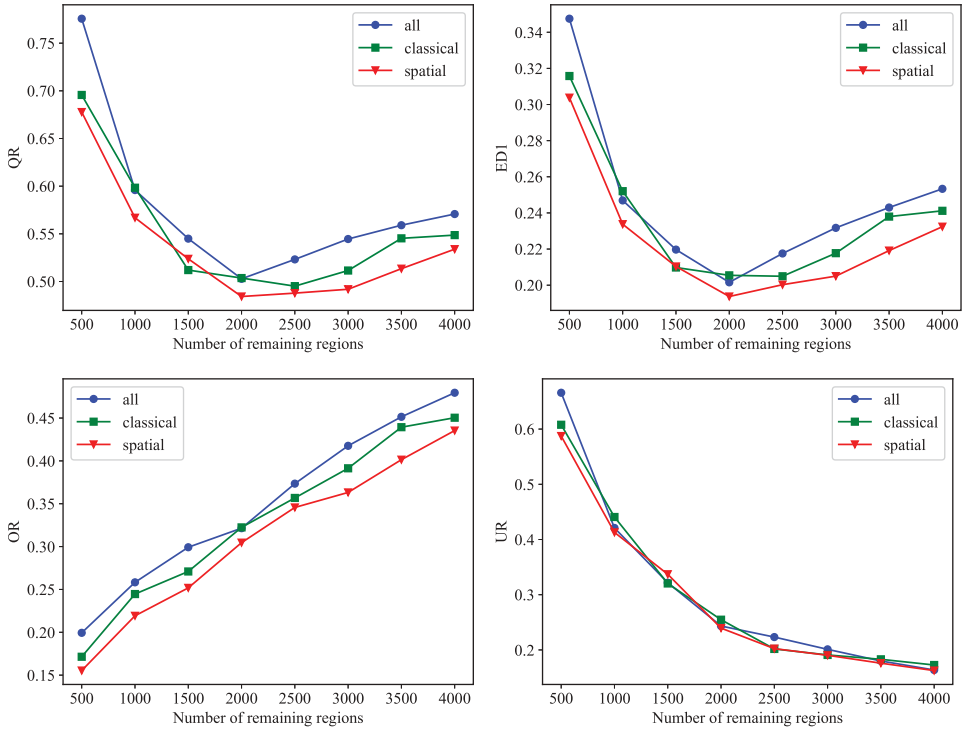


Figure 7. Comparison of the accuracy assessment of the segmentation results using different feature selection methods at different scales. The *y*-axis is the value of the accuracy assessment index and the *x*-axis is the number of remaining objects.

discriminant power for every *x*-centered region. Furthermore, the traditional SV rule matching method was improved to take into account spatial autocorrelation of objects.

There are two main advantages of the proposed rough set-based model. First, the model is more appropriate for analyzing spatial data than the CRS because it takes into account spatial heterogeneity and spatial autocorrelation. Compared with the CRS, both experiments showed that the classification accuracy or segmentation accuracy was improved using the proposed model. Second, a spatial reduct requires much less time than a classical reduct for large-scale spatial data. For example, in the second experiment, the SAR took only 1.19 s, whereas the AR took 1.5 h for the same dataset.

Not only does the CRS not take into account the spatial distribution of objects but also almost all rough set extensions, such as rough fuzzy sets, fuzzy rough sets and variable precision rough sets, share the assumption that all objects are completely randomly distributed over space by default. Accordingly, it is necessary to add spatial information into these extensions to model spatial data in the near future. This paper mainly discussed the basic concepts of spatial rough sets and showed how to apply this model to real-world spatial datasets. Little attention has been paid to the theoretical aspects of spatial rough sets. In our future work, we will further study the theoretical aspects of spatial rough sets and the potential application of these theoretical results, and apply this model to more real-world spatial data.

Furthermore, it is also important to extend the current approach to manage large-scale spatial data. The proposed model has the potential to be used to manage large spatial

datasets; however, there are still some difficulties that need to be addressed. For example, the calculating of the spatial reduct can be decomposed into two steps: calculation of *x*-local reducts and calculating of the minimal set cover. Because there are generally few neighboring objects of *x* and spatial rough sets concentrate only on the neighboring objects for each geographical object, the first step is not time-consuming and can be easily sped up using parallel computing resources. However, the second step is a typical set covering problem, which is an NP-hard issue and is time-consuming for large-scale datasets.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The work is supported by the Strategic Priority Research Program of the Chinese Academy of Science under Grant [XDA19040501], the National Natural Science Foundation of China under Grant [41871286], the National Natural Science Foundation for Distinguished Young Scholars of China under Grant [41725006] and the Natural Science Foundation of Shanxi Province of China under Grant [201701D121055].

Notes on contributors

Hexiang Bai is an associate professor in Computer Science at the school of computer and informatin technology, Shanxi University. His research interests include rough set based granular computing and spatial statistics. E-mail: baihx@sxu.edu.cn.

Deyu Li is a professor in Computer Science at the school of computer and informatin technology, Shanxi University. His research interests include rough set based granular computing and computational intelligence. E-mail: lidy@sxu.edu.cn.

Yong Ge is a professor in Geograhphical Information Sciences at the Institute of Geograhpic Sciences and Natural Resources Research, Chinese Academy of Science. Her research interests include spatial statistics spatial scale transformation. E-mail: gey@lreis.ac.cn.

Jinfeng Wang is a professor in Geograhphical Information Sciences at the Institute of Geograhpic Sciences and Natural Resources Research, Chinese Academy of Science. His research interest is spatial statistics. E-mail: wangjf@lreis.ac.cn.

References

∅Hrn, A., 1999. *Discernibility and rough sets in medicine: tools and applications*. Thesis (PhD). Norwegian University of Science and Technology, Norway.

Ahlqvist, O., 2005. Using uncertain conceptual spaces to translate between land cover categories. *International Journal of Geographical Information Science*, 19 (7), 831–857. doi:10.1080/13658810500106729

Ahlqvist, O., Keukelaar, J., and Oukbir, K., 2000. Rough classification and accuracy assessment. *International Journal of Geographical Information Science*, 14 (5), 475–496. doi:10.1080/13658810050057605

Ahlqvist, O., Keukelaar, J., and Oukbir, K., 2003. Rough and fuzzy geographical data integration. *International Journal of Geographical Information Science*, 17 (3), 223–234. doi:10.1080/13658810210157750

Albanese, A., Pal, S.K., and Petrosino, A., 2014. Rough sets, kernel set, and spatiotemporal outlier detection. *IEEE Transactions on Knowledge and Data Engineering*, 26 (1), 194–207. doi:10.1109/TKDE.2012.234

Allard, D., 1998. Geostatistical classification and class kriging. *Journal of Geographic Information and Decision Analysis*, 2 (2), 77–90.

Anselin, L., 1988. *Spatial econometrics: methods and models*. Dordrecht: Springer.

Anselin, L., 1992. *Spatial Data Analysis with GIS: An Introduction to Application in the Social Sciences (92-10)*. National Center for Geographic Information and Analysis.

Anselin, L., 1995. Local indicators of spatial association–LISA. *Geographical Analysis*, 27 (2), 93–115. doi:10.1111/j.1538-4632.1995.tb00338.x

Anselin, L., 2002. Under the hood: Issues in the specification and interpretation of spatial regression models, *Agricultural Economics*, 27 (3), 247–267. Spatial Analysis for Agricultural Economists: Concepts, Topics, Tools and Examples. doi:10.1111/j.1574-0862.2002.tb00120.x

Bai, H., *et al*., 2010. Using rough set theory to identify villages affected by birth defects: the example of Heshun, Shanxi, China. *International Journal of Geographical Information Science*, 24 (4), 559–576. doi:10.1080/13658810902960079

Bai, H., *et al*., 2014. A method for extracting rules from spatial data based on rough fuzzy sets. *Knowledge-Based Systems*, 57, 28–40. doi:10.1016/j.knosys.2013.12.008

Bai, H., *et al*., 2016. Detecting nominal variables' spatial associations using conditional probabilities of neighboring surface objects' categories. *Information Sciences*, 329, 701–718. Special issue on Discovery Science. doi:10.1016/j.ins.2015.10.003

Bai, H. and Ge, Y., 2009. Handling spatial-correlated attribute values in a rough set. *In*: O. Gervasi, *et al*., eds.. *Computational science and its applications iccsa 2009*. Lecture Notes in Computer Science. Vol. 5592. Springer: Berlin Heidelberg, 467–478.

Banu, S.K. and Tripathy, B.K., 2018, Neighborhood Rough-Sets-Based Spatial Data Analytics. *In*: D. Mehdi Khosrow-Pour, ed.. *Encyclopedia of Information Science and Technology*. *Fourth* ed. IGI Global. Pennsylvania: Hershey, 1835–1844. doi:10.4018/978-1-5225-22553.ch160

Beaubouef, T., Ladner, R., and Petry, F., 2004. Rough set spatial data modeling for data mining. *International Journal of Intelligent Systems*, 19 (7), 567–584. doi:10.1002/(ISSN)1098-111X

Beaulieu, J.M. and Goldberg, M., 1989. Hierarchy in picture segmentation: a stepwise optimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11 (2), 150–163. doi:10.1109/34.16711

Boots, B., 2003. Developing local measures of spatial association for categorical data. *Journal of Geographical Systems*, 5 (2), 139–160. doi:10.1007/s10109-003-0110-3

Chen, D., *et al*., 2012. Sample pair selection for attribute reduction with rough set. *IEEE Transactions on Knowledge and Data Engineering*, 24 (11), 2080–2093. doi:10.1109/TKDE.2011.89

Cliff, A.D. and Ord, J.K., 1970. Spatial autocorrelation: a review of existing and new measures with applications. *Economic Geography*, 46, 269–292. doi:10.2307/143144

Clinton, N., *et al*., 2010. Accuracy assessment measures for object-based image segmentation goodness. *Photogrammetric Engineering & Remote Sensing*, 76 (3), 289–299. doi:10.14358/PERS.76.3.289

Cressie, N., 1993. *Statistics for spatial data (revised edition)*. New York: WileyInterscience.

Dash, M. and Liu, H., 2003. Consistency-based search in feature selection. *Artificial Intelligence*, 151 (1âĂŞ2), 155–176. doi:10.1016/S0004-3702(03)00079-1

De Smith, M., Goodchild, M., and Longley, P., 2007. *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*. Leicester: Troubador Publishing Ltd.

Dubois, D. and Prade, H., 1990. Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems*, 17 (2–3), 191–209. doi:10.1080/03081079008935107

Dunn, D., Higgins, W.E., and Wakeley, J., 1994. Texture segmentation using 2-D gabor elementary functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16 (2), 130–149. doi:10.1109/34.273736

Fiedukowicz, A., 2015. Fuzzy Rough Sets Theory Reducts for Quantitative Decisions Approach for Spatial Data Generalization. *In*: M. Kryszkiewicz, S. Bandyopadhyay, H. Rybinski and S. Pal, eds. *PReMI 2015: Pattern Recognition and Machine Intelligence*. Springer International Publishing. 314–324.

Fisher, M. and Wang, J., 2011. *Spatial data analysis: Problems, techniques and applications*. Berlin: Springer.

Fortin, M. and Dale, M., 2005. *Spatial analysis: A guide for ecologists*. Cambridge, United Kingdom: Cambridge University Press.

Ge, Y., *et al*., 2012. Assessing the quality of training data in the supervised classification of remotely sensed imagery: a correlation analysis. *Journal of Spatial Science*, 57 (2), 135–152. doi:10.1080/14498596.2012.733616

Ge, Y. and Bai, H., 2011. Multiple-point simulation-based method for extraction of objects with spatial structure from remotely sensed imagery. *International Journal of Remote Sensing*, 32, 2311–2335. doi:10.1080/01431161003698278

Getis, A. and Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24 (3), 189–206. doi:10.1111/j.1538-4632.1992.tb00261.x

Gong, P., 1992. An efficient contextual classifier for land-use classification. *In*:*Geoscience and Remote Sensing Symposium, 1992. IGARSS '92*, Houston, Texas, USA. Institute of electrical and electronics engineers Inc PiscatawayNJ, 555–557.

Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*. Oxford, United Kingdom: Oxford University Press.

Griffith, D.A., 1989. Spatial econometrics: methods and models. *Economic Geography*, 65 (2), 160–162. doi:10.2307/143780

Guo, L., *et al*., 2013. Global and local indicators of spatial association between points and polygons: a study of land use change. *International Journal of Applied Earth Observation and Geoinformation*, 21 (4), 384–396. doi:10.1016/j.jag.2011.11.003

Haining, R., 2003. *Spatial data analysis: theory and practice*. Cambridge, United Kingdom: Cambridge University Press.

Haining, R.P., 1990. *Spatial data analysis in the social and environmental sciences*. Cambridge, United Kingdom: Cambridge University Press.

Hall, M., 2000. Correlation-based feature selection for discrete and numeric class machine learning. *In*:*Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)*. Morgan Kaufmann, 359–366.

Haralick, R.M., Shanmugam, K., and Dinstein, I., 1973. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC, 3 (6), 610–621. doi:10.1109/TSMC.1973.4309314

Jindal, S., 2017. *Rough Sets for Satellite Image Classification*. LAP LAMBERT. Saarbrücken, Germany: Academic Publishing.

Lei, T., Wan, S., and Chou, T., 2008. The comparison of pca and discrete rough set for feature extraction of remote sensing image classification: a case study on rice classification, taiwan. *Computational Geosciences*, 12, 1–14. doi:10.1007/s10596-007-9057-7

Leslie, T.F. and Kronenfeld, B.J., 2011. The colocation quotient: a new measure of spatial association between categorical subsets of points. *Geographical Analysis*, 43 (3), 306–326. doi:10.1111/j.1538-4632.2011.00821.x

Leung, Y., *et al*., 2007. A rough set approach to the discovery of classification rules in spatial data. *International Journal of Geographical Information Science*, 21 (9), 1033–1058. doi:10.1080/13658810601169915

Leung, Y. and Li, D., 2003. Maximal consistent block technique for rule acquisition in incomplete information systems. *Information Sciences*, 153, 85–106. doi:10.1016/S0020-0255(03)00061-6

Liang, J., *et al*., 2012. An efficient rough feature selection algorithm with a multigranulation view. *International Journal of Approximate Reasoning*, 53 (6), 912–926. doi:10.1016/j.ijar.2012.02.004

Liao, Y., *et al*., 2009a. Risk assessment of human neural tube defects using a bayesian belief network. *Stochastic Environmental Research and Risk Assessment*, 24 (1), 93–100. doi:10.1007/s00477-009-0303-5

Liao, Y., *et al*., 2009b. Identifying environmental risk factors for human neural tube defects before and after folic acid supplementation. *BMC Public Health*, 9, 391. doi:10.1186/1471-2458-9-391

Liu, H., Motoda, H., and Yu, L., 2004. A selective sampling approach to active feature selection. *Artificial Intelligence*, 159 (1–2), 49–74. doi:10.1016/j.artint.2004.05.009

Liu, J., *et al*., 2011. A rough set approach to analyze factors affecting landslide incidence. *Computers & Geosciences*, 37 (9), 1311–1317. doi:10.1016/j.cageo.2011.02.010

Liu, Y., *et al*., 2012. Discrepancy measures for selecting optimal combination of parameter values in object-based image analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 68, 144–156. doi:10.1016/j.isprsjprs.2012.01.007

Moran, P.A.P., 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10 (2), 243–251.

Murgante, B., Las Casas, G., and Sansone, A., 2008. A spatial rough set for extracting the periurban fringe. *Revue des nouvelles technologies de l'information*, 857, 101–125.

Páez, A., *et al*., 2012. Measuring ethnic clustering and exposure with the Q statistic: an exploratory analysis of Irish, Germans, and Yankees in 1880 Newark. *Annals of the Association of American Geographers. Association of American Geographers*, 102 (1), 84–102. doi:10.1080/00045608.2011.620502

Pal, S.K. and Mitra, P., 2002. Multispectral image segmentation using the rough-setinitialized EM algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 40, 2495–2501. doi:10.1109/TGRS.2002.803716

Patra, S., Modi, P., and Bruzzone, L., 2015. Hyperspectral band selection based on rough set. *IEEE Transactions on Geoscience and Remote Sensing*, 53, 5495–5503. doi:10.1109/TGRS.2015.2424236

Pawlak, Z., 1982. Rough sets. *International Journal of Computer & Information Sciences*, 11 (5), 341–356. doi:10.1007/BF01001956

Petry, F. and Elmore, P., 2015. Geospatial Uncertainty Representation: Fuzzy and Rough Set Approaches. *In*: D.E. Tamir, N.D. Rishe, and A. Kandel, eds.. *Fifty Years of Fuzzy Logic and its Applications*. No. 326 in Studies in Fuzziness and Soft Computing. Cham, ZG: Springer International Publishing, 483–497.

Qian, Y., *et al*., 2010. Positive approximation: an accelerator for attribute reduction in rough set theory. *Artificial Intelligence*, 174 (9–10), 597–618. doi:10.1016/j.artint.2010.04.018

Ripley, B.D., 1976. The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13 (2), 255–266. doi:10.2307/3212829

Sharmila Banu, K. and Tripathy, B.K., 2016. Rough set based similarity measures for data analytics in spatial epidemiology. *International Journal of Rough Sets and Data Analysis (IJRSDA)*, 3 (1), 114–123. doi:10.4018/IJRSDA.2016010107

Sheikhian, H., Delavar, M.R., and Stein, A., 2015. Uncertainty handling in disaster management using hierarchical rough set granulation. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3, 271–276. doi:10.5194/isprsannals-II-3-W5-271-2015

Shin, K., Fernandes, D., and Seiya, M., 2011. Consistency measures for feature selection: *a formal Definition, relative sensitivity comp*arison, and a fast Algorithm. *In*: *IJCAI 2011,Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, Barcelona, Catalonia, Spain. IJCAI/AAAI.

Slowiński, R., 15–18 June 1993. Rough set learning of preferential attitude in multi-criteria decision making. *In*: J. Komorowski and Z.W. Raś, eds. *Methodologies for intelligent systems: 7th international symposium, ismis'93 trondheim, norway, proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 642–651.

Ursin, B., 1983. Spatialfiltering of marine seismic data. *Geophysics*, 48, 1611. Available from: http://adsabs.harvard.edu/abs/1983Geop...48.1611U. doi:10.1190/1.1441443

Vedaldi, A. and Soatto, S., 2008. Quick shift and kernel methods for mode seeking. *In*: D. Forsyth, P. Torr, and A. Zisserman, eds. *Computer vision – eccv 2008: 10th european conference on computer vision, marseille, france, october 12-18, 2008, proceedings, part iv*. Berlin, Heidelberg: Springer Berlin Heidelberg, 705–718.

Wang, J., *et al*., 2002. Spatial sampling design for monitoring the area of cultivated land. *International Journal of Remote Sensing*, 23 (2), 263–284. doi:10.1080/01431160010025998

Wang, J., *et al*., 2010b. Assessing local determinants of neural tube defects in the Heshun Region, Shanxi Province, China. *BMC Public Health*, 10, 52. doi:10.1186/1471-2458-10-52

Wang, J.F., *et al*., 2010a. Geographical detectors–based health risk assessment and its application in the neural tube defects study of the heshun region, China. *International Journal of Geographical Information Science*, 24 (1), 107–127. doi:10.1080/13658810802443457

Wang, J.F., Zhang, T.L., and Fu, B.J., 2016. A measure of spatial stratified heterogeneity. *Ecological Indicators*, 67 (Supplement C), 250–256. doi:10.1016/j.ecolind.2016.02.052

Wikipedia, 2016. Spatial analysis — Wikipedia, the free encyclopedia. Available from: https://en.wikipedia.org/wiki/Spatial_analysis#Spatial_dependency_or_auto-correlation.

Wikipedia, 2017. Confusion matrix — Wikipedia, the free encyclopedia. Available from: https://en.wikipedia.org/wiki/Confusion_matrix.

Wu, J., *et al*., 2004. Exploratory spatial data analysis for the identification of risk factors to birth defects. *BMC Public Health*, 4, 23. doi:10.1186/1471-2458-4-23

Wu, W. and Zhang, W., 2002. Neighborhood operator systems and approximations. *Information Sciences*, 144 (1–4), 201–217. doi:10.1016/S0020-0255(02)00180-9

Yan, H.Y., *et al*., 2016. Spatial and temporal relation rule acquisition of eutrophication in Da'ning River based on rough set theory. *Ecological Indicators*, 66, 180–189. doi:10.1016/j.ecolind.2016.01.032

Yang, X., *et al*., 2017. A unified framework of dynamic three-way probabilistic rough sets. *Information Sciences*, 420, 126–147. doi:10.1016/j.ins.2017.08.053

Yun, O. and Ma, J., 2006. Land cover classification based on tolerant rough set. *International Journal of Remote Sensing*, 27 (14), 3041–3047. doi:10.1080/01431160600702368

Zhang, J., *et al*., 2016. Efficient parallel boolean matrix based algorithms for computing composite rough set approximations. *Information Sciences*, 329, 287–302. doi:10.1016/j.ins.2015.09.022

Zhang, J., Li, T., and Chen, H., 2014a. Composite rough sets for dynamic data mining. *Information Sciences*, 257, 81–100. doi:10.1016/j.ins.2013.08.016

Zhang, X., *et al*., 2014b. Hybrid region merging method for segmentation of highresolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 98, 19–28. doi:10.1016/j.isprsjprs.2014.09.011

Zhao, Z. and Liu, H., 2007. Searching for interacting features. *In*: *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence*, January, Hyderabad, India. IJCAI/AAAI.

Ziarko, W., 1993. Variable precision rough set model. *Journal of Computer and System Sciences*, 46 (1), 39–59. doi:10.1016/0022-0000(93)90048-2

Ziarko, W., 2005. Probabilistic rough sets. *In*: D. Ślęzak, *et al*., eds. *Rough sets, fuzzy sets, data mining, and granular computing: 10th international conference, rsfdgrc 2005, Regina, Canada, august 31 september 3, 2005, proceedings, part i*. Berlin, Heidelberg: Springer Berlin Heidelberg, 283–293.