

## Article

# An Approach for a Spatial Data Attribute Similarity Measure Based on Granular Computing Closeness

Weihua Liao <sup>1</sup>, Daizhong Hou <sup>2,\*</sup> and Weiguo Jiang <sup>3,\*</sup><sup>1</sup> College of Mathematics and Information Science, Guangxi University, Nanning 530004, China<sup>2</sup> School of Mathematics and Statistics, Nanning Normal University, Nanning 530001, China<sup>3</sup> State Key Laboratory of Remote Sensing Science, Beijing Normal University, Beijing 100875, China

\* Correspondence: lzhdz@163.com (D.H.); jiangweiguo@bnu.edu.cn (W.J.)

Received: 1 June 2019; Accepted: 25 June 2019; Published: 28 June 2019



**Abstract:** This paper proposes a spatial data attribute similarity measure method based on granular computing closeness. This method uses the distance and membership degree of different index levels of spatial entities to measure the similarity of attributes. It not only reflects the degree of similarity of spatial entity types at different index levels but also reflects the integration similarity between spatial entity types under a comprehensive index. This method embodies the layered idea of granular computing and can provide a basis for spatial problem decision making and for spatial entity classification. Finally, the feasibility and applicability of the method are verified by taking the similarity measure of the land-use type attribute in Guangxi as an example.

**Keywords:** closeness; granular computing; similarity measure; spatial data; land use

## 1. Introduction

Spatial similarity is an important theoretical issue of Geographic Information System (GIS), and it is necessary to conduct in-depth research [1]. Spatial problems can be explained by similar spatial phenomena, and the knowledge obtained from the research examples can be used to understand other phenomena, as well as to help explain certain scene phenomena [2,3]. Before the advent of computer graphics, the understanding of the law of spatial similarity was a mostly intuitive and subjective judgement. It did not achieve precision, quantification and formulation, and there was no strict mathematical model for the understanding of similar laws [4]. Similarity provides a basis for judging how to classify objects, form concepts, solve problems and integrate ideas [5].

Spatial similarity belongs to a type of spatial relationship that is juxtaposed with topological relations, directional relations, distance relationships and related relationships [6]. Due to the development of computer technology, scholars have carried out various research efforts in recent years on the basic theory and application of spatial data similarity. Because many similarity measure methods cannot measure the relationship between apertured polygons, scholars have proposed a method based on a positional map to describe the distribution of complex geometric shapes [7]. Scholars have developed and tested a spatial geometric consistency algorithm based on two spatial surface objects in the Euclidean plane for the continuous function of increasing the consistency of the position, orientation, size and shape of the spatial object [8]. A geometric similarity measure method for complex solid objects with surface holes was introduced to describe local and global features by constructing a multi-level bending radius complex function [9]. The abovementioned method is used to analyse spatial similarity in the geometry of spatial targets, and there is much similar literature regarding similar geometric shapes [10–12].

Spatial similarity, in addition to geometric similarities, is similar in structure to spatial objects (groups). It is often used to study the spatial distributions and layouts of geographic phenomena, as well

as to study the internal structures of geographic entities in geoscience research. The similarities of such structures are the basis for classification. Using the vegetation temperature condition index, scholars have applied structural similarity to analyse the spatial characteristics of the vegetation temperature condition index time series in plain areas and considered that structural factors are the main factors affecting the change [13]. The community structure and spatial distribution of macrobenthos in the Bering Sea shelf area were studied by scholars [14]. The scholars used the nearest-neighbour method to describe the K-function and box dimension of the spatial distributions of species and the spatial similarities, and verified the method with an experimental region [15]. To explore the insufficiency of habitat structure for the quantification of bird biodiversity, the researchers proposed a structural similarity method to determine the relationship between vertical and horizontal habitat structure indexes and bird biodiversity patterns and calculated the richness of all birds and of three specific bird species for each route of the habitat [16]. There are many papers on the distribution law of spatial entities using spatial structure similarity [17,18].

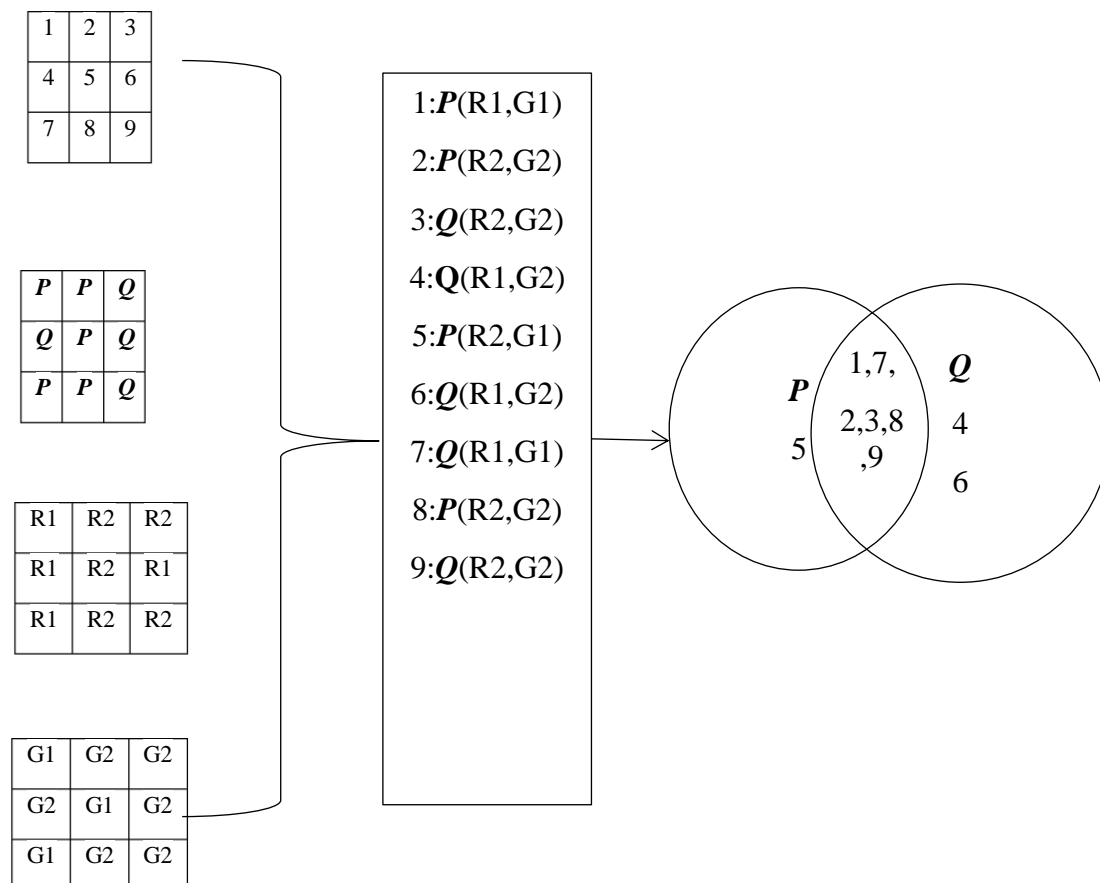
Due to the similarities between spatial entities in spatial relationships, spatial data based on spatial locations may also have similar feature attributes, and there is presently little theoretical or applied research on the similarity of feature attributes for spatial entities. Similarity is an uncertainty problem, and its measurement methods are mainly based on fuzzy sets [19], rough sets [20] and granular computing [21]. Similar theory descriptions and applied calculations require a foundation of mathematical methods, and it is necessary to integrate new methods on the basis of considering the nature of geoscience space. Granular computing, which was proposed in 1998 [22], is a new way of thinking about problems that comprise a superset of various methods in the field of artificial intelligence, such as rough sets, fuzzy sets, quotient spaces and cloud models [23]. Two kinds of distance and similarity metrics to measure granules were proposed for the searching and matching of granules, then researchers designed a granule classifier by the similarity metric [24]. Through the transformation of the traditional fuzzy set approximation axiomatization definition, a method to measure lattice closeness similarity reflecting the proximity between two particle sets has been proposed [25]. In addition to the geometric form, spatial entities have some natural–society–economic characteristics, and generally few different spatial types have identical characteristics [26]. These features constitute the attributes of spatial entities. Scholars have proposed a fast and effective two-step technique to measure similarity for the retrieval of spatio-temporal patterns, and used a database formed of 88 Landsat images to retrieve similar patterns of land cover and land use [27]. Characterizing and defining the similarities and closeness between different levels of different attributes of these spatial entities can be accomplished by means of the relevant definitions and the methods of granular computing. This paper will use granular computing closeness to calculate the similarity between spatial entities, introduce the granulation concept to study the spatial problem, expand the application field of granular computing and enrich the spatial data mining method system.

## 2. Prerequisite Knowledge

The similarity measure is the degree of closeness between two things. The closer the two things are, the larger their similarity measure value is, while the more distant the two things are, the smaller their similarity measure value is [28]. There are many different methods for measuring similarity, and they are generally selected according to actual problems. Common similarity measures have correlation coefficients and similarity coefficients. The correlation coefficient is used to measure the closeness between variables (e.g., the relationship between various counties' GDPs, population densities and other indexes within a provincial administrative division). Obviously, when calculating the correlation coefficient of two indexes, each index has the same number of samples.

To measure the proximity between spatial entity samples, similarity coefficients should be used, such as the measure between economic (GDP) and social (population density) indexes between different land types. The similarity coefficient is a quantitative description of the similarity between classification units, so the degree of similarity between these things must be described by quantitative methods.

Things often need to be characterized by multiple variables. For example, in Figure 1 there are nine land units divided into *P* and *Q* land types, each with a population density (represented by *R*, where *R*1 and *R*2 represent two values of population density) and GDP (represented by *G*, where *G*1 and *G*2 represent two values of GDP). How to measure the similarity of land use indexes between the *P* and *Q* land types in this region requires the introduction of similarity measure methods. The general similarity is measured based on different distances. The distance measure has a distance based on continuous index data, and there is also a set distance based on discrete data. Spatial entity attribute index discrete values are used, and a set distance is used in this study.



**Figure 1.** Spatial entity set distance solution process diagram.

Let  $X = \{x_1, x_2, \dots, x_n\}$  be a universe with two sets, *P* and *Q*,  $P, Q \in F(X)$ . The distance measure between the sets *P* and *Q* is defined by the following formula [29]:

$$D(P, Q) = 1 - \frac{|P \cap Q|}{|P \cup Q|} \quad (1)$$

The set distance reflects the degree of difference between the two sets. When  $P \neq Q$  and  $P \cap Q = \Phi$ , the difference between the sets *P* and *Q* is the largest, and the sets *P* and *Q* are completely different. Conversely, if  $P = Q$ , the sets *P* and *Q* are completely similar. The distance of the sets reflects the degree of difference between the sets caused by different relationships in the same universe. The distance formula shows that the fewer elements there are in the two sets of common intersections, the larger the distance is, while the more intersecting elements there are, the smaller the distance is. The *P* and *Q* land types represent two types of sets as in Figure 1, where *P* contains land units {1, 2, 5, 7, 8} and *Q* contains land units {3, 4, 6, 9}. Similarly, the population density indicator also has two sets, *R*1 {1, 4, 6, 7} and *R*2 {2, 3, 5, 8, 9}, and the GDP indicator has two sets, *G*1 {1, 5, 7} and *G*2 {2, 3, 4, 6, 8, 9}, so that

different land types composed of each land unit can be obtained, and different measure sets can be obtained. For example, the first unit is  $P \{R1, G2\}$ , which means that its land type is  $P$ , the population density index is R1, the GDP index is G2, and it is the same as the seventh land unit belonging to the set  $P$  and  $Q$  as the public part of the two indicators,  $P \cap Q$ . Details are shown in Figure 1. According to Formula (1), the set distances of the land types  $P$  and  $Q$  measured by population density and GDP in Figure 1 are  $D(P, Q) = 1 - \frac{||\{1,2,3,7,8,9\}||}{||\{1,2,3,4,5,6,7,8,9\}||} = 3/9$ .

The set distance describes the degree of difference between sets of spatial entities. Spatial similarity is the closeness relationship between sets, and it is complementary to distance; therefore, it is necessary to introduce the concept of the degree of set closeness. Given different spatial sets,  $P$  and  $Q$ , the definition of the closeness degree for sets  $P$  and  $Q$  is [30]:

$$n(P, Q) = \frac{|P \cap Q|}{|P \cup Q|}, 0 \leq n(P, Q) \leq 1 \quad (2)$$

In particular, if  $P = Q$ , the set closeness of  $P$  and  $Q$  reaches a maximum value of 1; that is,  $P$  and  $Q$  are completely close. If  $P \neq Q$  and  $P \cap Q \neq \emptyset$ , the set closeness of  $P$  and  $Q$  reaches the minimum value,  $P$  and  $Q$  are not close at all. It can be seen that the closeness of the collection reflects the degree of closeness between the sets and the common part between the two sets (i.e., the more intersecting parts there are, the closer the elements are, while the fewer intersecting parts there are, the closer the elements are). Thus, the closeness of  $P$  and  $Q$  in Figure 1 is  $n(P, Q) = \frac{||\{1,2,3,7,8,9\}||}{||\{1,2,3,4,5,6,7,8,9\}||} = 4/9$ .

### 3. Granular Computing Measure

The fuzzy set based on the set distance can only measure in one dimension; that is, it can only measure the closeness between the two spatial sets, with no way to measure the comprehensive closeness under complex mutual relationships. For example, there are 23 types of land use in an actual land use evaluation. Measuring the similarity coefficient between different land use types by using multi-dimensional economic indexes is a very difficult task. To overcome this deficiency, a multi-dimensional index closeness similarity measure based on granular computing is introduced. Granular computing is a new form of conceptual analysis and its basic idea is to use to solve hierarchical problems in the problem solving process. Granular computing first needs to granulate information, and there are many methods for doing so. This paper uses the atomic formula to granulate information. Let  $IS = (U, A)$  be an information system and  $U$  a universe, where  $a \in A$  represents the attribute in  $A$ , and  $v$  is the attribute value of  $a$  on the individual  $x$  on  $U$ ,  $v = a(x)$ . In this case,  $(a, v)$  or  $a_v$  is called an atomic formula in granular computing [31]. The land types  $P$  and  $Q$  have two measure indexes in Figure 1, R and G, which form the attribute set  $A$ . The value of each indicator (i.e., R1 and R2) is the value of the different land units in relation to the index R.  $(R, R1)$  or  $R = R1$  is an atomic formula. Thus, through each atomic formula, granular information is obtained, and the calculation of the atomic formula is the information granulation process. Different information granulation approaches can obtain different types of granular information, such as the index R, through the atomic formulas  $R = R1$  and  $R = R2$ , the land use area can be granulated into two information granules,  $\{1,4,6,7\}$  and  $\{2,3,5,8,9\}$ , while through the atomic formulas  $G = G1$  and  $G = G2$ , the land use area can be granulated into two information granules,  $\{1,5,7\}$  and  $\{2,3,4,6,8,9\}$ .

Different atomic formulas form different information particles, and how these information particles perform set operations requires some mathematical descriptions and logic rules. Atomic formulas can be described in the form of logical negations, conjunctions and disjunctions. By combining many of the classic logical conjunctions  $(a, v)$  or  $a_v$ , the formula for the calculation  $\psi$  of the granularity can be obtained. Let  $(\varphi, m(\varphi))$  and  $(\varphi', m(\varphi'))$  be two information granules, and the definition of the operation of logical connectives can be calculated using the following forms [31]:

- (1)  $\sim (\varphi, m(\sim \varphi)) = (\sim \varphi, U - m(\varphi))$ ;
- (2)  $(\varphi, m(\varphi)) \oplus (\varphi', m(\varphi')) = (\varphi \vee \varphi', m(\varphi) \cup m(\varphi'))$  ;.

- (3)  $(\varphi, m(\varphi)) \otimes (\varphi', m(\varphi')) = (\varphi \vee \varphi', m(\varphi) \cap m(\varphi'))$  ;
- (4)  $(\varphi, m(\varphi)) c(\varphi', m(\varphi')) = (\varphi \rightarrow \varphi', m(\varphi) \subseteq m(\varphi')) \vee m_*(\varphi) \subseteq m_*(\varphi') \wedge m^*(\varphi) \subseteq m^*(\varphi')$  ;
- (5)  $(\varphi, m(\varphi)) \equiv (\varphi', m(\varphi')) = (\varphi \leftrightarrow \varphi', m(\varphi) \subseteq m(\varphi')) \wedge m(\varphi') \subseteq m(\varphi) \vee ((m_*(\varphi) \subseteq m_*(\varphi')) \wedge m_*(\varphi') \subseteq m_*(\varphi)) \wedge m^*(\varphi) \subseteq m^*(\varphi') \wedge m^*(\varphi') \subseteq m^*(\varphi)$ ;

Feature transformation can be performed on multi-dimensional spatial types through the operation of logical conjunctions. Let  $U$  be a universe,  $\tilde{P} \subseteq U$ ; if the set has  $m$  vector objects,  $\tilde{P} = \{P_1, P_1, \dots, P_m\}$ ,  $P_k \in U (k = 1, 2, \dots, m)$ , and the feature information set of each vector object  $P_k = \{P_{k1}, P_{k2}, \dots, P_{ks}\}$  is an information granule, then  $s$  represents the feature information number of  $P_k$ . Some feature information of all vectors  $P_k$  forms a set by the value  $P'_k$  calculated by the formula  $\varphi$  [32]:

$$m_{\tilde{P}}(\varphi) = \{P'_k | P'_k \in \varphi, P_k \rightarrow P'_k, k = 1, 2, \dots, m\} \quad (3)$$

where  $m_{\tilde{P}}(\varphi)$  is a set of granules formed by the calculation of the set  $\tilde{P}$  through the formula  $\varphi$ ,  $P_k \rightarrow P'_k$  indicates that each object forms a one-to-one correspondence with its value and  $P'_k \in \varphi$  indicates the value obtained by the formula  $\varphi$  calculation from  $P_k$ .

There are two vectors formed by land types  $P$  and  $Q$  in Figure 1. The population density characteristics are R1 and R2, which means that the formula is  $R = R1$  or  $R = R2$ , with different sets of land types  $P$  and  $Q$  determined by two atomic formulas. The formula  $R = R1$  for land type  $P$  is  $\{1\}$ , and for land type  $Q$  is  $\{4,6,7\}$ ; the atomic formula  $R = R1$  gives the value  $\{\{1\}, \{4,6,7\}\}$  of land types  $P$  and  $Q$ , which is the value obtained by the formula  $R = R1$ . Similarly,  $R = R2$  for land type  $P$  is  $\{2,8\}$ , and for land type  $Q$  is  $\{3,5,9\}$ , while the value of land type  $P$  and  $Q$  obtained by the atomic formula  $R = R2$  is  $\{\{2,8\}, \{3,5,9\}\}$ . Obviously, Equation (3) can measure a single value of a single indicator, but it cannot address complex measurement relationships, such as simultaneously measuring atomic formulas  $R = R1$  and  $R = R2$ .

Since Equation (3) can only measure the similarity of single layers, it is necessary to redefine the formation of granular sets by atomic formulas at different levels to comprehensively measure the similarity of different spatial entity types at different levels of different indexes. For example, the similarity of different land types under different attribute values of high, medium or low GDPs, and high, medium or low population densities. Let  $U$  be a universe,  $\varphi_i$  is a collection of atomic formulas,  $\tilde{P} \subseteq U$ ,  $\tilde{P} = \{P_1, P_2, \dots, P_m\}$  is a set where  $P_k = \{p_{k1}, p_{k2}, \dots, p_{ks}\} (k = 1, 2, \dots, m)$ , and  $s$  indicates that each  $P_k$  contains  $s$  characteristic information; each vector object in the set is calculated by the formula and reaches the corresponding set of vector objects of the hierarchy, which is expressed as [32]:

$$m_{\tilde{P}}^j(\varphi_i) = \{P_k | P_k \in \tilde{P}, P_k \in \varphi_i, k = 1, 2, \dots, m, j = 1, 2, \dots, h\} \quad (4)$$

where  $h$  is the number of levels, such as high, medium or low levels of GDP, and the distance from a road is one of three levels: far, medium or near. All spatial entities are divided into different granular sets by atomic formulas with different index values according to Equation (4), and such granular sets are indistinguishable at different levels of the same indicator. It is assumed that there are 100 land units divided into three land types: S1, S2 and S3 (see Figure 2). There are two measurements of GDP and population density, and GDP is divided into three levels: G1, G2 and G3; while the population density is divided into three levels: P1, P2 and P3. Thus,  $\{S1, S2, S3\}$  constitutes  $\tilde{P}$ , where  $p_{ks}$  is composed of different sets on G1, G2, G3 from S1. To calculate the similarity, we need to define the membership of different atomic formulas in  $\tilde{P}$ .  $\tilde{P} = \{P_1, P_1, \dots, P_m\} \tilde{P} \in U, \forall \varphi_i \in \psi (i = 1, 2, \dots, r)$  is a vector object collection in universe  $U$ , The degree to which  $m_{\tilde{P}}^j(\varphi_i)$  belongs to a set is defined as [33]:

$$\mu_{\tilde{P}}^j(\varphi_i) = \frac{|m_{\tilde{P}}^j(\varphi_i) \cap \tilde{P}|}{|\tilde{P}|} \quad (i = 1, 2, \dots, r; j = 1, 2, \dots, h) \quad (5)$$

obviously,  $0 \leq \mu_{\tilde{P}}^j(\varphi_i) \leq 1$ ,  $\sum_{j=1}^h \mu_{\tilde{P}}^j(\varphi_i) = 1$ . Taking Figure 2 as an example, the membership degree is the ratio of the granular set formed by each land type under different index values to the number of land types. For example, if there are 20 land units in S1, when the atomic formula is GDP = G1 and there are six granular sets, then the atomic formula GDP = G1 forms a membership of the granular set of 0.3 (see Table 1). In the calculations of actual problems, the closeness calculation method based on the membership function has a distance method, a maximum and minimum method and an algebra and minimum method. The distance method is used in this study, and the distance between sets  $\tilde{P}$  and  $\tilde{Q}$  is defined as [33]:

$$d(\tilde{P}, \tilde{Q}) = \left( \sum_{j=1}^h \sum_{i=1}^r \left| u_{\tilde{P}}^j(\varphi_i) - u_{\tilde{Q}}^j(\varphi_i) \right|^\lambda \right)^{\frac{1}{\lambda}} \quad n(\tilde{P}, \tilde{Q}) = 1 - \frac{1}{(r \times h)} \left( \sum_{j=1}^h \sum_{i=1}^r \left| u_{\tilde{P}}^j(\varphi_i) - u_{\tilde{Q}}^j(\varphi_i) \right|^\lambda \right)^{\frac{1}{\lambda}} \quad (6)$$

where  $r$  is the number of types of entities,  $h$  is the number of levels of different indexes,  $\lambda$  is a formula for calculating distances,  $\lambda = 1$  is the closeness based on the Hamming distance and  $\lambda = 2$  is the closeness based on the Euclidean distance;  $\lambda = 1$  is used in this paper. Thus, according to Equation (6) for Figure 2, the similarity calculations of the following processes are performed separately.

$$d(s1, s2) = |0.3 - 0.21| + |0.6 - 0.6| + |0.1 - 0.19| + |0.2 - 0.16| + |0.55 - 0.67| + |0.25 - 0.18| = 0.41$$

$$n(s1, s2) = 1 - 0.41/6 = 0.932$$

$$d(s1, s3) = |0.3 - 0.26| + |0.6 - 0.52| + |0.1 - 0.22| + |0.2 - 0.13| + |0.55 - 0.65| + |0.25 - 0.22| = 0.44$$

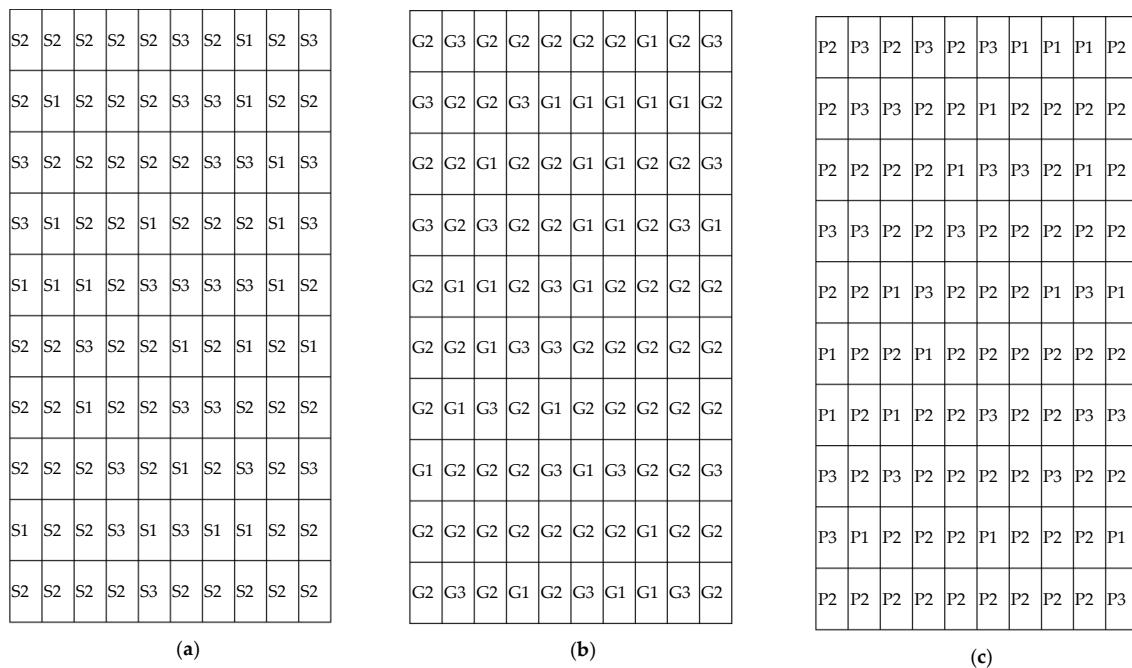
$$n(s1, s3) = 1 - 0.44/6 = 0.927$$

$$d(s2, s3) = |0.21 - 0.26| + |0.6 - 0.52| + |0.19 - 0.22| + |0.16 - 0.13| + |0.67 - 0.65| + |0.18 - 0.22| = 0.25$$

$$n(s2, s3) = 1 - 0.25/6 = 0.958$$

**Table 1.** Three spatial entity attribute membership degree comparison table.

		$\phi 1$			$\phi 2$		
		G1	G2	G3	P1	P2	P3
S1	$m_{\tilde{P}}^j(\varphi i)$	6	12	2	4	11	5
	$\mu_{\tilde{P}}^j(\varphi i)$	0.30	0.60	0.10	0.20	0.55	0.25
S2	$m_{\tilde{P}}^j(\varphi i)$	12	34	11	9	38	10
	$\mu_{\tilde{P}}^j(\varphi i)$	0.21	0.60	0.19	0.16	0.67	0.18
S3	$m_{\tilde{P}}^j(\varphi i)$	6	12	5	3	15	5
	$\mu_{\tilde{P}}^j(\varphi i)$	0.26	0.52	0.22	0.13	0.65	0.22



**Figure 2.** Spatial distribution of three land use types and spatial distribution of indexes.

Therefore,  $n(s2, s3) > n(s1, s2) > n(s1, s3)$ , and this shows that the distance between S2 and S3 is the smallest and the closeness is the largest (i.e., the similarity between GDP and population density is the largest between the two land types). The distance between S1 and S3 is the largest and the closeness is the smallest; that is, the similarity between GDP and population density indexes is the smallest between the two land types. Similarity value generator is designed in this study, and the process of the algorithm is shown in Algorithm 1.

---

**Algorithm 1** Similarity Value Generator

---

Input	membership matrix
Output	Similarity matrix
1	function B=simi(R)
2	[n,k]→size(R);
3	B←zeros(n);
4	for i = 1→n
5	for j = i→n
6	for p = 1→k
7	Calculating distance B(i,j);
8	end
9	B(i,j)←1-B(i,j)/k;
10	B(j,i)←B(i,j);
11	end
12	end

---

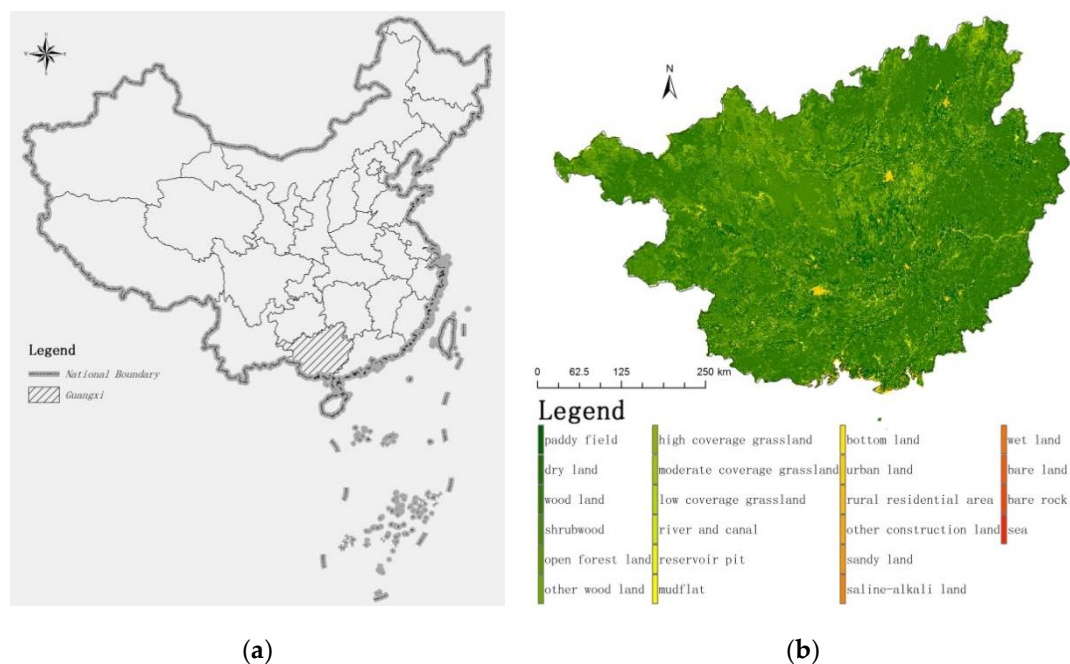
## 4. Case Study

### 4.1. Research Area and Data Sources

Guangxi is a sea passage in southwestern China, with the advantage being located along the border, along the coast and along the Yangtze River. Guangxi has a complete array of land types, with high and low landforms including mountains, plateaus, middle mountains, hills, terraces and plains.



Having entered a new era of industrialization and rapid economic development in Guangxi, land use has changed dramatically. This paper will use the 2010 data on land use in Guangxi (Figure 3b), the Guangxi GDP kilometre network data in 2010 (Figure 4a), the data of the Guangxi population density kilometre network in 2010 (Figure 4b), the data of Guangxi DMSP/OLS in 2010 (Figure 4c), Guangxi Digital Elevation Model (DEM) data (Figure 4d) and Guangxi road data (Figure 4e), using the method proposed in this paper to verify the spatial similarity of different land types in Guangxi. The road data source of this study is the open street map (<https://www.openstreetmap.org>), and other data are from the data collection of the Resource and Environmental Science Data Center of the Chinese Academy of Sciences (<http://www.resdc.cn/Default.aspx>). There were seven types of first-level land use and 22 types of second-level land use in Guangxi in 2010, where cultivated land, forest land and grassland areas dominated, and the proportion of unused land and sea areas was small, as is seen in Table 2.

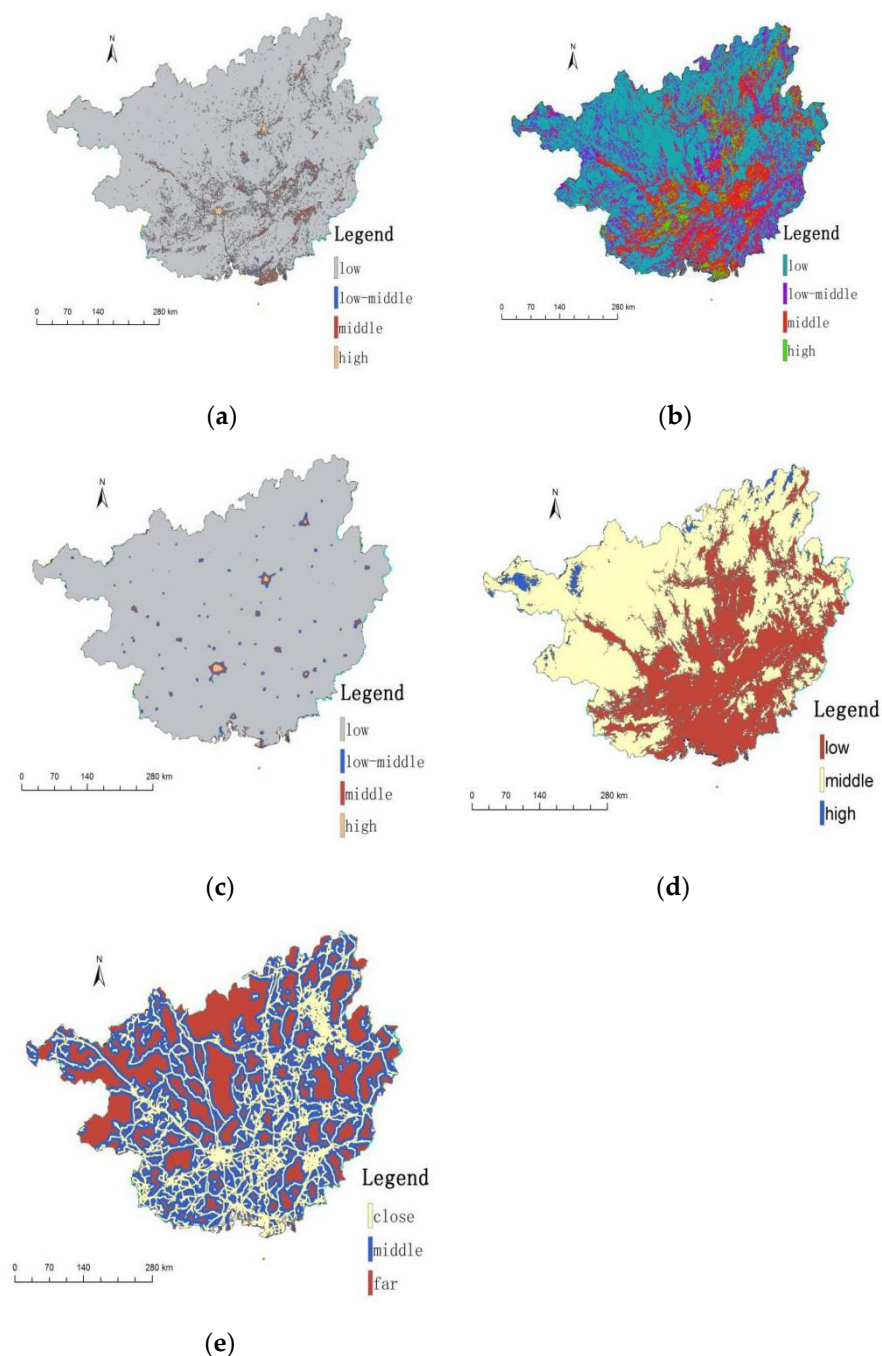


**Figure 3.** Location of Guangxi (a) and the spatial distribution map of land use types (b) in Guangxi 2010.

**Table 2.** Land use type area in 2010 (unit: km<sup>2</sup>).

Land Use Type	Area	Land Use Type	Area
paddy field	25,131	mudflat	316
dryland	26,409	bottom land	294
woodland	85,513	urban land	986
shrubwood	36,614	rural residential area	3386
open forest land	28,539	other construction land	349
other woodland	5094	sandy land	7
high-coverage grassland	17,699	saline-alkali land	3
moderate coverage grassland	2890	wet land	2
low-coverage grassland	108	bare land	14
river and canal	1622	bare rock	11
reservoir pit	1702	sea	28



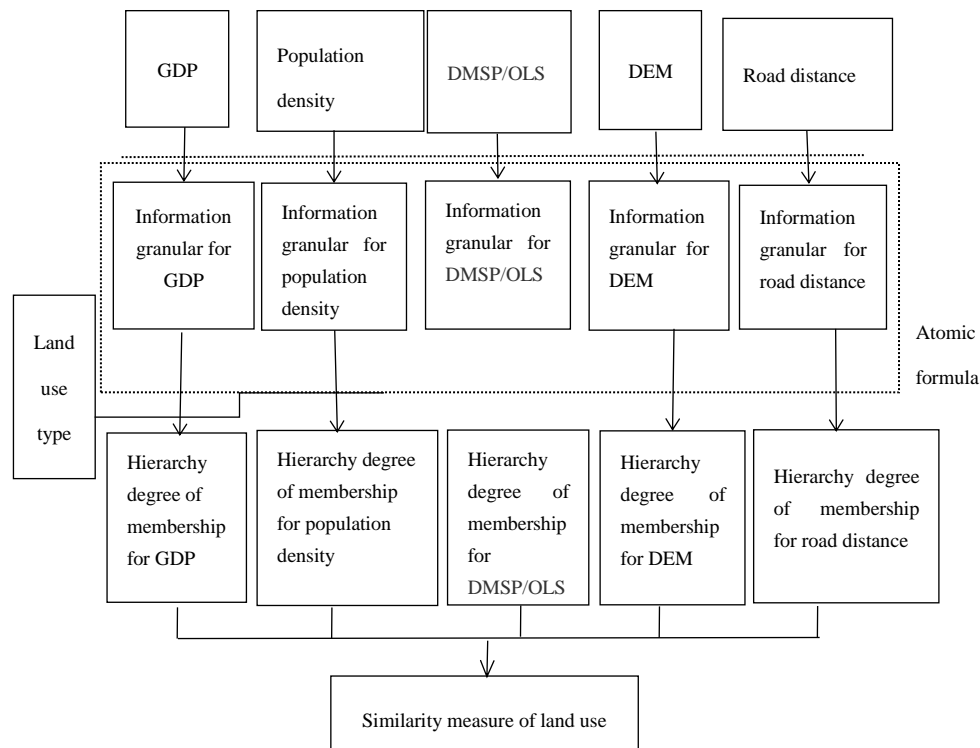


**Figure 4.** Guangxi GDP (a); population density (b); lighting data (c); DEM (d); road distance (e) spatial distribution map.

#### 4.2. Computing Framework

Using the granular similarity measure method proposed in this paper combined with the 22 land use change data in Guangxi and the spatial distribution data of five socio-economic attribute indexes, we designed the similarity measure of the land use change type attribute computing framework in Guangxi (Figure 5). The five socio-economic indexes are continuous data, and they were separately resampled and discretized to obtain spatially discrete data of the five indexes to form hierarchical data of the index sets. Then, different values of different indexes were used as atomic formulas to form the granular information set of each index, and the number of statistical indexes at different levels in the granular information concentration was calculated by determining the index membership degree for

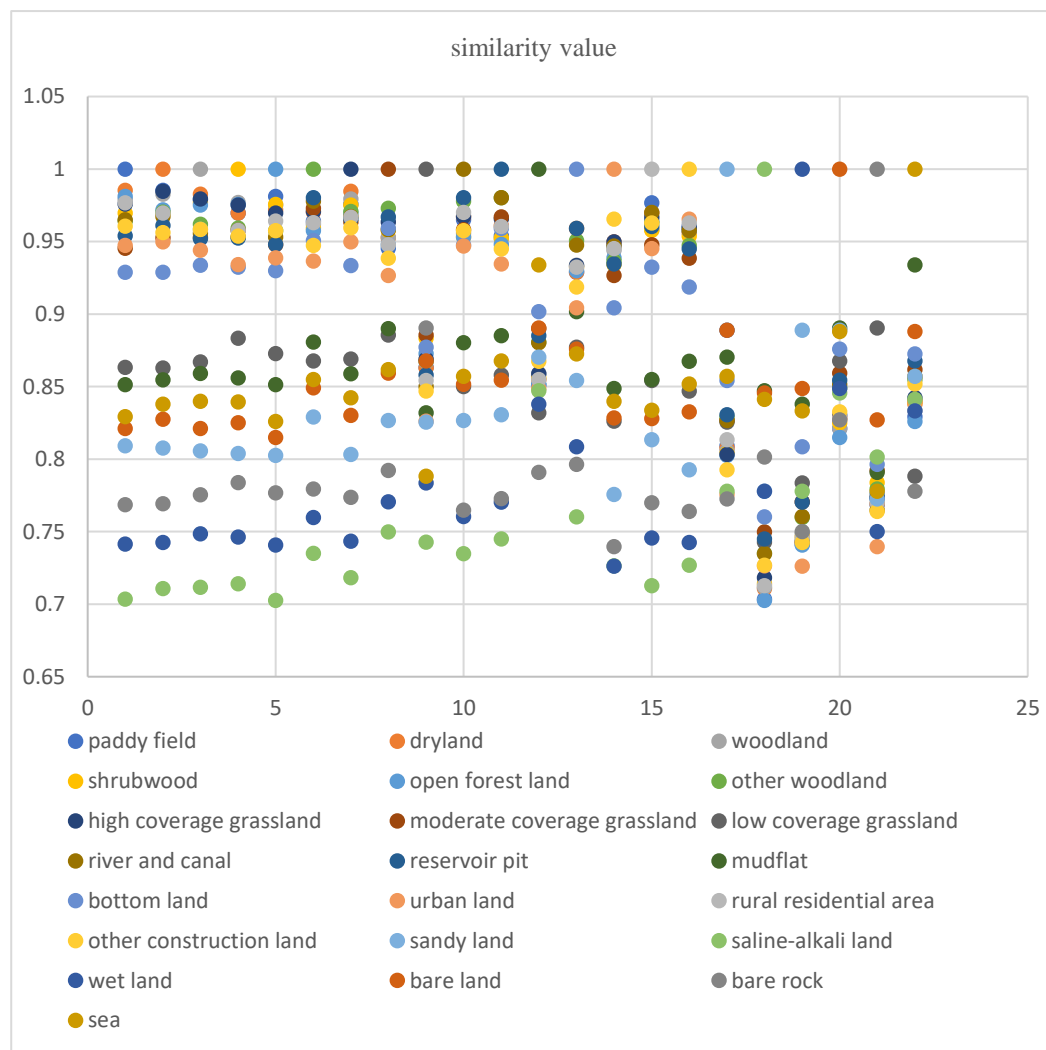
different land use types. Finally, the land use type membership function for each index provided the socio-economic similarity measure between 22 land use types in Guangxi.



**Figure 5.** Computing framework for the similarity measures of land use change types in Guangxi based on granular computing closeness.

#### 4.3. Results Analysis

Through the calculation of the similarity measure of land use change types based on granular computing closeness proposed in this paper, the similarity matrix of Figure 6 was obtained by calculating of the similarity of land use economic and social indexes in the Guangxi Zhuang Autonomous Region. The abscissa in the figure, from 1 to 22, respectively, represents paddy field, dryland, woodland, shrubwood, open forest land, other woodland, high-coverage grassland, moderate coverage grassland, low-coverage grassland, river and canal, reservoir pit, mudflat, bottom land, urban land, rural residential area, other construction land, sandy land, saline-alkali land, wet land, bare land, bare rock and sea. It can be seen that the similarity coefficient for each land type and itself was 1, and the similarity coefficient was relatively high among types with larger land areas; the similarity coefficient between land types with less land area and other land types was low. The high similarity coefficient indicates that the two types of land use had similar proportions among different levels of different indexes, such as distances from a road being closer or further, GDP being higher or land being cultivated. Due to the variety of land types, the following discussion focuses on the similarity coefficients of paddy fields, forested land and urban land use.



**Figure 6.** Scatter plot of similarity measures of land use types in Guangxi.

Paddy fields refer to cultivated land with a water source guarantee and irrigation facilities that can be irrigated in typical years to grow aquatic crops such as rice and lotus root, and that includes cultivated land for rice and dry crop rotation. Guangxi paddy fields account for more than 75% of Guangxi's cultivated land, and are mainly distributed in the plains, terraces and hilly areas of east and southeast Guangxi, as well as in the west of Guangxi and the mountainous areas of northern Guangxi, especially in the karst mountains. The distribution of paddy fields in Guangxi is uneven and scattered in the valleys of the mountains. The five land types with the largest similarity coefficient to the paddy field land type were dryland, open forest land, woodland, rural residential area and high-coverage grassland; the five land types with the smallest similarity coefficient to the paddy field land type were saline-alkali land, wet land, bare rock, sandy land and bare land. Paddy fields are generally located in agricultural areas, and are generally part of rural settlements outside of towns. Population density, GDP, distance from the road and sea level were all highly similar, and the brightness of night lights in rural areas also showed a high degree of similarity. The paddy fields in Guangxi are mainly planted with rice, and to ensure the irrigated conditions of the paddy fields, especially in hilly areas, they are commonly distributed around the topographic conditions of reservoirs, ponds and rivers. Guangxi paddy fields also have a high degree of similarity with forested land and other wooded land, which are also related to the topographical conditions in Guangxi, and these three types are mainly distributed in low-altitude, low-slope areas. Guangxi paddy fields and low-coverage grassland have a

small similarity coefficient. Low coverage grassland refers to natural grassland with a coverage of 5–20% that is distributed in areas where water is scarce, grass is sparse and conditions are poor for animal husbandry. Low coverage grassland has completely the opposite natural conditions to paddy fields, so the spatial distribution is completely different, resulting in smaller distances between roads, night light values, population density, GDP and other indexes at different levels. For the other land types with smaller similarity coefficients with the paddy fields in Guangxi, due to the small values of the coefficients, it is not feasible to explain the similarity measurement problem. Guangxi needs to persist in making good use of land and resources policies, fully support the alleviation of poverty, unswervingly adhere to the red line of cultivated land quantity and quality, strictly implement the balance of cultivated land occupation and compensation, vigorously implement land remediation and the upgrading of cultivated land and stimulate the vitality of rural land resources.

Woodland refers to natural forests and plantations with a canopy density >30%, including timber forests, economic forests and shelter forests. Guangxi woodland accounts for 54.9% of the forest land in Guangxi, is a major type of forest land in Guangxi and is mainly distributed in the mountains and hills of Guangxi and north Guangxi. The five land types with the largest similarity coefficient to the woodland type were dryland, high-coverage grassland, paddy field, shrubwood, open forest land and other woodland; the five land types with the smallest similarity coefficient to the woodland type were saline-alkali land, wetland, bare rock, sandy land and sea. The similarity coefficient between woodland, paddy field and dryland in Guangxi was high, which indicates that woodland in Guangxi is located around cultivated land. Because woodland is composed of some economic forests and timber forests, the participation of human social activities is high, and this land type has very similar features to cultivated land in terms of social and economic factors, as well as for natural landform. The five types with low similarity coefficients to woodland were the same as for the paddy fields, and because of the small number of land types, they cannot explain the similarity measurement problem. The similarity coefficients of woodland require that Guangxi insist on giving full play to natural conditions and resource advantages, scientific planning, accurate positioning, highlighting characteristics, implementing comprehensive management of rocky desertification areas and developing precious tree species and characteristic economic forests according to local conditions. Abundant forest resources provide very high ecological and economic value, and would consolidate the role of forestry in promoting and safeguarding Guangxi's economic and social development.

Urban land refers to land for large, medium and small cities and built-up areas in the county, and these areas are distributed around the sea and roads, with high population density, high GDP, high night light values and relatively low altitude. The five land types with the largest similarity coefficients to the urban land type were other construction land, dryland, high-coverage grassland, paddy field and river and canal. The five land types with the smallest similarity coefficients to the urban land type were wet land, saline-alkali land, bare rock, sandy land and low-coverage grassland. Other construction land refers to factories, mines, large industrial areas, oil fields, salt fields, quarries and other land, as well as traffic roads, airports and special land, and is very similar to the layout conditions of urban land. Urban land has a high similarity with dryland, high-coverage grassland, and paddy fields, indicating the possibility that Guangxi will be converted from these types of land in future urban expansion. The urban layout itself has the characteristics of rivers and coastal areas, and as such has a high similarity with the river and canal types in terms of a social economy. The land types with small similarity coefficients to urban land use in Guangxi were mainly grassland and shrub. These land types are far away from the city, and the requirements for water sources and natural conditions are not like those for forest land and cultivated land. Therefore, there is a low coefficient of similarity measure in nature, society and economy. In terms of urban land use, Guangxi's innovative land management model is need to optimize land use approval procedures, improve the land benefit distribution mechanism, promote of the formation of unified planning, produce city integration and benefit sharing, promote local social and economic development, ensure land use demand for the new

industrialization and urbanization of Guangxi and deepen the structural reform of the supply side of land resources.

If a traditional fuzzy set is used to calculate the similarity of the case area, only the similarity measure between the two indexes can be calculated, and the comprehensive hierarchical measure of different indexes cannot be performed. A comparative calculation of the land use change type lattice closeness and rough set method in the experimental area is also carried out in this paper. The results of the proposed method are consistent with the similarity ranking results based on the two methods. This paper only studies the socio-economic similarities of the case area in 2010. More work can be used to study the similarities of long-term sequences in the case area using remote sensing data, such as in the literature [27,34].

## 5. Conclusions and Discussion

### 5.1. Conclusions

Granular computing closeness was used to measure the similarities among spatial entities in this paper, and the land use change data in Guangxi was taken as an example. The three main conclusions are as follows:

- (1) The similarity problem of spatial data attribute calculation in granular computing is introduced in this paper, and a similarity measure method for spatial entity type attributes based on granular computing closeness is proposed. The spatial information particles, atomic formulas, and formulas are defined and characterized. Based on fuzzy set theory, the similarity criteria of spatial entity attributes are discussed. The concept of distance between granular information and the similarity measure of granularity is proposed based on the principle of simple set membership degree and distance, so that the atomic formula can be used to granulate spatial entities and spatial attribute indexes to form different spatial entity granular sets. This method can measure the similarity between different types of spatial entities, which is a similarity problem between samples, and it can provide decision-making guidance for determining the similarities of socio-economic attributes between spatial entities as a basis for spatial classification. This study can also provide a unified theoretical framework for the study of spatial similarity in granular computing, enrich the spatial information science computing method and broaden the application field of granular computing.
- (2) This method of measurement uses the formula to divide different sets of indexes and overcomes the drawback that traditional fuzzy sets can only measure between the two types. The size of the method dimension does not depend on the number of spatial sets but instead depends on the number of indexes of the spatial evaluation object. Therefore, this method not only reflects the degree of similarity of spatial entity types at different indicator levels, but also reflects the comprehensive similarity between spatial entity types under the consideration of comprehensive indexes.
- (3) Taking the secondary land use types of the Guangxi Zhuang Autonomous Region in 2010 as an example, a land use change calculation framework based on granular computing closeness is established that also applies to index increases. In the example of the similarity measure of the socio-economic indexes of the land use types in Guangxi, the similarity coefficient between urban land and grassland was low, but the similarity coefficient between other land for construction and river types was higher. The similarity coefficient between forestland and agricultural land rural settlements was higher, while the similarity coefficient between grassland and wetland was lower. The similarity measure between these different land use types illustrates the proximity of different land use types in Guangxi to social and economic attributes. These laws are basically consistent with the actual situation in Guangxi and can provide reference for land planning and decision support in Guangxi. Land use is a function process determined by the coordination of land quality characteristics and social land demand, and is also an economic process that interacts

with the economy and society. The land use types with high similarity values of measurement indexes have similar social and economic service values.

## 5.2. Discussion

Due to the limitations of data acquisition, only five indexes were selected in the case area, which is not sufficient to comprehensively measure the attribute similarities between land use types in Guangxi. The similarity of attributes is more likely to be able to provide the basis for classification problems. Based on the method of granular computing closeness, the classification of land use types from the perspective of socio-economic attributes will be a new perspective to explore in our future work.

**Author Contributions:** Conceptualization, W.L. and W.J.; methodology, W.L.; software, D.H.; validation, W.L., D.H. and W.J.; formal analysis, D.H.; investigation, D.H.; resources, W.J.; data curation, D.H.; writing—original draft preparation, W.L.; writing—review and editing, D.H. and W.J.; visualization, W.L.; supervision, D.H.; project administration, W.L.; funding acquisition, W.J.

**Funding:** This research was funded by the National Key Research and Development Program of China [2016YFC0503002], the National Natural Science Foundation of China [41571077], the Guangxi Key Research and Development Program [AB18126007].

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Ding, H. A study on Spatial Similarity Theory and Calculation Model. Ph.D. Thesis, Wuhan University, Wuhan, China, 2004.
- Elkwaie, E.A.; Kabuka, M.R. A robust framework for content-based retrieval by spatial similarity in image databases. *Acm Trans. Inf. Syst.* **1999**, *17*, 174–198. [\[CrossRef\]](#)
- Mansoor, K. Spatial Databases: Concepts of spatial similarity relations with the view point of fuzzy sets. In Proceedings of the IEEE Seventeenth UKSim-AMSS International Conference on Computer Modelling and Simulation, Cambridge, UK, 25–27 March 2015.
- Mark, D.M.; Freksa, C.; Hirtle, S.C.; Lloyd, R.; Tversky, B. Cognitive models of geographical space. *Int. J. Geogr. Inf. Sci.* **1999**, *13*, 747–774. [\[CrossRef\]](#)
- Holt, A. Understanding environmental and geographical complexities through similarity matching. *Complex. Int.* **2000**, *7*, 1–16.
- Schwering, A. Approaches to semantic similarity measurement for geo-spatial data: A survey. *Trans. Gis* **2008**, *12*, 5–29. [\[CrossRef\]](#)
- Xu, Y.; Xie, Z.; Chen, Z.; Wu, L. Shape similarity measurement model for holed polygons based on position graphs and fourier descriptors. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 253–279. [\[CrossRef\]](#)
- Zhao, Z.; Stough, R.R.; Song, D. Measuring congruence of spatial objects. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 113–130. [\[CrossRef\]](#)
- Fu, Z.L.; Fan, L.; Yu, Z.Q.; Zhou, K.C. A Moment-Based Shape Similarity Measurement for Areal Entities in Geographical Vector Data. *ISPRS Int. J. Geo Inf.* **2018**, *7*, 1–19. [\[CrossRef\]](#)
- Zhao, Y.; Zhou, X. Version similarity-based model for volunteers' reputation of volunteered geographic information: A case study of polygon. *Acta Geodaetica Cartographica Sinica*. **2015**, *44*, 578–584.
- Hu, M.K. Visual pattern recognition by moment invariants. *IRE Trans. Inf. Theory*. **1962**, *8*, 179–187.
- Zhang, D.; Lu, G. Review of shape representation and description techniques. *Pattern Recognit.* **2004**, *37*, 1–19. [\[CrossRef\]](#)
- Bai, X.J.; Wang, P.X.; Jie, Y.; Wang, L.; He, P. Spatial distribution characteristics of droughts in guanzhong plain based on structural similarity. *Trans. Chin. Soc. Agric. Mach.* **2015**, *46*, 345–351.
- Wang, J.J.; He, X.B.; Lin, J.H.; Huang, Q.Y.Q.; Zheng, C.X.; Zheng, F.W.; Li, R.G.; Jiang, J.X. Community structure and spatial distribution of macrobenthos in the shelf area of the bering sea. *Acta Oceanol. Sin.* **2014**, *33*, 74–81. [\[CrossRef\]](#)
- Gao, M.; Wang, X.; Wang, D. Species spatial distribution analysis using nearest neighbor methods: Aggregation and self-similarity. *Ecol. Res.* **2014**, *29*, 341–349. [\[CrossRef\]](#)



16. Culbert, P.D.; Radeloff, V.C.; Flather, C.H.; Kellndorfer, J.M.; Rittenhouse, C.D.; Pidgeon, A.M. The influence of vertical and horizontal habitat structure on nationwide patterns of avian biodiversity. *The Auk* **2013**, *130*, 656–665. [[CrossRef](#)]
17. Natalia, M.K.; Nagengast, B. The influence of the spatial structure of hydromacrophytes and differentiating habitat on the structure of rotifer and cladoceran communities. *Hydrobiologia* **2006**, *559*, 203–212.
18. Gidi, N.; Izhaki, I. Stability of pre- and post-fire spatial structure of pine trees in aleppo pine forest. *Ecography* **2006**, *21*, 535–542.
19. Wu, D.; Mendel, J.M. A comparative study of ranking methods, similarity measures and uncertainty measures for interval type 2—Fuzzy sets. *Inf. Sci.* **2009**, *179*, 1169–1192. [[CrossRef](#)]
20. Huang, B.; Guo, C.X.; Li, H.X.; Feng, G.F.; Zhou, X.Z. An intuitionistic fuzzy graded covering rough set. *Knowl. Based Syst.* **2016**, *107*, 155–178. [[CrossRef](#)]
21. Butenkov, S.; Zhukov, A.; Nagorov, A.; Krivsha, N. Granular computing models and methods based on the spatial granulation. *Procedia Comput. Sci.* **2017**, *103*, 295–302. [[CrossRef](#)]
22. Lin, T.Y. Granular computing on binary relations (ii): Rough set representations and belief functions. *Rough Sets Knowl. Discov.* **1998**, *1*, 121–140.
23. Yang, J.; Wang, G.; Zhang, Q. Knowledge distance measure in multigranulation spaces of fuzzy equivalence relations. *Inf. Sci.* **2018**, *22*, 408–423. [[CrossRef](#)]
24. Jiang, H.; Chen, Y. Neighborhood Granule Classifiers. *Appl. Sci.* **2018**, *8*, 2646. [[CrossRef](#)]
25. Ma, Y.Y.; Meng, H.L.; Xu, J.C.; Zhu, M. Normal distribution of lattice close-degree based on granular computing. *J. Shandong Univ. (Nat. Sci.)* **2014**, *49*, 107–110.
26. Liu, Y.F.; Yang, Z.Z.; Sun, X.L.; Si, R.C. Temporal semantic characteristics of spatial entities' attributes and an algebraic framework. *Geomatics Inf. Sci. Wuhan Univ.* **2013**, *38*, 1097–1102.
27. Radoi, A.; Burileanu, C. Retrieval of Similar Evolution Patterns from Satellite Image Time Series. *MDPI Appl. Sci.* **2018**, *8*, 2435. [[CrossRef](#)]
28. Rodgers, J.L.; Nicewander, W.A. Thirteen ways to look at the correlation coefficient. *Am. Stat.* **1988**, *42*, 59–66. [[CrossRef](#)]
29. Seifoddini, H.; Djassemi, M. The production data-based similarity coefficient versus jaccard's similarity coefficient. *Comput. Ind. Eng.* **1991**, *21*, 263–266. [[CrossRef](#)]
30. Schockaert, S.; Cock, M.D.; Cornelis, C.; Kerre, E.E. Fuzzy region connection calculus: An interpretation based on closeness. *Int. J. Approx. Reason.* **2008**, *48*, 332–347. [[CrossRef](#)]
31. Liu, H.; Xiong, S.; Fang, Z. Fl-grcca: A granular computing classification algorithm based on fuzzy lattices. *Comput. Math. Appl.* **2011**, *61*, 138–147. [[CrossRef](#)]
32. Liao, H.; Xu, Z.; Zeng, X.J. Distance and similarity measures for hesitant fuzzy linguistic term sets and their application in multi-criteria decision making. *Inf. Sci.* **2014**, *271*, 125–142. [[CrossRef](#)]
33. Patra, B.K.; Nandi, S.; Viswanath, P. A distance based clustering method for arbitrary shaped clusters in large datasets. *Pattern Recognit.* **2011**, *44*, 2862–2870. [[CrossRef](#)]
34. Fonji, S.F.; Taff, G.N. Using Satellite Data to Monitor Land-Use Land-Cover Change in North-Eastern Latvia. *Springerplus* **2014**, *3*, 61. [[CrossRef](#)] [[PubMed](#)]

