

Article

Similarity Measurement of Metadata of Geospatial Data: An Artificial Neural Network Approach

Zugang Chen ^{1,2,3} , Jia Song ^{1,2}  and Yaping Yang ^{1,2,4,*}

¹ State Key Laboratory of Resources and Environmental Information System, Beijing 100101, China; czgwan@gmail.com or czgbjy@yeah.net (Z.C.); songj@igsrr.ac.cn (J.S.)

² Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

³ University of Chinese Academy of Sciences, Beijing 100049, China

⁴ Jiang Su Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

* Correspondence: yangyp@igsrr.ac.cn; Tel.: +86-010-6488-9045

Received: 24 December 2017; Accepted: 7 March 2018; Published: 9 March 2018

Abstract: To help users discover the most relevant spatial datasets in the ever-growing global spatial data infrastructures (SDIs), a number of similarity measures of geospatial data based on metadata have been proposed. Researchers have assessed the similarity of geospatial data according to one or more characteristics of the geospatial data. They created different similarity algorithms for each of the selected characteristics and then combined these elementary similarities to the overall similarity of the geospatial data. The existing combination methods are mainly linear and may not be the most accurate. This paper reports our experiences in attempting to learn the optimal non-linear similarity integration functions, from the knowledge of experts, using an artificial neural network. First, a multiple-layer feed forward neural network (MLFFN) was created. Then, the intrinsic characteristics were used to represent the metadata of geospatial data and the similarity algorithms for each of the intrinsic characteristics were built. The training and evaluation data of MLFFN were derived from the knowledge of domain experts. Finally, the MLFFN was trained, evaluated, and compared with traditional linear combination methods, which was mainly a weighted sum. The results show that our method outperformed the existing methods in terms of precision. Moreover, we found that the combination of elementary similarities of experts to the overall similarity of geospatial data was not linear.

Keywords: artificial neural networks; geospatial data; similarity; metadata; intrinsic characteristics; combination

1. Introduction

Geospatial data play an important role in enhancing the capability of humans to monitor and understand society and nature [1]. They are widely used for decision making, Earth system science research, and so on [2]. In the past decades, billions of gigabytes of geospatial data have been produced from multiple Earth orbit missions, ground surveys, and in situ measurements and made available to the public through the spatial data infrastructures (SDIs, e.g., catalogs and portals) by government agencies and other stakeholders [3]. A major challenge has become how to help the user find the most relevant datasets in the ever-growing global SDIs.

Metadata is ‘data about data’ [4]. It is a structured description of the necessary properties of an object [5]. Most existing SDIs adopt metadata to describe, manage, discover, and exchange data [6,7]. To help users discover the relevant spatial datasets in SDIs, some solutions that are based on metadata of geospatial data have been proposed, such as linked geospatial data [8,9], data recommendation

systems [10], and so on. Among them, assessing the similarity of metadata of geospatial data and then recommending or linking geospatial data according to the similarity is demonstrated to be an attractive approach [10–13]. For example, for the geospatial data ‘2005 land use dataset of San Francisco Bay Area on 1:100,000’ (a), ‘2000 land use dataset of Texas on 1:100,000’ (b), and ‘2005 land use dataset of California on 1:100,000’ (c), the existing method [10–13] can compute a quantitative similarity between (a) and (c) that is higher than that of (a) and (b) based on the metadata information: the thematic contents of (a), (b), and (c) are the same but the spatial coverage of (c) contains that of (a) and the time coverage of (a) and (c) is the same. Then, the relevant data to (a) can be recommended and ranked by similarity.

Geospatial data have many characteristics, such as thematic context, spatial coverage, temporal coverage, topic category, data type, spatiotemporal precision, provenance, and so on. According to the roles of these characteristics in data discovery, the characteristics of geospatial data can be divided into two types: intrinsic and morphologic characteristics. Intrinsic characteristics refer to the basic ‘what, where, when’ triple features of geospatial data, namely, the thematic content, spatial and temporal coverage. These characteristics make geospatial data distinguishable from one another. Morphologic characteristics represent the structural and shape features of geospatial data, such as data type, format, and spatiotemporal precision. Morphologic characteristics can be transformed with more or less information loss without affecting the nature of geospatial data [9]. Both the intrinsic and morphologic characteristics are generally described by metadata formally [14,15]. Researchers have assessed the similarity of metadata of geospatial data based on one or several data characteristics [9,12,13]. They built different similarity measures for each of the selected characteristics and then combined these elementary similarities to the overall similarity of geospatial datasets. (The similarity of geospatial data refers to the similarity of metadata of geospatial data hereafter.) For example, for the geospatial data (a) and (b), the similarity of (a) and (b) can be computed by integrating the elementary similarities of their characteristics of thematic content, spatial and temporal coverage. The main issue in the integration of several similarity approaches into one similarity function is how different measures can be combined. Many integration schemes have been proposed in the literature. These schemes can be divided into three categories: standard combinations, linear combinations, and non-linear combinations. Standard combinations calculate the maximal, minimal, or median values of elementary similarities of characteristics of geospatial data [16]. Linear combinations assign a weight value to each of the elementary similarities and then take the sum of the weighted similarity scores as the final results [17]. The performances of standard combinations and linear combinations have been studied extensively [10,12,13,18–20]. In contrast to standard and linear combinations, non-linear combinations allow elementary similarities to be combined in more complex manners. Artificial neural network is an important approach to the learning of non-linear similarity functions [21,22]. Many experimental results have shown that significant improvements in similarity measures could be achieved by combining multiple similarities non-linearly [16,17]. However, there is no previous work on the non-linear integration of elementary similarities of characteristics of geospatial data, which is the main focus of this paper.

This paper reports our experiences in attempting to learn optimal similarity integration functions of geospatial data from the knowledge of experts using an artificial neural network. The performance of our approach has been compared with the traditional linear combination method. The results show that our method can achieve a higher precision than the existing methods and demonstrate that the integration of elementary similarities of experts to the overall similarity of geospatial data is not linear.

The remainder of this article is organized as follows. Section 2 surveys relevant literature on geospatial data similarity. Section 3 details the artificial neural network algorithms and different similarity measures for each of the selected characteristics of geospatial data. The artificial neural network is trained and evaluated in Section 4. We conclude with a summary and discussion of directions for future research in Sections 5 and 6.

2. Background

Advances in linked geospatial data [9], geographic information retrieval (GIR), and the uptake of the spatial data infrastructure initiative have led to urgent requirements on assessing the similarity of geospatial data. Some similarity measures about geospatial data have been proposed. These measures can be classified into two main families: similarity measures of user's query to geospatial data and geospatial data to geospatial data.

For example, in the context of geographic information retrieval, Lacasta et al. [23] aggregated users' search results of geospatial datasets by identifying the implicit spatial and thematic relations between the metadata records of geospatial datasets as similarity to offer complete answers for a user's query. Martins et al. [16] ranked geospatial data retrieval results according to a combination of thematic and geographical similarity between a user's query and the geospatial datasets. They proposed four combination methods (a standard, two linear, and a non-linear combination). Hu and Ge [17] presented an approach that learned GIR ranking functions using genetic programming (GP) methods based on textual statistics and geographic properties derived from the metadata of geospatial data and user queries. These three methods were used for geographic information retrieval and are not suitable for assessing the similarity between geospatial data and geospatial data. Andrade et al. [18] proposed several similarity metrics to solve spatial, semantic, and temporal queries and combined them by a weighted sum method. Al-Bakri and Fairbairn [24] measured semantic, structural, and data type similarities between categories of formal data and volunteered geographic information (VGI) and obtained the overall similarity based on a weighted sum combination of these three measures. Besides being used for GIR, these two methods combined elementary similarities by a linear method.

In the context of linked geospatial data, Zhao et al. [12,13] used the intrinsic characteristics of geospatial datasets to link geospatial datasets and quantified the overall interlinking as similarity that considered all data characteristics. Zhu et al. [9] proposed a multidimensional and quantitative interlinking approach for geospatial datasets that considered the characteristics of theme, category, spatial coverage, temporal coverage, spatial precision, temporal granularity, type, and format of geospatial datasets. For these two methods, the elementary similarities of selected characteristics are combined to the overall similarity of geospatial data by a weighted sum method.

Since these similarity integration functions were intuitively and empirically derived, they might not be the true integration function. If the true integration function is found or simulated, significant improvements in the precision of similarity measure can be achieved [25]. One way to obtain the optimal function is to learn it from the knowledge of experts [25]. Artificial neural networks (ANNs) have a remarkable ability to learn any linear or non-linear function from input and output data. Therefore, they are widely used in domains of search engines [26], power systems [27], transportation [28], agriculture [29], meteorology [30], and so on. In this article, we use the artificial neural network to learn the optimal functions from the knowledge of experts and combine the elementary similarities of selected characteristics to the overall similarity of geospatial data, aiming to improve the precision of the similarity measures of geospatial data.

In the next section, we will detail the artificial neural network algorithms and the similarity measures for intrinsic characteristics of geospatial data.

3. Methodology

3.1. Basic Idea

The proposed approach aims to integrate the elementary similarities of characteristics of geospatial data to overall similarity by using artificial neural networks. Artificial neural networks or neural networks are general terms for computer algorithms built as imitations of biological neural networks interconnected by a number of artificial neuron nodes ('neurons' hereafter). Artificial neural networks have remarkable capabilities in pattern recognition and trend predictions. They can learn laws from data that is complicated or imprecise and, thus, they have been widely used in various domains [31].

Before an artificial neural network can work, the prior knowledge that is used to train the network is required. The prior knowledge consists of the input data and output data.

The details of the proposed method are as follows: First, the intrinsic characteristics are selected to represent geospatial data for the sake of simplicity and generalization. Then, the quantitative similarity algorithms for each of the selected characteristics are built to obtain the input data for the ANN. To obtain the output data of the prior knowledge, some geospatial data experts are asked to rate the similarity for the designed geospatial data pairs according to the intrinsic characteristics. A multiple-layer feedforward neural network (MLFFN) is then created and trained by using the overall similarity of geospatial data as the desired correct output value, which is given by experts, and the elementary similarities of intrinsic characteristics as the input values, which is calculated by corresponding algorithms. The trained artificial neural network will be evaluated and compared with existing methods in terms of precision. After the evaluation, the trained artificial neural networks can be used to calculate the overall similarity of inter-geospatial data. The basic idea is shown in Figure 1.

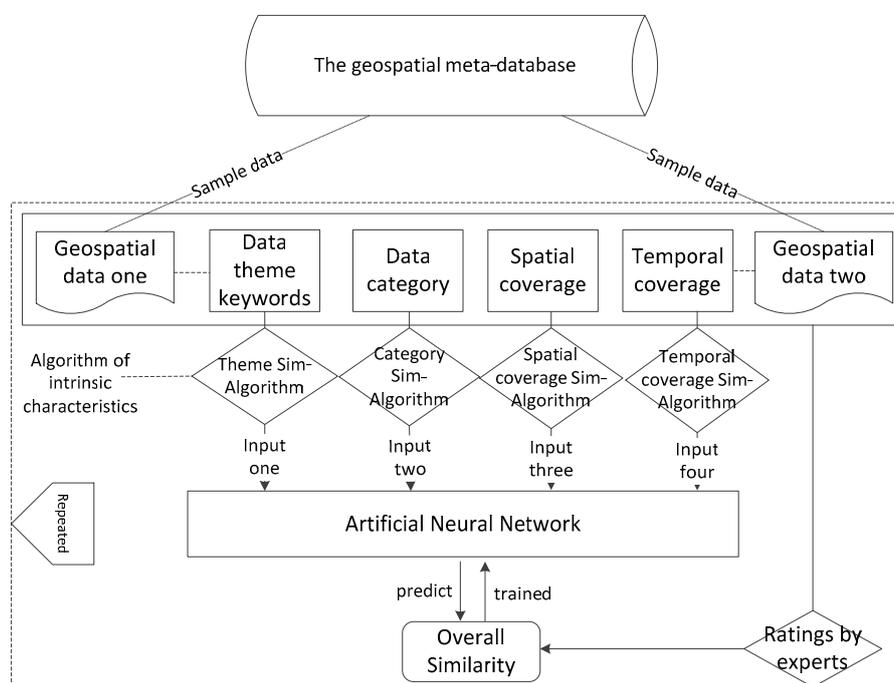


Figure 1. The basic ideal of our approach.

3.2. Artificial Neural Network Algorithm

Of the family of ANN algorithms, multiple-layer feedforward neural networks (MLFFNs) are quite popular because of their ability to model complex relationships between output and input data. Adding more hidden units to a network makes it possible for a MLFFN to represent any continuous, or even discontinuous, function of the input parameters. Moreover, compared to deep neural networks, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), the MLFFN requires less training samples, less training time, and lower computational and processing ability of the computer. It is a lightweight neural network [32]. Hence, a MLFFN is the best choice for our research. The structure of a MLFFN can be designed based on the problem to be solved. Figure 2 shows one structure of an MLFFN.

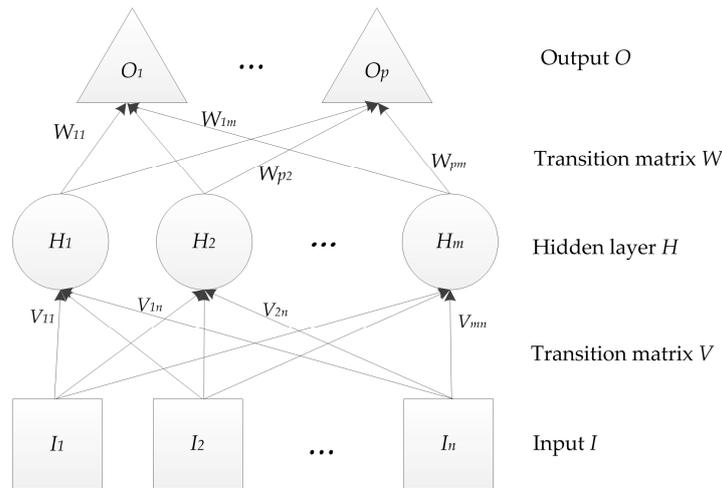


Figure 2. One structure of a multiple-layer feedforward neural network (MLFFN) (adapted from [33]).

The core of the algorithm is backpropagation and forward propagation, where backpropagation is used to train the neural net to get a stable transition matrix V , which transmits information from the input nodes to the hidden nodes, and W , which transmits information from the hidden nodes to the output nodes. Forward propagation is used to measure the difference between predicted output and the desired output using current V and W . The MLFFN uses the mean square error (MSE) as the error metric between the output O_k and the desired correct output C_k [33,34]:

$$E = \frac{1}{p} \sum_{k=1}^p (C_k - O_k)^2 \quad (1)$$

where p is the number of neurons in the output layer O . The detailed algorithm is as follows:

- (1) Initialize V and W with the given boundaries
- (2) Input data D (a set of input vectors)
- (3) For each element in D ,
 - a. Perform forward propagation as follows:

$$H_i = \Phi\left(\sum_{j=1}^n V_{ij}I_j - \theta_i\right) \quad (i = 1, 2, \dots, m) \quad (2)$$

$$O_k = \Phi\left(\sum_{j=1}^m W_{kj}H_j - \theta_k\right) \quad (k = 1, 2, \dots, p) \quad (3)$$

where $\Phi(x)$ is the transfer function; I is the input value; H_i is the output value of the hidden layer H ; θ is the bias; O_k is the output of layer O .

- b. Calculate the mean square error (MSE) between each neuron's output and its desired correct output in layer O . If the MSE is lower than the given good-minimum-error, then the network has completed the training, returning V and W as two transition matrices. If the MSE is not lower than the given good-minimum-error, perform backpropagation.

$$\Delta W_{kj} = \frac{\delta E}{\delta W_{kj}} = \frac{2}{p} (O_k - C_k) O_k (1 - O_k) H_j \quad (k = 1, 2, \dots, p; j = 1, 2, \dots, m) \quad (4)$$

$$\Delta \theta_k = \frac{\delta E}{\delta \theta_k}; \Delta V_{ij} = \frac{\delta E}{\delta V_{ij}}; \Delta \theta_i = \frac{\delta E}{\delta \theta_i} \quad (5)$$

$$\begin{cases} W = W - \omega \times \Delta W \\ \theta_k = \theta_k - \omega \times \Delta \theta_k \\ V = V - \omega \times \Delta V \\ \theta_i = \theta_i - \omega \times \Delta \theta_i \end{cases} \quad (6)$$

where ω is the learning rate of the MLFFN.

- c. Repeat a and b until the MSE is lower than the given good-minimum-error.

Neuroph [35] is a lightweight Java neural network framework to develop common neural network architectures. It contains a well-designed, open-source Java library with a small number of basic classes that correspond to the basic neural network concepts. In this article, we use the Neuroph open-source Java library to create MLFFNs based on the Eclipse Kepler 2 Integrated Development Environment. The experiment was performed on a computer with 8 G memory, 4 CPU, and a Windows 7 64-bit operating system (Dell (China) Co. Ltd., Kunshan, China).

3.3. Similarity for Intrinsic Characteristics of Geospatial Data

Because an artificial neural network is a numerical algorithm, its input and output are all numerical. It is necessary to build the similarity algorithms for the intrinsic characteristics of geospatial data to obtain the elementary similarities. Then, the MLFFN integrates these similarities by learning from the knowledge by which data experts assess the similarity of the geospatial data according to data's intrinsic characteristics. In this article, the intrinsic characteristics refer to the theme, category, spatial coverage, and temporal coverage of geospatial data that can be derived from the metadata [14,15].

3.3.1. Theme Similarity

The theme of geospatial data is represented by thematic keywords. Geospatial data generally have a few thematic keywords. Each of the thematic keywords can be seen as a word vector. We proposed the following method to compute the similarity of thematic keywords.

Set the keyword set of geospatial data A to be $(f_{A1}, f_{A2}, \dots, f_{Am})$ and geospatial data B to be $(f_{B1}, f_{B2}, \dots, f_{Bn})$. The thematic keyword similarity between A and B is $SimK(A, B)$. The keywords in keyword sets $(f_{A1}, f_{A2}, \dots, f_{Am})$ and $(f_{B1}, f_{B2}, \dots, f_{Bn})$ are segmented to be word vector sets $(\vec{f}_{A1}, \vec{f}_{A2}, \dots, \vec{f}_{Am})$ and $(\vec{f}_{B1}, \vec{f}_{B2}, \dots, \vec{f}_{Bn})$. If a keyword f_{Ax} in $(f_{A1}, f_{A2}, \dots, f_{Am})$ is segmented into the word vector $\vec{f}_{Ax} = (a_1, a_2, \dots, a_k)(x = 1, 2, 3, \dots, m)$, a keyword f_{By} in $(f_{B1}, f_{B2}, \dots, f_{Bn})$ is segmented into the vector $\vec{f}_{By} = (b_1, b_2, \dots, b_s)(y = 1, 2, 3, \dots, n)$, and the similarity between \vec{f}_{Ax} and \vec{f}_{By} is $sim(\vec{f}_{Ax} \rightarrow \vec{f}_{By})$. Then, according to the algorithm presented by Corley and Mihalcea [36],

$$SimK(A, B) = \begin{cases} \frac{\sum_{x=1}^m \text{Max}(sim(\vec{f}_{Ax} \rightarrow \vec{f}_{B1}), sim(\vec{f}_{Ax} \rightarrow \vec{f}_{B2}), \dots, sim(\vec{f}_{Ax} \rightarrow \vec{f}_{Bn}))}{m} & (m \geq n) \\ \frac{\sum_{y=1}^n \text{Max}(sim(\vec{f}_{By} \rightarrow \vec{f}_{A1}), sim(\vec{f}_{By} \rightarrow \vec{f}_{A2}), \dots, sim(\vec{f}_{By} \rightarrow \vec{f}_{Am}))}{n} & (n > m) \end{cases} \quad (7)$$

where

$$sim(\vec{f}_{Ax} \rightarrow \vec{f}_{By}) = \begin{cases} \frac{\sum_{i=1}^k \text{Max}(sim(a_i, b_1), sim(a_i, b_2), \dots, sim(a_i, b_s))}{k} & (k \geq s) \\ \frac{\sum_{i=1}^s \text{Max}(sim(b_i, a_1), sim(b_i, a_2), \dots, sim(b_i, a_k))}{s} & (s > k) \end{cases} \quad (8)$$

The similarity $sim(a_i, b_j)$ can be computed by a WordNet-based method. In this article, the Patwardhan and Pedersen's vector method is used to get $sim(a_i, b_j)$, because this measure outperforms other measures in terms of precision according to [37].

3.3.2. Category Similarity

Geospatial dataset A and B generally have several categories from different classification systems. Their categories must be consistently converted to the designated category system. Global Change Master Directory [38] is considered the unified topic category in this work. The category similarity between A and B is set to be $SimC(A, B)$. According to the latest research [9], the category similarity $SimC(A, B)$ is computed by Equation (9):

$$SimC(A, B) = \frac{1}{n} \sum_{i=1}^n \text{Max}_{j=1}^m (sim(C_{Ai}, C_{Bj})) \tag{9}$$

where $sim(C_{Ai}, C_{Bj})$ is computed by Equation (10), which was given by Wu and Palmer [39]:

$$sim(C_{Ai}, C_{Bj}) = \frac{2 \times N(C_{AB})}{N(C_{Ai}) + N(C_{Bj}) + 2N(C_{AB})} \tag{10}$$

where C_{Ai} and C_{Bj} refer to the categories in a classification system, C_{AB} is the closest parent node of C_{Ai} and C_{Bj} , $N(C_{Ai})$ is the number of edges from C_{Ai} to C_{AB} , $N(C_{Bj})$ is the number of edges from C_{Bj} to C_{AB} , and $N(C_{AB})$ is the number of edges from C_{AB} to the root node of the classification system.

For example, geospatial data ‘Dar es Salaam Land Use and Informal Settlement Dataset’ (A) has two categories: ‘Global Change Master Directory > Human Dimensions > Human settlements > Urban areas’ (C_{A1}) and ‘Global Change Master Directory > Land Surface > Land Use/Land Cover > Land Use Classes’ (C_{A2}). ‘ISLSCP II Global Population of the World’ (B) has three categories: ‘Global Change Master Directory > Human Dimensions > Population > Population Distribution’ (C_{B1}), ‘Global Change Master Directory > Human Dimensions > Population > Population Size’ (C_{B2}), and ‘Global Change Master Directory > Land Surface > Land Use/Land Cover > Land Use/Land Cover Classification’ (C_{B3}). The $sim(C_{A1}, C_{B1}) = \frac{2 \times 1}{2 + 2 + 2 \times 1} = 0.333$, $sim(C_{A1}, C_{B2}) = 0.333$, $sim(C_{A1}, C_{B3}) = 0$, $sim(C_{A2}, C_{B1}) = 0$, $sim(C_{A2}, C_{B2}) = 0$, $sim(C_{A2}, C_{B3}) = 0.667$, and $SimC(A, B) = \frac{1}{2} \times (0.333 + 0.667) = 0.5$.

3.3.3. Spatial Coverage Similarity

The spatial coverage of geospatial data is usually represented by the minimum enclosing rectangle of the dataset [9]. The minimum enclosing rectangle extent can be regarded as a geospatial polygon. Therefore, how to compute the similarity of geospatial polygons is the key to computing the spatial coverage similarity. In this article, we use the topological and metric relations of inter-geospatial polygons to compute the similarity [40]. The similarity of geospatial coverages is calculated by Equation (11):

$$SimP(A, B) = W_{t0} + W_t \times C(A, B) \tag{11}$$

where $SimP(A, B)$ refers to the similarity of spatial coverage; W_{t0} refers to the minimum similarity of geospatial coverages under a specified topology relation; W_t is the weight of the metric relationship of the geospatial coverages under the corresponding topology relation; and $C(A, B)$ is the function of the metric relationship between two spatial coverages.

In this study, the topology relations between geospatial polygons are grouped into six categories, as shown in Table 1.

Table 1. The topology relations of geospatial polygons (adapted from [40]).

Topo. Relation	Equals	Contains	Within	Overlaps	Touches	Disjoints
Diagram						

Using the weight measurement method (WMM) of the analytical hierarchy process (AHP) (hereafter referred to as AHP-WMM) [41], we obtain the values of W_{t0} and W_t . The detailed steps of AHP-WMM are as follows. First, we establish a pairwise comparison matrix of the relative importance of all factors that affect the same upper-level goal. Then, domain experts establish pairwise comparison scores using a 1–9 preference scale. The normalized feature vector of the pairwise comparison matrix is regarded as the weight of the factors. If the number of factors is more than two, a consistency check is required. The standard to pass the consistency check is that the consistency ratio (CR) is less than 0.1. The weights of W_{t0} and W_t calculated by AHP-WMM are shown in Table 2.

Table 2. The values of W_{t0} and W_t in different spatial topology relations.

No.	Topology	W_{t0}	W_t
1	Equals	1	0
2	Within	0.667	0.333
3	Contains	0.667	0.333
4	Overlaps	0.5	0.167
5	Touches	0.333	0.167
6	Disjoint	0	0.333

We define the spatial distance as the Euclid distance between the geometric centers of the geospatial polygons. $C(A, B)$ is calculated by Equation (12).

$$C(A, B) = \begin{cases} \frac{Area(E_A \cap E_B)}{Area(E_B)}, & E_A \text{ within } E_B \\ \frac{Area(E_A \cap E_B)}{Area(E_A)}, & E_A \text{ contains or overlaps } E_B \\ \frac{Len(E_A \cap E_B)}{Len(E_A)}, & E_A \text{ touches } E_B \\ \frac{1}{1 + D(E_A, E_B)}, & E_A \text{ disjoints } E_B \end{cases} \quad (12)$$

where E_A and E_B refer to two geospatial polygons, $Area(E_A)$ and $Area(E_B)$ are the areas of E_A and E_B ; $Area(E_A \cap E_B)$ is the overlapping area of E_A and E_B ; $Len(E_A)$ is the perimeter length of E_A ; $Len(E_A \cap E_B)$ is the length of the intersection of E_A and E_B ; and $D(E_A, E_B)$ is the spatial distance between E_A and E_B .

3.3.4. Temporal Coverage Similarity

Temporal coverage consists of a textual time description, such as “the fifties of the twentieth century” and “September 2009.” Temporal coverage generally has two aspects: the beginning date and the ending date. However, sometimes the ending date is null, which means that the temporal coverage of geospatial dataset is an instant. An instant and an interval are relative and convertible to each other under different timescales; thus, we change an instant to an interval through time downscaling and unify the two intervals to the minimum timescale. For example, if the timescale of one geospatial data is “year” (for example, 2010) and the other data’s timescale is “month” (for example, September 2009–July 2012), the “year” timescale should be transferred to “month” (January 2010–December 2010) to maintain a consistent timescale between the two datasets to calculate their temporal coverage similarity.

The similarity of two time intervals can also be calculated according to the time topology and metric relation. The topology relation of two time intervals is the relation of containing, within, overlapping, touching, and disjointing [42], and their metric relation refers to the length of the overlap, the distance of the interval between them. The time topology relations are shown in Table 3. We propose using Equation (13) to calculate the temporal similarity $SimT(A, B)$.

$$SimT(A, B) = \begin{cases} 1 & T_A \text{ Equal } T_B \\ W_{T0} + W_{t0} \times \frac{Len(T_A \cap T_B)}{Max(Len(T_A), Len(T_B))} & T_A \text{ Contains or Within } T_B \\ W_{T1} + W_{t1} \times \frac{Len(T_A \cap T_B)}{Max(Len(T_A), Len(T_B))} & T_A \text{ Overlaps } T_B \\ W_{T2} + W_{t2} \times \frac{2}{Len(T_A) + Len(T_B)} & T_A \text{ Touches } T_B \\ W_{t3} \times \frac{1}{Dis(T_A, T_B)} & T_A \text{ Disjoints } T_B \end{cases} \quad (13)$$

where T_A and T_B are two time intervals; $Len(T_A)$ and $Len(T_B)$ are the lengths of T_A and T_B ; $Len(T_A \cap T_B)$ is the overlapping length of T_A and T_B ; $Dis(T_A, T_B)$ is the time distance between T_A and T_B , which is equal to the middle of T_A minus the middle of T_B ; W_{T0} , W_{T1} , and W_{T2} are the topology weights when the topology relation between T_A and T_B is “Contains/Within,” “Overlaps,” and “Touches,” respectively; and W_{t0} , W_{t1} , W_{t2} , and W_{t3} are the metric relation weights when the topology relation between T_A and T_B is “Contains/Within,” “Overlaps,” “Touches,” and “Disjoints,” respectively. Using AHP-WMM, we get the values of W_{T0} , W_{T1} , and W_{T2} , which are equal to 0.667, 0.5, and 0.333, respectively, and the values of W_{t0} , W_{t1} , W_{t2} , and W_{t3} , which are equal to 0.333, 0.167, 0.167, and 0.333, respectively.

Table 3. The topology relations among time intervals.

Time Topo. Relation	Diagram	Time Topo. Relation	Diagram
Equal		Overlaps	
Contains		Touches	
Within		Disjoints	

Time interval one: time interval two:

4. Experiment and Results

4.1. Materials

The National Earth System Science Data Sharing Infrastructure (NESSDSI, <http://www.geodata.cn>) is one of the national science and technology infrastructures in China. It provides one-stop data sharing and open service. As of 15 November 2017, NESSDSI has shared 15,142 multi-disciplinary datasets, including geography, geology, hydrology, geophysics, ecology, and astronomy and the page view of the website has exceeded 21,539,917.

NESSDSI utilizes the ISO19115-based metadata to describe geospatial datasets. The metadata of NESSDSI includes the dataset title, dataset language, a set of thematic keywords, abstract, category, spatial coverage, temporal coverage, format, provenance, and so on. All the metadata and datasets can be openly accessed. We selected 1700 geospatial datasets and their metadata from NESSDSI whose contents were about basic geographic information, land use/cover, population, social economy, regionalization, landform, terrain, soil, desert, body of water, wetland, vegetation, environment, disaster, and natural resources. The intrinsic characteristics of these datasets, which were the thematic keywords, category, spatial coverage and temporal coverage, were extracted to build an intrinsic characteristic database of geospatial data (ICGDatabase for short). We used these selected datasets to create geospatial data pairs and asked geoscience experts to determine the similarity of these data pairs, which will be the prior knowledge of the artificial neural networks and the evaluation baseline for different similarity combination methods.

4.2. The Acquisition of Prior Knowledge

Prior knowledge acts as the training dataset for the artificial neural networks. It determines how well the transition matrices can be built into the machine learning process. Although a neural net is highly tolerant of noisy data, completeness and representativeness of the prior knowledge is still significant for the accuracy of the similarity computation of geospatial data. The training of a neural network is the process by which the neural net learns the laws and features contained in the prior knowledge (or sample data); if the sample data can represent the population excellently, the trained neural net will be accurate when it is used to make predictions.

4.2.1. The Features of Geospatial Data

As mentioned before, in this article, we use the thematic content, geospatial coverage, temporal coverage to represent the geospatial data. The detailed intrinsic characteristics of geospatial data include theme keywords, category, spatial coverage, and temporal coverage, which can be directly derived from the metadata of geospatial data. For each of the detailed intrinsic characteristics of geospatial data, different features between them will affect the similarity of the geospatial data. For example, compared with the “2000 land use dataset of San Francisco Bay” (A), the “2000 land use dataset of California” (B) is more similar to it than the “2000 land use dataset of Nevada” (C) because the feature between A and B in terms of spatial coverage is “within” while that between A and C is “disjoint.” There are five features between two spatial coverages: “same,” “contains or within,” “overlaps,” “touches,” and “disjoint.” There are also five features between two temporal coverages: “same,” “contains or within,” “overlaps,” “touches,” and “disjoint.” There are three features between the theme keywords of two geospatial datasets: same, similar, and non-similar. There are three features between two categories: same, parent and child, and sibling and other. The features of each detailed intrinsic characteristic are shown in Table 4. There are $3 \times 3 \times 5 \times 5 = 225$ combined features between two geospatial datasets.

Table 4. Features of each detailed intrinsic characteristic.

Overall Characteristics	Intrinsic Characteristics	Detailed Intrinsic Characteristics	Features
Geospatial data overall characteristic	Thematic content	Theme keywords	Same Theme Similar None-Similar
		Category	Same Parent and Child Sibling and Other
	Spatial coverage	Spatial topology	Same Contains or Within Overlaps Touches Disjoint
		Temporal coverage	Temporal topology

4.2.2. Survey Design and Results

Given the features that affect the similarity of geospatial datasets, the following experiment was designed to obtain prior knowledge for the similarity computation of geospatial data. We selected geospatial datasets and their detailed intrinsic characteristics from the ICGDatabase and created

geospatial data pairs. All the geospatial data pairs with different feature combinations of four detailed intrinsic characteristics formed the prior knowledge samples. Thus, there are $3 \times 3 \times 5 \times 5 = 225$ data pairs in the similarity rating questionnaire, which are called sample pairs. In order to evaluate the geospatial data similarity measures, another 20 pairs of geospatial data with different feature combinations (called evaluation pairs) were also added to the similarity rating questionnaire. We asked the geospatial data experts to rate the similarity for each pair of geospatial data in the similarity rating questionnaire on a 100-point scale from 0 to 100: 0 represents no relevance at all and 100 the same data pair. The ordering of the 245 pairs was randomly determined for each subject.

We received 37 complete responses. Intra-rater reliability (IRR) refers to the relative consistency in ratings provided by multiple judges of multiple targets [43]. In contrast, intra-rater agreement (IRA) refers to the absolute consensus in scores furnished by multiple judges for one or more targets [44]. Pearson's r is usually the index for IRR [45]. Kendall's W is usually for IRA [46]. Table 5 shows the plausible indices of IRR and IRA for our similarity ratings from experts.

Table 5. The indices of intra-rater reliability (IRR) and intra-rater agreement (IRA) for our similarity ratings.

Index Type	Index Name	Minimum	Median	Max.	Means	Value
IRR	Pearson's r	0.86	0.942	0.96	0.945	
IRA	Kendall's W					0.812

The indices of IRR and IRA in Table 5 indicate that the responses of experts possess a high reliability and are in agreement. The correlation is satisfactory and is better than analogous surveys [47].

For each pair of geographic datasets, we computed the mean ratings of the 37 experts and normalized them in the interval $[0, 1]$ as similarity scores. Among them, 225 similarity scores of the sample pairs were used to train the MLFFN and the other 20 scores of the evaluation pairs were used to evaluate the similarity measures. Given the small size of the evaluation pairs, an inspection about whether there was a distribution difference between the sample pairs and the evaluation pairs needed to be done. The Mann-Whitney U test [48] is the most commonly used nonparametric procedure in comparing two distributions based on independent samples. It is especially useful when the assumption of normality is not met. Using a Mann-Whitney U test, we obtained a p -value of 0.101 and concluded that there is no evidence to suggest that the error distributions in the two groups are different.

The training dataset was derived by the following method: for every pair of geospatial datasets in sample pairs, we used Equations (7), (9), (11) and (13) to compute the similarities of theme, category, spatial coverage, and temporal coverage. The four values were the input data of the training dataset. The overall similarity score of this pair of geospatial datasets, which was given by experts, was the output data of the training dataset. By the input and output data in the training dataset, an MLFFN can be trained to compute the similarity of inter-geospatial data based on the elementary similarities of intrinsic characteristics.

4.3. The Creating and Training of MLFFN

Once the training datasets had been collected, an MLFFN could be created and trained. To create an MLFFN with the best performance, two important factors must be considered: the architecture and the learning rate of the MLFFN.

4.3.1. Prediction Accuracy vs. the Architecture of MLFFN

The architecture of an MLFFN determines the number of connection weights (free parameters) and the way information flows through the network. Determination of an appropriate network architecture is one of the most important, but also one of the most difficult, tasks in the ANN model

building process. This is generally done by fixing the number of hidden layers and choosing the number of nodes in each of these layers [49]. The number of nodes in the input layer is fixed by the number of model inputs, whereas the number of nodes in the output layer equals the number of model outputs. Therefore, the input and output nodes of our MLFFN are 4 and 1, respectively, as listed in Tables 6 and 7. It has been shown that MLFFNs with one hidden layer can approximate any function [50]. However, in practice, many functions are difficult to approximate with one hidden layer, requiring a prohibitive number of hidden layers [51]. The use of more than one hidden layer provides greater flexibility and enables approximation of complex functions with fewer connection weights in many situations [51,52]. Flood and Kartam [51] suggested using two hidden layers as a starting point. Moreover, the rule of thumb says that the number of connections between neurons should not exceed the number of training samples [53] and larger networks (more than two hidden layers) generally require a large number of training samples to achieve good generalization ability [54]. There are 225 pairs of training data in our research, which is not a very large size; hence, two hidden layers were enough for our MLFFN. The number of nodes in the hidden layers was determined by the following method: MLFFNs with a different number of hidden layer nodes were evaluated and then the Pearson product-moment correlation coefficient r and root mean squared error (RMSE) were used [55,56] to determine the optimum network topology.

Table 6. Input parameters for our MLFFN and their ranges.

Input Parameter	Range
Theme similarity	0–1
Category similarity	0–1
Spatial coverage similarity	0–1
Temporal similarity	0–1

Table 7. Output parameters for our MLFFN and their ranges.

Output Parameter	Range
Overall similarity of inter-geospatial datasets	0–1

The goal of training an MLFFN is to maximize the coefficient r and minimize the RMSE and the iteration times. The initial parameters used for training the network are shown in Table 8. The tested architectures and evaluation results of the MLFFNs are listed in Table 9.

Table 8. Initial training parameters.

No.	Parameter	Value
1	Max. iterations	300,000
2	Max. error	10^{-4}
3	Initial learning rate	0.1
4	Momentum coefficient	0.25

In Table 8, parameter 1 is the largest number of steps that the MLFFN can run. Parameter 2 is measured by the mean square error (MSE), and a value of 10^{-4} means that the MLFFN will stop iterating if $MSE < 10^{-4}$. Parameter 3 is the initial learning rate, and the learning rate was set to different values in each training process in the next experiment. The introduction of parameter 4 cuts down the learning time and efficiently prevents the networks from remaining at local optima.

Evaluation of different MLFFNs resulted in a 4-10-5-1 network topology (Table 9). The Pearson's r and RMSE are equal to 0.958 and 0.0143, respectively.

Table 9. Evaluation results of different MLFFNs.

No.	Architecture	RMSE	Pearson's r	Iteration Times
1	4-25-2-1	0.0159	0.957	115,517
2	4-20-2-1	–	–	Over 300,000
3	4-17-3-1	–	–	Over 300,000
4	4-15-3-1	0.0149	0.957	74,479
5	4-13-4-1	0.0144	0.957	58,642
6	4-10-4-1	0.0144	0.957	71,596
7	4-10-5-1	0.0143	0.957	72,251
8	4-9-5-1	0.0146	0.957	70,996
9	4-8-6-1	0.0147	0.957	134,707
10	4-8-7-1	0.0153	0.957	234,536
11	4-7-7-1	–	–	Over 300,000
12	4-8-3-1	–	–	Over 300,000

Note: an architecture of 4-25-2-1 means two hidden layers for the MLFFN. The first hidden layer has 25 neurons and the second has 2 neurons. The number of connection weights is less than 225.

4.3.2. Quickness of Convergence vs. Learning Rate

The learning rate controls the speed of MLFFN learning by affecting the changes being made to the weights of transition matrices at each step. The performance of the ANN algorithm is very sensitive to proper setting of the learning rate [57]. If the learning rate is too low, the algorithm will take too long to converge. If the value is too large, the ANN becomes unstable and oscillates around the error surface. When we tried to get the optimal network topology, we set the learning rate to be 0.1, which is a small number. It may not be the optimal value. In the study, we need to adjust the learning rate to test the number of iterations and measure the prediction accuracy so that the optimal learning rate can be found.

By gradually increasing the learning rate, we recorded the number of iterations when the MLFFN with the network topology of 4-10-5-1 completed training. Then, the correlation coefficient r and RMSE between the prediction values of MLFFN and the desired correct output value given by experts were computed on the sample pairs. Figure 3 shows the experimental results of the learning rate by the number of iterations, Pearson's r and RMSE. The X-axis indicates a learning rate ranging from 0.1 to 0.9 with intervals of 0.1, whereas the Y-axis (left side) indicates the number of iterations when the MLFFN converges and the Y-axis (right side) indicates the values of the correlation coefficient r and the RMSE.

By analyzing Figure 3, we find that as the learning rate continually increases, the RMSE increases. The correlation coefficient between the prediction values of MLFFN and the desired correct output values given by experts on sample pairs keeps quite stable, ranging from 0.956 to 0.957. The number of iterations decreases first and then increases, although there is an abnormal value when the learning rate is equal to 0.4. This can be interpreted as follows: when the learning rate increases, the MLFFN can learn more quickly, but when the learning rate is increased to some degree, the algorithm becomes unstable, oscillates around the error surface, and takes more time to converge. When the learning rate is equal to 0.9, the MLFFN cannot converge at all. To ensure that our MLFFN has a high accuracy, we choose 0.1 as the optimal learning rate when r is 0.957 and the number of iterations is 72,251 (the number of iterations will vary depending on different computation conditions).

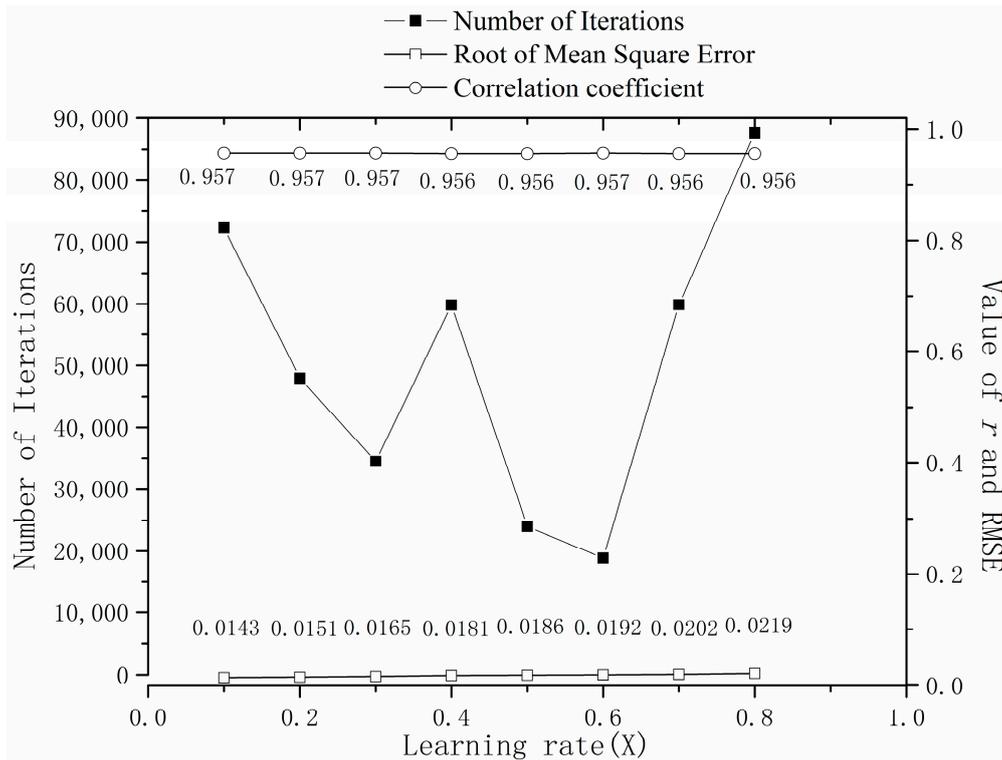


Figure 3. The relationship between the learning rate, number of iterations, and prediction accuracy of the MLFFN (Note: when the learning rate is 0.9, the MLFFN cannot be converged).

4.4. Comparison and Evaluation

Given the trained MLFFN, we now compared it with the existing methods to demonstrate the advances of it in improving the precision of similarity measure of geospatial data. The existing methods are mainly the weighted sum of the elementary similarities of the characteristics of geospatial data, for example, the methods of [9,10,13]. In this article, we use Equation (14) as the representative of the traditional methods:

$$S = \sum_{i=1}^n (W_{subi} \times S_{subi}) \tag{14}$$

where S is the overall similarity of geospatial data; S_{subi} denotes the similarity of i th detailed intrinsic characteristic of the geospatial data and W_{subi} is the corresponding weight; and n is the number of detailed intrinsic characteristics.

In this article, four detailed intrinsic characteristics were selected to represent geospatial data. According to [13], the weights of W_{subi} are 0.2378, 0.1722, 0.35, and 0.24 for theme, category, spatial coverage, and temporal coverage similarity, respectively.

The Pearson product-moment correlation coefficient r and RMSE of our trained MLFFN and traditional method on 20 pairs of geospatial data of evaluation pairs were computed, as shown in Table 10.

Table 10. The comparison of precision for our MLFFN and weighted sum methods.

Index Name	MLFFN	Weighted Sum
Pearson’s r	0.943	0.929
RMSE	0.015	0.075

We find that our MLFFN outperforms the traditional linear method in terms of combining elementary similarities of characteristics of geospatial data to overall similarity, though the precision of traditional weighted sum method is also high.

In order to give precise indications of the practical applicability of our proposed method, it was necessary to analyze the spatiotemporal computational complexity of the approach. For both our ANN-based method and the weighted sum method, the four elementary similarities of intrinsic characteristics must be computed first. The computational complexity of the step is equal for the two methods. We do not compare them here. After obtaining the four elementary similarities of the intrinsic characteristics of a pair of geospatial data, the weighted sum method gives the overall similarity at once. For n pairs of geospatial data, the weighted sum method has a linear complexity $O(n)$. For our trained MLFFN, the computational complexity is equal to that of forward propagation. As the network topology of our MLFFN is 4-10-5-1, for the worst case the computational complexity is $O(n(4 \cdot 10 + 10 \cdot 5 + 5 \cdot 1)) = O(95n)$, which is still a linear complexity. Space or memory complexity of our MLFFN is negligible since there are only $4 + 10 + 5 + 1 = 20$ neurons and $4 \cdot 10 + 10 \cdot 5 + 5 \cdot 1 = 95$ weights of connections that require memory space allocation. Although our MLFFN increased the computational complexity and memory cost, it does not constitute an obstacle for the practical usage of the technique.

5. Discussion

5.1. Interpretation of the Performance of Our Method

Why can the non-linear combination methods of elementary similarities of intrinsic characteristics of geospatial data, which is represented by our MLFFN, improve the precision of similarity computation of geospatial data? We ranked the 20 pairs of datasets from the evaluation pairs by the similarity scores generated by our MLFFN and the traditional weighted sum method, respectively. We found that most orders of the two groups of geospatial data are the same, but some are different. The different pairs are shown in Table 11.

Table 11. The data pairs in different orders ranked by MLFFN and weighted sum methods.

No.	Dataset A	Dataset B	MLFFN Similarity	Weighted Sum Similarity
1	2000 Tibet Plateau land use data	2000 Tibet Plateau sub-regional climate data	0.57	0.67
2	1980 Tibet Plateau land use data	2015 ShangHai land use data	0.59	0.43
3	1952–1993 Tibet Plateau weather and climate data	1951–2000 China average wind velocity data on a 1 km grid	0.66	0.73
4	1952–1993 Tibet Plateau weather and climate data	1980–1981 China agricultural phenology data	0.83	0.71

By analyzing dataset pairs 1 and 2 in Table 11, we found that our MLFFN deems that a pair of geospatial data is more similar when their theme contents are the same, even though their spatial coverage and time coverage are completely different. The weighted sum method gives a relatively high similarity score for two datasets with the same spatial coverage and temporal coverage even though their theme content is thoroughly different, which is in contrast to the expert's knowledge. Therefore, we could infer that the combination of elementary similarities of intrinsic characteristics of geospatial data to overall similarity must not be linear, although we cannot derive the functions explicitly.

5.2. Factors Affecting the Precision of Similarity Computation of Geospatial Data

Although our non-linear combination of elementary similarities of intrinsic characteristics of geospatial data achieved higher precision, there are still limitations that affect the similarity computation precision of geospatial data.

For example, for the theme similarity, our proposed method is incapable of computing all geospatial dataset's theme similarities because some geographic terminologies, such as "phenology," "foredune," "regionalization," "semi-arid climate," "periglacial landform," are not recorded in the WordNet database. How to build a geographic semantic web large enough and realize similarity computation of terms is still an urgent issue to tackle. Moreover, for the spatial coverage similarity algorithm when the topology relations gradually change, the corresponding similarities of spatial coverage change discontinuously. Table 12 shows the similarities of geospatial coverages computed by our method in Section 3.3.3, whose topology relations are ranging for "same," "contains/within," "overlaps," "touches," to "disjoint," but the similarities are ranging from 1, 0.82, 0.53, 0.34, to 1.03×10^{-6} . We know that an ANN has a better performance in fitting continuous functions than discontinuous ones of the input parameters. Therefore, we should create new algorithms for geospatial coverage similarity to get continuous results and further improve the performance of the MLFFN.

Table 12. Spatial coverage pairs with different topology relations and similarities.

Spatial Cover One	Spatial Cover Two	Topology Relation	Similarity
Jiang Su Province	Jiang Su Province	Same	1
Jiang Su Province	Yangtze River Delta	Within/Contains	0.82
Jiang Su Province	Taihu Basin	Overlaps	0.53
Jiang Su Province	Zhe Jiang Province	Touches	0.34
Jiang Su Province	Fu Jian Province	Disjoint	1.03×10^{-6}

Jiang Su, Zhe Jiang, and Fu Jian are provinces of China; Zhe Jiang is between Jiang Su and Fu Jian.

6. Conclusions

In this study, we built an artificial neural network and the similarity algorithms for intrinsic characteristics of geospatial data to combine the elementary similarities to overall similarity non-linearly. The prior knowledge was obtained from domain experts. The MLFFN was trained and evaluated. The results show that our proposed method achieves a high precision in terms of similarity computation of geospatial data and outperforms the traditional combination method of the weighted sum.

We first integrated the elementary similarities of intrinsic characteristics of geospatial data to an overall similarity by using an artificial neural network and demonstrated that the combination pattern in human rating process is not linear. Our method can be used as an accurate measure to assess the similarity of geospatial data.

As the study involves numerous research domains, there are still some problems that need to be solved. (1) Due to limited vocabularies of WordNet, particularly in the domain of geosciences, a new similarity measure of keywords should be proposed. (2) A new similarity algorithm for geospatial coverage must be presented to achieve continuous similarity results. (3) In this research, we considered only the intrinsic characteristics of geospatial data. If more characteristics are considered, the similarity of geospatial data can be assessed more comprehensively. (4) As the training of neural networks is a time-consuming process, we should take parallel computation into consideration to accelerate the training speed.

Acknowledgments: This work was supported by the Branch Center Project of Geography, Resources and Ecology of Knowledge Center for Chinese Engineering Sciences and Technology (No. CKCEST-2017-1-8), the National Earth System Science Data Sharing Infrastructure (No. 2005DKA32300), the Multidisciplinary Joint Scientific Expedition Project in International Economic Corridor Across China, Mongolia and Russia (No. 2017FY101300), the

Construction Project of Ecological Risk Assessment and Basic Geographic Information Database of International Economic Corridor Across China, Mongolia and Russia (No. 131A11KYSB20160091), and the National Natural Science Foundation of China (No. 41631177). We would like to thank the editors and the anonymous re-viewers for their very helpful suggestions, all of which have improved the article.

Author Contributions: Zugang Chen is the leading author of this work. He conceived the core ideas and carried out the implementation. Jia Song revised the paper. Yaping Yang is the deputy supervisor of Zugang Chen and she offered the experiment platform. They gave substantial contributions to the design and analysis of this work and to the critical review of the article.

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

References

1. Xu, C.; Yang, C. Introduction to big geospatial data research. *Ann. GIS* **2014**, *20*, 227–232. [[CrossRef](#)]
2. Guo, H.; Liu, Z.; Jiang, H.; Wang, C.; Liu, J.; Liang, D. Big earth data: A new challenge and opportunity for digital earth's development. *Int. J. Digit. Earth* **2017**, *10*, 1–12. [[CrossRef](#)]
3. Li, W.; Goodchild, M.F.; Raskin, R. Towards geospatial semantic search: Exploiting latent semantic relations in geospatial data. *Int. J. Digit. Earth* **2012**, *7*, 1–21. [[CrossRef](#)]
4. Coyle, K. Understanding metadata and its purpose. *J. Acad. Librariansh.* **2005**, *31*, 160–163. [[CrossRef](#)]
5. Baca, M.; Gilliland, A.J.; Gill, T.; Woodley, M.S.; Whalen, M. *Introduction to Metadata*; Getty Research Institute: Los Angeles, CA, USA, 2008.
6. Goodchild, M.F.; Fu, P.; Rich, P. Sharing geographic information: An assessment of the geospatial one-stop. *Ann. Assoc. Am. Geogr.* **2007**, *97*, 250–266. [[CrossRef](#)]
7. Wang, G.; Sun, C.; He, X.; Zhu, J.; Chen, X. A study on meteorological metadata catalogue service system. *J. Geogr. Inf. Sci.* **2009**, *11*, 24–29.
8. Goodwin, J.; Dolbear, C.; Hart, G. Geographical linked data: The administrative geography of great britain on the semantic web. *Trans. GIS* **2008**, *12*, 19–30. [[CrossRef](#)]
9. Zhu, Y.; Zhu, A.-X.; Song, J.; Yang, J.; Feng, M.; Sun, K.; Zhang, J.; Hou, Z.; Zhao, H. Multidimensional and quantitative interlinking approach for linked geospatial data. *Int. J. Digit. Earth* **2017**, *10*, 1–21. [[CrossRef](#)]
10. Zhu, Y.; Zhu, A.-X.; Feng, M.; Song, J.; Zhao, H.; Yang, J.; Zhang, Q.; Sun, K. A similarity-based automatic data recommendation approach for geographic models. *Int. J. Geogr. Inf. Sci.* **2016**, *31*, 1403–1424. [[CrossRef](#)]
11. Ristoski, P.; Mencía, E.L.; Paulheim, H. A hybrid multi-strategy recommender system using linked open data. *Comm. Comput. Inf. Sci.* **2014**, *475*, 150–156.
12. Zhao, H.; Zhu, Y.; Hou, Z.; Yang, H. Construction of geospatial metadata association network. *Sci. Geogr. Sin.* **2016**, *36*, 1180–1189.
13. Zhao, H.; Zhu, Y.; Yang, H.; Luo, K. The semantic relevancy computation model on essential features of geospatial data. *Geogr. Res.* **2016**, *35*, 58–70.
14. FGDC. Content Standard for Digital Geospatial Metadata. Available online: https://www.fgdc.gov/standards/projects/metadata/base-metadata/v2_0698.pdf (accessed on 20 December 2017).
15. ISO-19115. *Geographic Information—Metadata*; International Organization for Standardization: Geneva, Switzerland, 2003.
16. Martins, B.; Silva, M.J.; Andrade, L. Indexing and ranking in geo-ir systems. In Proceedings of the 2005 Workshop on Geographic Information Retrieval, Bremen, Germany, 4 November 2005; pp. 31–34.
17. Hu, Y.-H.; Ge, L. Learning ranking functions for geographic information retrieval using genetic programming. *J. Pract. Inf. Technol.* **2009**, *41*, 39–52.
18. De Andrade, F.G.; Baptista, C.D.S., Jr.; Davis, C.A. Improving geographic information retrieval in spatial data infrastructures. *Geoinformatica* **2014**, *18*, 793–818. [[CrossRef](#)]
19. Daoud, M.; Huang, J.X. Mining query-driven contexts for geographic and temporal search. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 1530–1549. [[CrossRef](#)]
20. Schwering, A.; Kuhn, W. A hybrid semantic similarity measure for spatial information retrieval. *Spat. Cogn. Comput.* **2009**, *9*, 30–63. [[CrossRef](#)]
21. Bartell, B.T.; Cottrell, G.W.; Belew, R. Automatic combination of multiple ranked retrieval systems. In Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland, 3–6 July 1994; pp. 173–181.
22. Trotman, A. Learning to rank. *Inf. Retr.* **2005**, *8*, 359–381. [[CrossRef](#)]

23. Lacasta, J.; Lopez-Pellicer, F.J.; Espejo-García, B.; Nogueras-Iso, J.; Zarazaga-Soria, F.J. Aggregation-based information retrieval system for geospatial data catalogs. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1583–1605. [[CrossRef](#)]
24. Al-Bakri, M.; Fairbairn, D. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1437–1456. [[CrossRef](#)]
25. Li, Y.; Bandar, Z.A.; McLean, D. An approach for measuring semantic similarity between words using multiple information sources. *IEEE Trans. Knowl. Data Eng.* **2003**, *15*, 871–882.
26. Yadav, S.; Sangwan, O.P. Neural network based approach for predicting user satisfaction with search engine. *Int. J. Comput. Appl.* **2011**, *18*, 16–21. [[CrossRef](#)]
27. Zeynelgil, H.L.; Demiroren, A.; Sengor, N.S. The application of ann technique to automatic generation control for multi-area power system. *Int. J. Electr. Power* **2002**, *24*, 345–354. [[CrossRef](#)]
28. Karmakar, P.; Roy, B.; Paul, T.; Manna, S. Target classification: An application of artificial neural network in intelligent transport system. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2012**, *2*, 242–245.
29. Li, X.; Chen, W.; Cheng, X.; Wang, L. A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ziyuan-3 stereo satellite imagery. *Remote Sens.* **2016**, *8*, 514. [[CrossRef](#)]
30. Rehman, S.; Mohandes, M.A. Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy* **2008**, *36*, 571–576. [[CrossRef](#)]
31. Zhang, L.; Yan, W.; Zheng, W.; Liu, Y.; Liang, X.; Gao, C.; Fang, X. Response of camphor forest soil respiration to nitrogen addition determined based on the ga-bp network. *Acta Ecol. Sin.* **2017**, *37*, 1–11.
32. Bengio, Y. Learning deep architecture for ai. *Found. Trends Mach. Learn.* **2009**, *1*, 1–127. [[CrossRef](#)]
33. Li, W.; Raskin, R.; Goodchild, M.F. Semantic similarity measurement based on knowledge mining: An artificial neural net approach. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1–21. [[CrossRef](#)]
34. Buscema, M. Back propagation neural networks. *Subst. Use* **1998**, *32*, 233–270. [[CrossRef](#)]
35. Neuroph. Available online: <http://neuroph.sourceforge.net/index.html> (accessed on 1 January 2018).
36. Corley, C.; Mihalcea, R. Measuring the semantic similarity of texts. In Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment, Ann Arbor, MI, USA, 30 June 2005; pp. 13–18.
37. Using Wordnet-Based Context Vectors to Estimate the Semantic Relatedness of Concepts. Available online: https://www.researchgate.net/publication/233779111_Using_WordNet-based_Context_Vectors_to_Estimate_the_Semantic_Relatedness_of_Concepts (accessed on 20 December 2017).
38. GCMD Keywords, Version 8.1. Available online: <https://gcmd.nasa.gov/learn/keywords.html> (accessed on 20 December 2017).
39. Wu, Z.; Palmer, M. Verb semantics and lexical selection. In Proceedings of the Annual Meeting on Association for Computational Linguistics, Las Cruces, NM, USA, 24 June 1994; pp. 133–138.
40. Categorizing Binary Topological Relations between Regions, Lines, and Points in Geographic Databases. Available online: <https://www.semanticscholar.org/paper/Categorizing-Binary-Topological-Relations-Between-Egenhofer-Herring/b30339af3f0be6074f7e6ac0263e9ab34eb84271?tab=abstract> (accessed on 20 December 2017).
41. Saaty, T.L. How to make a decision: The analytic hierarchy process. *Eur. J. Oper. Res.* **1990**, *48*, 9–26. [[CrossRef](#)]
42. Allen, J.F. Maintaining knowledge about temporal intervals. In *Qualitative Reasoning about Physical Systems*; Daniel, S.W., Kleer, J.D., Eds.; Morgan Kaufmann Publishers: San Francisco, CA, USA, 1983; Volume 26, pp. 361–372.
43. Lebreton, J.; Burgess, J.R.D.; Kaiser, R.B.; Atchley, E.K.P.; James, L.R. The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organ. Res. Methods* **2003**, *6*, 80–128. [[CrossRef](#)]
44. James, L.R.; Demaree, R.G.; Wolf, G. An assessment of within-group interrater agreement. *J. Appl. Psychol.* **1993**, *78*, 306–309. [[CrossRef](#)]
45. Rodgers, J.; Nicewander, A. Thirteen ways to look at the correlation coefficient. *Am. Stat.* **1988**, *42*, 59–66. [[CrossRef](#)]
46. Kendall, M.G.; Smith, B.B. The problem of m rankings. *Ann. Math. Stat.* **1939**, *10*, 275–287. [[CrossRef](#)]
47. Ballatore, A.; Bertolotto, M.; Wilson, D.C. An evaluative baseline for geo-semantic relatedness and similarity. *Geoinformatica* **2014**, *18*, 747–767. [[CrossRef](#)]

48. Mann, H.B.; Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1947**, *18*, 50–60. [[CrossRef](#)]
49. Maier, H.R.; Dandy, G.C. Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environ. Model. Softw.* **2000**, *15*, 101–124. [[CrossRef](#)]
50. Hecht-Nielsen, R. Kolmogorov's mapping neural network existence theorem. In Proceedings of the IEEE First Annual International Conference on Neural Networks, San Diego, CA, USA, 21–24 June 1987; pp. 11–13.
51. Flood, I.; Kartam, N. Neural networks in civil engineering: Principles and understanding. *J. Comput. Civil Eng.* **1994**, *8*, 131–148. [[CrossRef](#)]
52. Tamura, S.; Tateishi, M. Capabilities of a four-layered feedforward neural network: Four layers versus three. *IEEE Trans. Neural Netw.* **1997**, *8*, 251–255. [[CrossRef](#)] [[PubMed](#)]
53. Rogers, L.L.; Dowla, F.U. Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. *Water Resour. Res.* **1994**, *30*, 457–481. [[CrossRef](#)]
54. Bebis, G.; Georgiopoulos, M. Feed-forward neural networks: Why network size is so important. *IEEE Potential* **1994**, *4*, 27–31. [[CrossRef](#)]
55. Resnik, P. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *J. Artif. Intell. Res.* **1999**, *11*, 95–130.
56. Rada, R.; Mili, H.; Bicknell, E.; Blettner, M. Development and application of a metric on semantic nets. *IEEE Trans. Syst. Man Cybern.* **1989**, *19*, 17–30. [[CrossRef](#)]
57. Amini, J. Optimum learning rate in back-propagation neural network for classification of satellite images (irs-1d). *J. Electrocardiol.* **2008**, *15*, 558–567.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).