

Article

A Knowledge Graph Convolutional Networks Method for Countryside Ecological Patterns Recommendation by Mining Geographical Features

Xuhui Zeng ^{1,†}, Shu Wang ^{2,†} , Yunqiang Zhu ², Mengfei Xu ¹ and Zhiqiang Zou ^{1,3,*} 

¹ College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

² Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing 100101, China

³ Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

* Correspondence: zouzq@njupt.edu.cn

† These authors contributed equally to this work.

Abstract: The recommendation system is one of the hotspots in the field of artificial intelligence that can be applied to recommend suitable ecological patterns for the countryside. Countryside ecological patterns mean advanced patterns that can be recommended to those developing areas which have similar geographical features, which provides huge benefits for countryside development. However, current recommendation methods have low recommendation accuracy due to some limitations, such as data-sparse and ‘cold start’, since they do not consider the complex geographical features. To address the above issues, we propose a geographical Knowledge Graph Convolutional Networks method for Countryside Ecological Patterns Recommendation (KGCN4CEPR). Specifically, a geographical knowledge graph of countryside ecological patterns is established first, which makes up for the sparsity of countryside ecological pattern data. Then, a convolutional network for mining the geographical similarity of ecological patterns is designed among adjacent countryside, which effectively solves the ‘cold start’ problem in the existing recommended methods. The experimental results show that our KGCN4CEPR method is suitable for recommending countryside ecological patterns. Moreover, the proposed KGCN4CEPR method achieves the best recommendation accuracy (60%), which is 9% higher than the MKR method and 6% higher than the RippleNet method.

Keywords: recommendation system; geospatial artificial intelligence; geographical knowledge graph; knowledge graph convolutional networks; countryside ecological patterns



Citation: Zeng, X.; Wang, S.; Zhu, Y.; Xu, M.; Zou, Z. A Knowledge Graph Convolutional Networks Method for Countryside Ecological Patterns Recommendation by Mining Geographical Features. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 625.

<https://doi.org/10.3390/ijgi11120625>

Academic Editors: Wolfgang Kainz and Danielle Ziebelin

Received: 31 October 2022

Accepted: 13 December 2022

Published: 15 December 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Countryside ecological patterns are exemplary ecological patterns that can guide people to reasonably develop production as well as consider local geographical features and protect the geographical environment. Generally speaking, CEP could be summarized into ten main patterns, among which industrial development pattern and leisure tourism pattern are two typical patterns. CEP is able to be described in a hierarchical way, such as the top level consisting of basic information, economic information and ecological information, the above three kinds of information could be further divided into fine-grained information on the second level. Recommending suitable countryside ecological patterns for a specific countryside, which means providing a pattern with similar features of geography, society, economy, and culture, is a realistic guide to the rapid and sustainable development of the countryside. However, the countryside ecological pattern involves many complex features, including human and natural geographical features related to ecology and geography. The former features include drought, climate, water density, and other basic information, while the latter features contain the relevant topography, landform, industrial structure, and

cultural customs information. How to make use of these human and natural geographical features in a recommendation system is challenging and meaningful work.

Referring to the current mainstream recommendation methods, there are three types of methods for countryside ecological patterns recommendation: content-based recommendation, collaborative filtering recommendation, and hybrid recommendation [1–3]. The content-based recommendation is a classic method, which calculates the similarity between two sets of items and sorts them into a recommendation list according to their similarity. The collaborative filtering method [4] takes advantage of a user's historical behavioral preference data to build a model to make recommendations by exploring the correlation between user and item. To further improve the recommendation accuracy, hybrid recommendation [5] tries to combine some AI algorithms with other auxiliary information (e.g., combining convolutional networks with knowledge graph). Among the three aforementioned methods, the optimal one is the hybrid recommendation method, which can achieve the best performance in film, music, and book recommendations. However, limitations still exist since they do not consider the geographical spatial features. Specifically, under the scenario of countryside ecological patterns recommendation, there still exist two main problems, one is the sparse problem and the other is the 'cold start' problem. The former makes it difficult to accurately calculate the similarity of these sparse features due to not enough features, and the latter leads to a low recommendation recall rate since the similarity cannot be calculated without historical records related to the geographical environment.

To address the above limitations, we introduce the Knowledge Graph Convolutional Network (KGCN) and establish the geographical knowledge graph of countryside ecological patterns (countryside knowledge graph, for short) through the ontology and entity construction of countryside ecological patterns. We propose a Knowledge Graph Convolutional Networks method for Countryside Ecological Patterns Recommendation (KGCN4CEPR). Specifically, our paper has four main points of contribution:

- (1) We establish a geographical knowledge graph of countryside ecological patterns by mining geographical features, which makes up for the sparsity of countryside ecological pattern data through the rich semantic information of the knowledge graph. The specific relationships and countryside ecological pattern embedding are weighted and calculated, so that the geographical personalized features of the countryside ecological pattern are effectively represented.
- (2) We design a convolutional network for sufficiently mining the geographical similarity of ecological patterns, which effectively solves the 'cold start' problem. The spatial features of neighborhood information are exploited through the neighborhood aggregation operation of convolutional networks.
- (3) We explore the geographical relationship features between the countryside and the countryside ecological pattern by considering the spatial scale of the neighborhood so that our method is more suitable for the recommended work under the countryside ecological pattern scenarios.
- (4) For the convenience of other researchers, we have published the code and dataset in this project on the Internet (<https://github.com/973866103/KGCN4CEPR>, accessed on 21 June 2022).

In general, the KGCN4CEPR method achieves the probability that the countryside suits a countryside ecological pattern based on the features in the knowledge graph related to the geographical environment. The countryside ecological pattern with a calculated probability closer to 1 is more likely to be recommended.

The rest of this paper is organized as follows. Section 2 reviews related work. The KGCN4CEPR method is presented in Section 3. Section 4 describes the experimental evaluations. Finally, our conclusions follow in Section 5.

2. Related Work

In this section, we will briefly introduce the ecological pattern, then we focus on the aforementioned methods related to recommendation systems.

2.1. Three Main Categories of Ecological Pattern

According to the different elements involved in the construction of ecological patterns, ecological patterns could be summarized into three categories: government-driven patterns, market-driven patterns, and individual consciousness patterns [6].

A government-driven pattern promotes the construction, implementation and expansion of the ecological pattern through the administrative power of the government, which relies on the administrative system. This pattern is the main pattern for the construction of ecological civilization in China at present, whose advantage is that it can make full the strength of China's central and local hierarchical management system. However, through the 'top-down' administrative force, there still is a conflict between local and national interests and it needs to be further optimized [7].

A market-driven pattern is a construction pattern widely applied in developed countries. It uses market mechanisms to regulate the dynamic balance of environment and resources, so as to achieve the simultaneous development of economy and ecology. Compared with the government-driven pattern, this pattern is considered as an effective way to construct ecological patterns because of its endogeneity. However, China's market mechanism and environmental economic policies are still inadequate; thus, how to fully exploit the endogeneity of the market-driven pattern is still a problem [8].

An individual consciousness pattern requires the full participation of each individual in society, which relies on the conscious action of society. When an individual's concept of ecological civilization is fully constructed, the individual conscious pattern will have a lasting momentum. The assumption of this pattern is that every individual in society needs to be actively involved. Nevertheless, our ecological education and propaganda are not sufficient so that it cannot meet this assumption well [9].

In summary, the government-driven pattern is dominant among three patterns in China. Therefore, we focus on government-driven patterns and provide countryside ecological patterns recommendations for government decisions and support.

2.2. Recommendation Systems

Recommendation is a critical, complex and challenging issue in artificial intelligence research. The existing recommendation methods could be divided into content-based methods, collaborative filtering methods, and hybrid methods [10].

2.2.1. Content-Based Recommendation

The proposal of the content-based recommendation method [11–13] originated from information retrieval research and is a kind of widely used traditional recommendation system. Content-based recommendation method mainly involves two types of data: content features of items and user profiles (interests). Its basic idea is to calculate the similarity between the items in a subset and the items in the whole set. The former is selected from the whole set according to explicit feedback (ratings, likes, etc.) and implicit feedback (search, click, buy, etc.). Then, this method sorts them into a recommendation list according to their similarity.

The disadvantage is that it relies heavily on domain experts and complex feature engineering when constructing content features and user profiles. In addition, the content-based recommendation method suffers from the 'cold start' problem for new users. For example, in the application of countryside ecological patterns recommendation, for a specified new countryside, it is difficult to accurately calculate the similarity since collecting massive and rich data of countryside in China is almost impossible. Therefore, it is impossible to directly use the content-based recommendation method to solve the problem of countryside ecological patterns recommendation.

2.2.2. Collaborative Filtering Recommendation

Collaborative filtering [14] recommendation is another widely used traditional recommendation system, which can provide interested content for a specific user. It firstly finds the users who are similar to this user, and then recommend the content to this user according to those users' interests. The collaborative filtering recommendation method can be further classified into memory-based collaborative filtering recommendation and model-based collaborative filtering recommendation. The former can be subdivided into item-based collaborative filtering and user-based collaborative filtering. The latter mainly includes SVD [4] and its variant, which can map data to a low dimension and calculate the similarities between items.

However, the collaborative filtering recommendation method still suffers from a data-sparse problem. In other words, when the interaction matrix of items and users is a sparsity matrix, it is difficult to find users' neighborhood information in the sparsity matrix, which leads to a lower recommendation accuracy. In general, the larger the data size, the smaller the item overlap and the sparser the matrix. In addition, it also has a 'cold start' problem. It cannot recommend a new product without historical data by collecting a user's behavior, such as browsing, clicking or buying. For example, under the scenario of countryside ecological patterns recommendation, the interaction matrix of countryside and patterns is sparse, since it is difficult to find enough natural geographical features and human features of the countryside. As a result, the collaborative filtering recommendation method is also not appropriate for countryside ecological patterns recommendation.

2.2.3. Hybrid Recommendation

To further improve the recommendation performance, especially in the case of sparse data and 'cold start', researchers proposed the hybrid recommendation method, whose core idea is to combine recommendation methods with auxiliary information. The sources of this auxiliary information are diverse, of which the more prominent is the knowledge graph [15,16]. Knowledge graph is a directed heterogeneous graph where the node represents entity and the edge represents relation [17], so that it can provide rich semantic relations between entities. Compared with KG-free methods, incorporating KG into recommendation benefits the results. The rich semantic relatedness among items in a KG can help explore their latent connections and improve the precision of results. Therefore, the KG-based hybrid recommendation methods have better performance than the KG-free methods.

RippleNet [9] is a KG-based hybrid recommendation method, which first fused path-based method and KG. MKR [18] is another KG-based hybrid recommendation method, which combined multi-task feature learning with KG. KGCN [19] aggregated neighborhood information in KG by using convolutional networks. The KGCN is chosen as a basic model due to its advantages in handling the sparsity and spatial problem. It achieved the best performance in film, music, and book recommendation among the three above methods [9,18,19]. However, these methods are not competent under the scenario with human and natural geographical features. In other words, under the scenario of countryside ecological patterns recommendation, the KGCN have a low accuracy since they do not consider the geographical spatial adjacency.

In fact, the geographical spatial adjacency refers to the Laws of Geography, "near things are more related to each other" and "The more similar geographic configurations of two areas, the more similar the values (processes) of the target variable at these two areas" [20,21]. Specifically, the closer countryside is in spatial location will have more similar geographical features and consequently have similar countryside ecological patterns. Therefore, geographical spatial adjacency is a key factor of geographical application, which should be considered in recommendation systems. If the geographical spatial adjacency can be used in recommendation systems, it could greatly improve the performance of the current recommendation method in geographical recommendation issues.

3. KGCN4CEPR Method

In this section, we describe the proposed KGCN4CEPR method in two parts. For the first part, we extend the KGCN recommendation method by establishing the geographical knowledge graph related to geographical environment through the ontology and entity construction of countryside ecological patterns. To compensate for sparsity and improve the performance of the recommendation system, we utilize the countryside knowledge graph so that the properties of countryside ecological patterns and countryside entities become richer than before. This is because the knowledge graph links up information with each other, improving the correlation of data. For the second part, we exploit convolutional networks to mine geographical spatial adjacency between neighborhood countryside areas in geographical space [20,21]. Meanwhile, we aggregated the neighbors for each entity in the knowledge graph and combine neighborhood information with bias when calculating the representation of a given entity. Convolutional network sufficiently enhances the prediction by choosing the most important information of the knowledge graph. In order to make it more suitable for the work of recommending the countryside ecological pattern in the geographical domain, we further explore the geographical relationship features between the countryside and the countryside ecological pattern.

3.1. Construction of Countryside Knowledge Graph

The construction of knowledge graph can be divided into two categories: top-down and bottom-up [22]. The top-down builds domain ontology of the knowledge graph first, extracting ontology and schema information from high-quality data sources, such as industry domains and encyclopedic websites. The bottom-up starts from the bottom layer entity, extracting new patterns by means of manual search and web crawlers.

Regarding the construction of the countryside knowledge graph, we use a combination of alternating top-down and bottom-up. We firstly construct the ontology base using the top-down method, and then extend the knowledge graph entities by extracting knowledge based on the bottom-up method. In this section, we will introduce the construction of ontology and entity of the countryside knowledge graph, respectively.

3.1.1. Ontology Construction of Countryside Knowledge Graph

We collect and organize a large amount of scientific literature, encyclopedic knowledge and government websites on countryside ecological patterns through manual search and web crawler. We sort out the ontological information of countryside ecological patterns and construct the ontology framework of countryside ecological patterns by using the Protégé tool, as shown in Figure 1. In Figure 1, the boxes with light color mean classes, while the boxes with deep color mean instances.

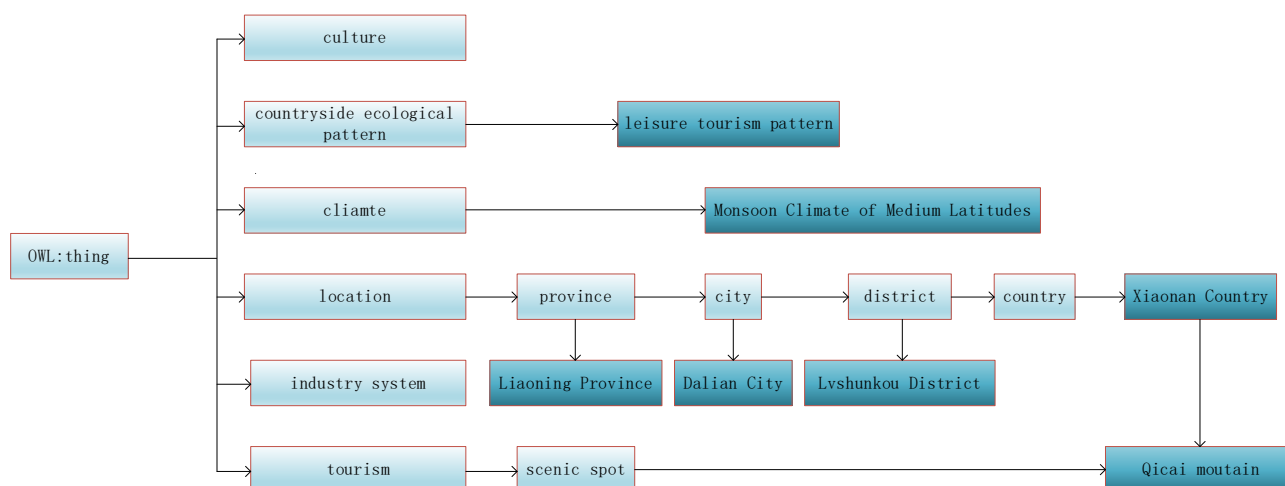


Figure 1. Ontology framework of countryside ecological pattern.

In order to provide guidance for the effective construction of the entity, we construct a countryside ecological patterns knowledge system according to three parts, i.e., ecological pattern concept, hierarchical information for describing pattern concept, and fine-grained information, as shown in Figure 2. For example, *Hancunhe* Countryside is suitable to develop *industrial development pattern*, its longitude and latitude is (115.961, 39.603), population is 2700, area size is 2.4 km², and road length is 1328.74 km.

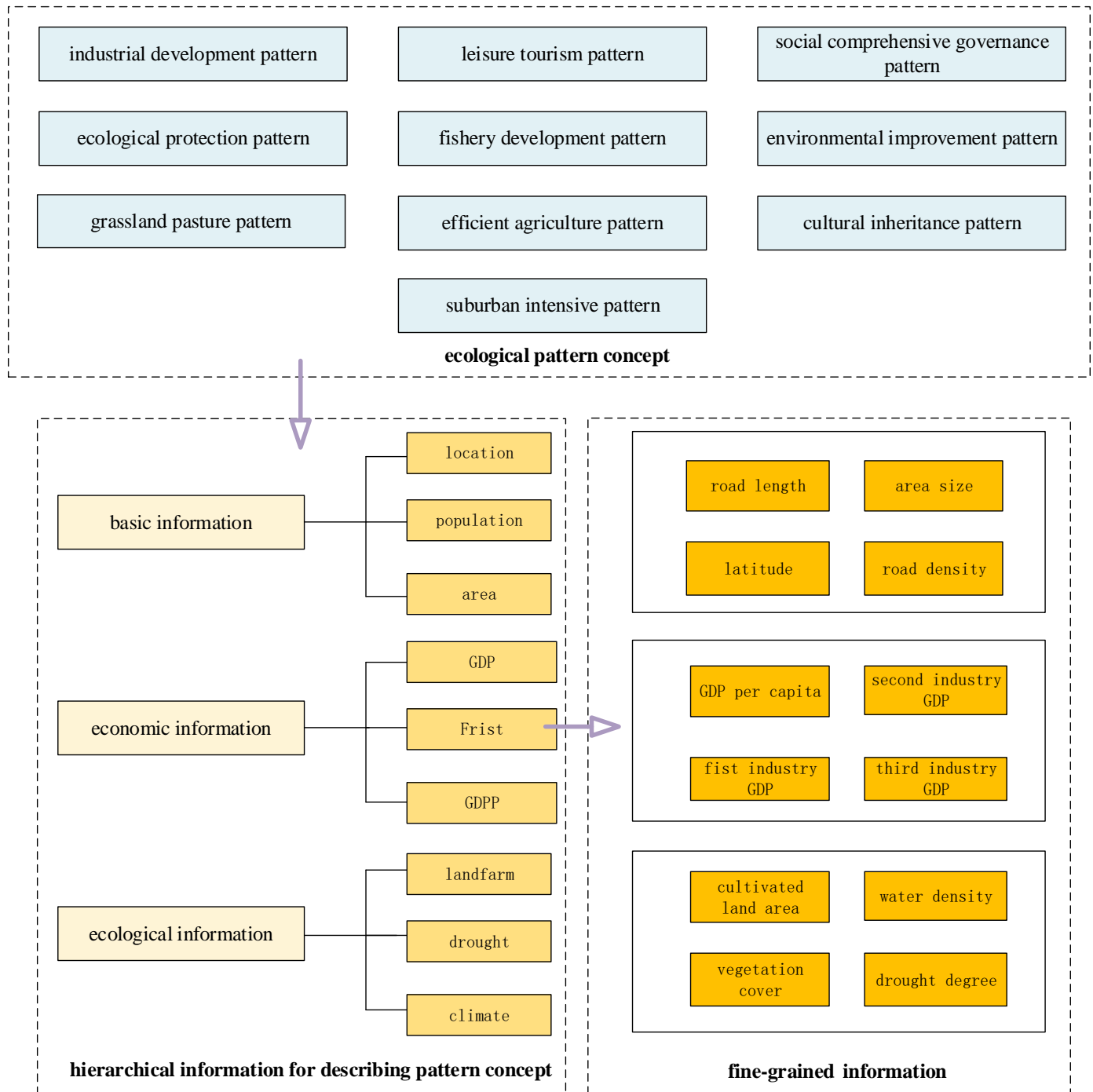


Figure 2. Knowledge system of countryside ecological pattern.

It can be seen that there are ten major countryside ecological patterns as follows: industrial development pattern, ecological protection pattern, suburban intensive pattern, social comprehensive governance pattern, cultural inheritance pattern, fishery develop-

ment pattern, grassland pasture pattern, environmental improvement pattern, efficient agriculture pattern, and leisure tourism pattern. Each pattern has some hierarchical relationships, which could be divided into basic, economic and ecological information. We will recommend one of the above-mentioned countryside ecological patterns for the specified countryside as suitable.

3.1.2. Entity Construction of Countryside Knowledge Graph

We collect and organize a large amount of structured, semi-structured and unstructured data to extend the entities of the countryside knowledge graph related to the geographical environment, according to Figure 2. Our data come from two main sources, i.e., Internet data and official statistics.

A portion of the data comes from the Internet. There is a lot of fragmented information on the Internet, and we collect data from scientific literature, encyclopedic knowledge, news reports, social media and other information by means of manual searches and web crawlers. For example, one instance of *Baidu* encyclopedia infobox is about the information of the *Hancunhe* Countryside, as shown in Figure 3. The *Baidu* encyclopedia is one of the sources of our countryside knowledge graph. We manually retrieve the name of the countryside and obtain the corresponding infobox information from the *Baidu* encyclopedia. We selected the category and filtered the contents according to Figure 2, such as the category of “basic information” and contents of “population and location” are from the *Baidu* encyclopedia. The collected information is further mapped to entity–relation–entity triples, such as (*Hancunhe* Countryside, countryside.locate.district, *Fangshan* District) and (*Hancunhe* Countryside, countryside.population, 2700).

Chinese Name	韩村河村	Area Code	010
English Name	Hancunhe Countryside	Zip Code	102488
Another Name	韩村河	Climate	Monsoon Climate of Medium Latitudes
Region category	Countryside	Population	2700
Region	North China	Airport	Beijing Capital International Airport
Location	N39° E115°	Train Station	Fangshan Station
Area	2.4 km ²	license plate	京0、1、2
		dialect	Beijing dialect

Figure 3. Encyclopedia infobox information of *Hancunhe* Countryside, *Fangshan* District, *Beijing*.

Another portion of the data comes from the township statistics of the China Statistical Yearbook (<https://data.cnki.net/yearbook/Single/N2021110004>, accessed on 15 March 2021) database, including the data of the township’s population, economy, area and so on. We map these data to the corresponding triples in two ways. On the one hand, we select important attributes related to the development of the countryside ecological patterns. On the other hand, we use the data with larger spatial scale to supplement the missing data with smaller spatial scale for the approximate attributes.

Finally, we create the countryside knowledge graph by mapping triples from the above data on a popular KG platform (Neo4j), which has 441 entities and 22 relationships, as shown in Figure 4.



Figure 4. Partial countryside knowledge graph.

3.2. KGCN4CEPR Recommendation Method

In this section, we firstly formulate the countryside ecological patterns recommendation problem, then present the implementation of KGCN4CEPR based on the countryside knowledge graph established in Section 3.1. Specifically, we use convolutional networks to mine the spatial adjacency between entities in the countryside knowledge graph and extract the features of the relationship between the countryside and the countryside ecological pattern, so that we could scientifically recommend the countryside ecological pattern for the specified countryside.

3.2.1. Problem Formulation

We formulate the countryside ecological patterns recommendation problem as follows. In the countryside ecological patterns recommendation scenario, we have a set of M patterns $P = \{p_1, p_2, \dots, p_M\}$ and a set of N countryside $V = \{v_1, v_2, \dots, v_N\}$. The pattern–countryside interaction matrix $Y = \{y_{p_1, v_1}, y_{p_1, v_2}, \dots, y_{p_M, v_N}\} \in \mathbb{R}^{M \times N}$ is defined depending on the real situation, where $y_{p, v} = 1$ indicates that countryside v is suitable for pattern p ; otherwise, $y_{p, v} = 0$. Additionally, we also have a countryside knowledge graph \mathcal{G} , which is comprised of entity–relation–entity triples (h, r, t) . Here $h \in \xi$, $r \in \mathcal{R}$ and $t \in \xi$ denote the head, relation, and tail of \mathcal{G} , respectively. Given an example, the triple (Xiaonan Countryside, countryside.locate.district, Lvshunkou District) means the fact that the Xiaonan Countryside is located in Lvshunkou District.

Given the pattern–countryside interaction matrix Y and the countryside knowledge graph \mathcal{G} , we purpose to predict whether countryside v matches pattern p . Our aim is to

learn a predication function $\hat{y}_{p,v}$, minimizing the difference between the predicted value and the true value:

$$\hat{y}_{p,v} = \mathcal{F}(p, v | \Theta, Y, \mathcal{G}) \quad (1)$$

where $\hat{y}_{p,v}$ denotes the possibilities that countryside v will match pattern p , and Θ denotes the parameters of function \mathcal{F} . The meanings of the symbols in KGCN4CEPR are shown in Table 1.

Table 1. The meanings of the symbols in KGCN4CEPR.

Variable	Meaning
P	Collections of patterns
V	Collections of countryside
p	An entity in P
v	An entity in V
\bar{v}	The vector of countryside entity v
$\bar{\bar{v}}$	The final vector of countryside entity v
$N(v)$	The set of all entities directly connecting to v
\mathcal{G}	Countryside knowledge graph
$\hat{y}_{p,v}$	Predicted value that countryside v suitable for pattern p
$y_{p,v}$	True value that countryside v suitable for pattern p
Θ	Parameters of function \mathcal{F}
$w_{r_i}^p$	The importance of relation r_i on pattern p
$g(p, r_i)$	The inner product function of relation r_i and pattern p
$\bar{w}_{r_i}^p$	The normalized $w_{r_i}^p$
e_i	The i th representation of neighborhood vector
k	Neighborhood aggregation times
W	Linear transformation matrix
b	Bias of the stitching aggregation
L	Loss function of $y_{p,v}$ and $\hat{y}_{p,v}$

3.2.2. Implementation of KGCN4CEPR Recommendation Method

This subsection first describes the framework of KGCN4CEPR and follows the four key implementation steps, as shown in Figure 5.

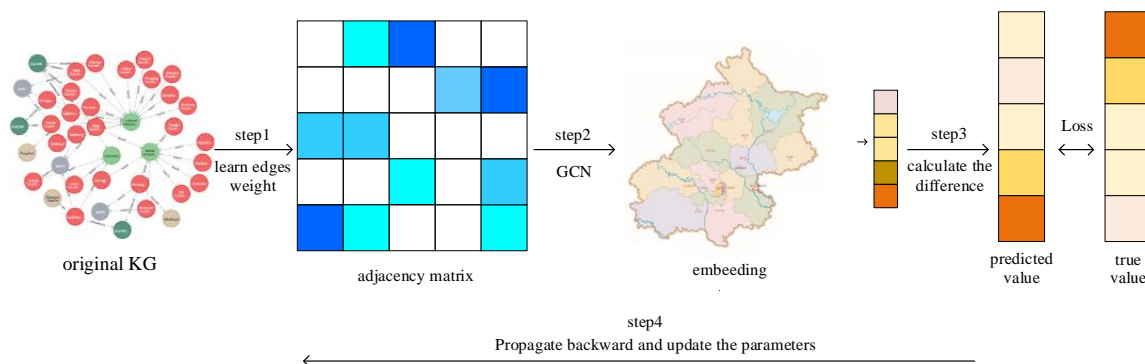


Figure 5. Framework of KGCN4CEPR.

Step 1. Weights adjacency matrix: Convert the original heterogeneous countryside knowledge graph into a countryside knowledge graph with weights, learn the weights of the edges $w_{r_i}^p$ to obtain an adjacency matrix.

Step 2. \bar{v} construction: Mine the geographical spatial adjacency features of countryside by graph convolutional networks, and then embed the geographical spatial adjacency information into the countryside entities v to obtain neighborhood vector \bar{v} , and finally aggregate into a countryside entity vector $\bar{\bar{v}}$.

Step 3. Iteration: Propagate backward and forward the error gradient to update the parameters by using the gradient descent algorithm and repeat the training several times so that the difference between the predicted value and the true value is minimized.

Step 4. Recommendation: Based on the minimized result in step 3, predict whether the countryside v matches pattern p .

In step 1, for the calculation of the countryside knowledge graph edge weights, we use Formula (2):

$$w_{r_i}^p = g(p, r_i) \quad (2)$$

where p denotes countryside ecological pattern vector, and r_i denotes the i th neighborhood relationship vector. The function $g: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is used to calculate the inner product of vector p and vector r_i . The weight is calculated by inner product function g . Take *industrial development pattern* as an example, according to the scoring of this pattern in the pattern–countryside interaction matrix, we can obtain countryside ecological pattern vector p . Correspondingly, we can find neighborhood relationship vector r_i on the basis of relationship linking with this pattern in the knowledge graph. In general, $w_{r_i}^p$ indicates the importance of relation r_i on pattern p . The results of importance are accurate since each calculation of $w_{r_i}^p$ takes into account the influence of relation r_i on pattern p instead of using a constant influence of relation.

Next, according to Formula (3), we perform the softmax normalization on $w_{r_i}^p$:

$$\tilde{w}_{r_i}^p = \text{softmax}(w_{r_i}^p) = \frac{\exp(w_{r_i}^p)}{\sum_{j \in N(v)} \exp(w_{r_j}^p)} w_{r_i}^p = g(p, r_i) \quad (3)$$

where $N(v)$ denotes the set of all entities directly connecting to the countryside entity v .

Then, we calculate the linear combination \bar{v} of v 's neighborhoods to obtain the predicted value $\hat{y}_{p,v}$ in step 2:

$$\bar{v} = \sum_{i \in S(v)} \tilde{w}_{r_i}^p e_i \quad (4)$$

where e_i denotes the i th representation of neighborhoods vector, $S(v)$ denotes the range of countryside entity selection.

According to the Laws of Geography, the closer the distance, the greater the correlation between objects. Therefore, there exist certain similarities among the countryside ecological patterns of geographic adjacent countryside. When there are too many neighbors of a countryside entity v in the countryside knowledge graph, it will bring too much computational pressure and non-essential neighborhood information. Therefore, we limit the visiting spatial scale of each countryside entity's neighborhoods, instead visiting the whole path-reachable entities $S(v)$. For each countryside entity, it only visits neighborhood nodes k times and obtains subset $N(v)$, referring to Formula (5). This ensures that the aggregation of neighborhood entities does not expand outward indefinitely while keeping enough information of high relevance:

$$N(v) \leftarrow \{v | v \in S(v)\} \text{ and } |N(v)| = k \quad (5)$$

where $N(v)$ is a subset of $S(v)$.

In step 2, the aggregation from \bar{v} to $\bar{\bar{v}}$ is shown in Formula (6):

$$\bar{\bar{v}} = \sigma(W = (\bar{v} + e) + b) \quad (6)$$

where W is linear transformation matrix, b is bias, vector e is generated by the countryside entity in the previous iteration, and σ is the nonlinear function, whose result is the final vector representation of the countryside entity $\bar{\bar{v}}$.

At last, a loss function is used to calculate the difference between the predicted value $\hat{y}_{p,v}$ and the true value $y_{p,v}$, as described in steps 3 and 4, referring to Formula (7):

$$loss = L(y_{p,v}, \hat{y}_{p,v}) \quad (7)$$

In the following, we take a countryside ecological patterns recommendation for specified countryside *Hancunhe* as an example to illustrate the implementation of KGCN4CEPR, which consist of four steps.

In step 1, we firstly calculate the weight of each edge that is adjacent to *Hancunhe* in the countryside knowledge graph. The weight $w_{r_i}^p$ represents the preference of the industrial development pattern for each relationship.

In step 2, the number of times of neighborhood aggregation k is 2, where the second layer contains: “Beijing city”, “road length” and “area”. We normalize and weight the entities to obtain the aggregated vector v of the “Fangshan District”. The first layer contains “Fangshan District”, “area”, “population” and “industrial development pattern”. Similarly, we obtain the neighborhood vector \bar{v} of the central location “Hancunhe Countryside”. Then, we perform a message aggregation of the neighborhood vector \bar{v} of “Hancunhe Countryside” and the countryside entity v of “Hancunhe Countryside” to obtain the final vector representation $\bar{\bar{v}}$ of “Hancunhe Countryside”. Figure 6 describes the above aggregation operation of KGCN4CEPR. In Figure 6, there are 2 kinds of objects, i.e., entities and properties. The circles represent different entities; the rectangles represent different properties; the arrow means the direction of aggregation operation; the objects with same color denote the objects involved in the same aggregation operation.

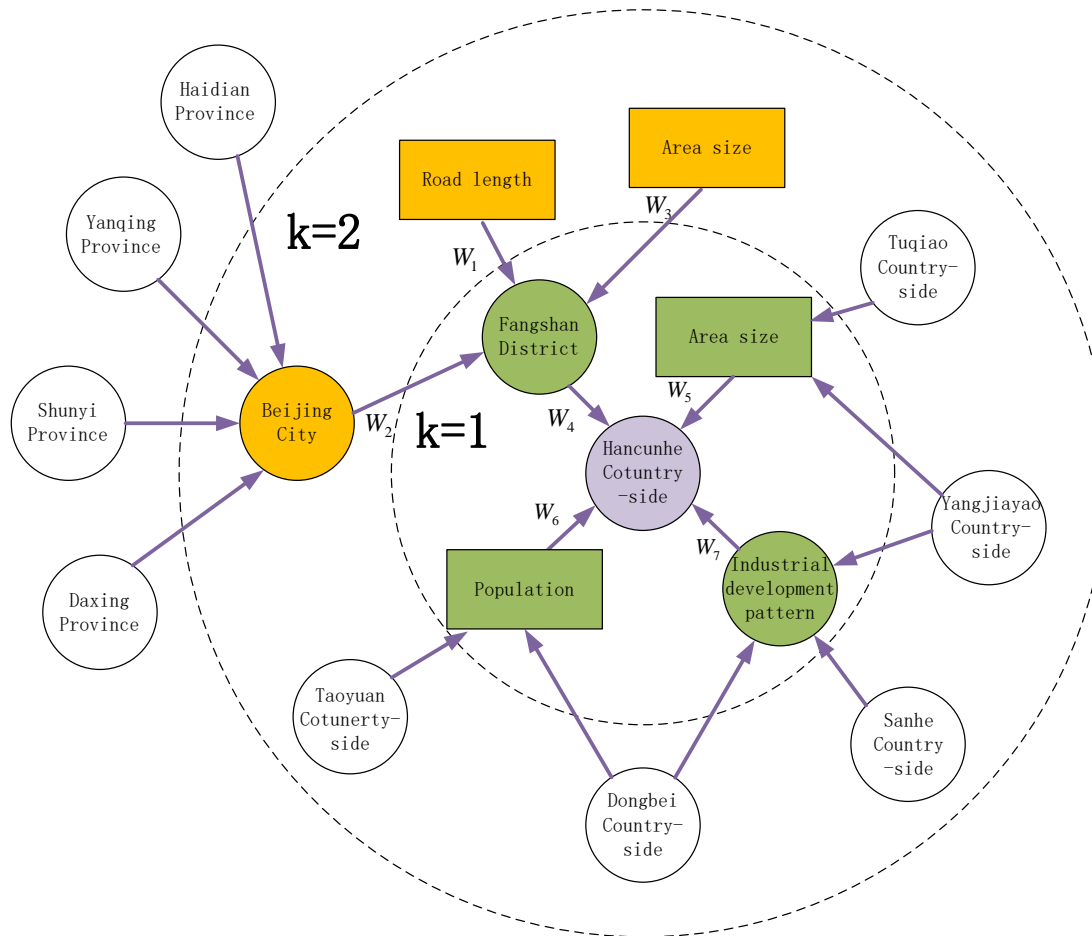


Figure 6. An example of aggregation operation in KGCN4CEPR.

In step 3, the minimum value of the difference between the predicted value and the true value of “Hancunhe Countryside” is iteratively obtained.

Finally, in step 4, the pattern corresponding to the minimized result will be recommended for the “Hancunhe Countryside”, i.e., *industrial development pattern*.

4. Experimental Evaluations and Discussion

To evaluate the performance of KGCN4CEPR, the experiment is performed based on the countryside knowledge graph.

4.1. Datasets

The dataset contains 100 countryside areas (https://www.sohu.com/a/133462746_731911, accessed on 1 November 2020), 441 entities and 22 relationships. The 441 entities contain 10 ecological patterns, 101 countryside areas and other entities. We filtered the raw data that we are interested in. The specific data we used are shown in Table 2. The large amounts of data collected are registered in the knowledge graph and input in the geographic recommendation system manually. We have many geographical elements in our model, such as water area and cultivated land area. Furthermore, we have added more geographical elements, such as countryside.locate.district, countryside.locate.city and countryside.locate.province. Through these elements, we can better enrich the geographical features of the model.

Table 2. The knowledge graph information list.

Category	Name	Description of Countryside
Basic Information	ID	Identifier
	QUXIAN	District
	SHI	City
	PROVINCE	Province
	POPULATION	Population
	ROAD	Road Length
	AREA	Area Size
Economic Information	ROADDENSITY	Road Density
	GDP	Gross Regional Product
	GDPP	GDP Per Capita
	FIRST	Primary Industry GDP
	SECOND	Second Industry GDP
	THIRD	Tertiary Industry GDP
Ecological Information	FARM	Cultivated Land Area
	GRASS	Grassland Area
	FROST	Woodland Area
	WATER	Water Area
	WATERDENSITY	Water Density
	LIVINGDENSITY	Biological Density
	VEGETATIONDENSITY	Vegetation Cover
	FROSTDENSITY	Forest Cover
	DROUGHT	Drought Degree

The ratio of training, evaluation and test set is 6:2:2. For the countryside ecological patterns interaction matrix, the threshold for each element in the matrix is 4. The matrix consists of the countryside’s rating of each of the 10 ecological patterns, the full score is 5. If an element value is greater than or equal to 4, the element is considered as a positive sample. Otherwise, it is considered as a negative sample. For the countryside knowledge graph, it holds the interactions between entities, where the entities include countryside and some countryside related information, for example: Xiaonan Countryside, countryside.locate.district, Lvshunkou District.

4.2. Experiment

4.2.1. Experiment Setup and Evaluation Criterion

Experimental environment: The hardware system utilized in this study contains: an AMD Ryzen 5 3600 CPU and a Nvidia GeForce RTX 2070 SUPER GPU. Software environment includes Python 3.6, TensorFlow 1.12.0 and Numpy 1.14.3. We use the average value after ten replicates for evaluation.

Experimental parameters: We optimize the set to determine the experimental parameters according to the testing experimental results (see Section 5 for detail). When training the KGCN4CEPR, we set the iteration number H to 2, dimension of embedding d to 8, neighbor aggregation time k to 2, batch size to 16, learning rate to $1e-3$. Finally, they repeatedly run until overfit and then stop.

Evaluation criterion: We compare the performance of our method with other baselines using the AUC as an evaluation metric [23,24]. Besides the above AUC , the common $Accuracy$, $Recall$ and $F1$ are also used as evaluation metrics.

AUC is defined as Area Under the receiver operating characteristic curve (ROC), we use AUC value as the evaluation standard of the method instead of ROC curve, because AUC can clearly explain which classifier is better [23].

$Accuracy$ is the most primitive evaluation metric in classification problems. $Accuracy$ is defined as the percentage of correct predicted outcomes over the total sample and is given by Formula (8), where TP means true positive, TN means true negative, FP denotes false positive and FN means false negative, respectively:

$$ACC = (TP + TN) / (TP + TN + FP + FN) \quad (8)$$

$Recall$ represents the percentage of positive samples that are completely identified, and the formula for $Recall$ is as follows:

$$Recall = TP / (TP + FN) \quad (9)$$

$Precision$ represents the percentage of positive samples that are correctly identified, and the formula for $Precision$ is as follows:

$$Precision = TP / (TP + FP) \quad (10)$$

$F1$ is the weighted average of $Precision$ and $Recall$, which can be interpreted as the overall performance considering the effects of completeness and noise, referring to Formula (11):

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (11)$$

KGCN4CEPR evaluation: We evaluate KGCN4CEPR in a countryside ecological patterns recommendation scenario. AUC and $F1$ are used to predict the accuracy of the recommendation possibility in the test set. In Top-K recommendation, $Recall$ is chosen to evaluate select K patterns with the highest predicted recommendation probability for each country in the test set.

4.2.2. Experiment Setup and Evaluation Criterion

We compare KGCN4CEPR with the following baselines, in order to verify the effectiveness and feasibility of KGCN4CEPR in the countryside ecological patterns recommendation scenario.

SVD [13] (Singular Value Decomposition) is a traditional collaborative filtering recommendation method based on latent factor models.

MKR [25] is a multi-task knowledge graph recommendation method, which can train recommendation tasks and knowledge graph embedding tasks in an alternating manner.

RippleNet [9] is a hybrid knowledge graph recommendation method that propagates entities' preferences, such as the degree to which one entity of a countryside is interested in another entity of a countryside ecological pattern.

KGCN4CEPR is our method, which is a countryside ecological patterns recommendation method based on KGCN [19]. Specifically, we establish a countryside knowledge graph to make up for the sparsity of countryside ecological patterns data. We design a convolutional network for mining the geographical adjacency of ecological patterns among adjacent countryside.

We compare KGCN4CEPR with the above baselines and take the Top-3 recommendation result of “Hancunhe Countryside” as an example. The real Top-3 patterns as labels suitable for this countryside are *industrial development pattern*, followed by *social comprehensive governance pattern* and *leisure tourism pattern*.

The first recommended result of SVD is *leisure tourism pattern*, which does not belong to any of the Top-3 patterns. This is because SVD only uses the decomposition of the interaction matrix instead of using the countryside knowledge graph to explore semantic relation between entities.

As for MKR and RippleNet based on the countryside knowledge graph, their Top-1 recommendation results are *industrial development pattern*, but they show significant differences in their Top-3 recommendations. The Top-3 recommendation results of both MKR and RippleNet are *industrial development pattern*, *suburban intensive pattern* and *environmental improvement pattern*. This is because they take into account the similarities between “Hancunhe Countryside” and other countryside by using the countryside knowledge graph. However, the geographical location of “Hancunhe Countryside” in the two above results is ignored, which leads to a wrong pattern being recommended, *suburban intensive pattern*.

As for our method, KGCN4CEPR, its Top-3 recommendation results are *industrial development pattern*, *leisure tourism pattern* and *social comprehensive governance pattern*. Although there is still a certain gap with the real pattern, the KGCN4CEPR result is the closest to the real label. This is because KGCN4CEPR not only exploits the countryside knowledge graph, but also designs a convolutional network for sufficiently mining the geographical similarity of ecological patterns. Furthermore, KGCN4CEPR explores the geographical adjacency of “Hancunhe Countryside” instead of indefinitely expanding the spatial scale. In fact, there is no *suburban intensive pattern* in Beijing and its surrounding countryside.

The results of the experiment are presented in Table 3. We can see that our method obtains the best performance in all metrics, *AUC*, *F1*, *ACC* and *Recall*. For instance, at *AUC*, our method is improved by ~3% compared with MKR; the *F1* is ~8% higher than SVD, reaching 63.79%; the *ACC* is ~3% higher than SVD and reaches 60.94%; while the *Recall* is ~3% higher than RippleNet and reaches 56.77%. The above experimental results demonstrate the feasibility and effectiveness of our method.

Table 3. The results of the experiment in recommendation (bold font is our method).

	<i>AUC</i>	<i>F1</i>	<i>Recall</i>	<i>ACC</i>
KGCN4CEPR	0.6237	0.6379	0.5677	0.6094
RippleNet	0.5933	0.6027	0.5307	0.5806
MKR	0.5843	0.5797	0.5012	0.5758
SVD	0.5015	0.5510	0.3274	0.4737

5. Discussion on Selecting Parameters

5.1. Discussion on Selecting Parameters

In this section, according to discussion of the experimental results, we optimize the method by selecting parameters and further analyze the impact of parameters: neighbor aggregation size s , iteration number H and embedding dimensions d referring to the classical experimental approach [25].

- **Effect of neighbor aggregation size, s :** To analyze the effect of neighbor aggregation size on recommendation, we conducted several experiments with different neighbor aggregation size. The results were significant, therefore, the s for the experiment was set to 2, which is based on the AUC values of the different experimental results, as shown in Table 4.
- **Effect of iteration number, H :** As demonstrated in Table 5, the AUC first increased and then decreased with an increase in iteration number. When the iteration number is excessively small, training effect is not optimal. When the number of iteration times is 3 or 4, the AUC shows a significant decrease, because it brings much noise. Therefore, the best iteration number for the countryside ecological patterns recommendation scenario is set to 2.
- **Effect of embedding dimensions, d :** In the experiment, we observed the effect of embedding dimensions on the utilization of the countryside knowledge graph. d is set to 8 because the AUC is maximum when the number of embedding dimensions is 8, as shown in Table 6.

Table 4. AUC results of KCN4CERP with different neighbor aggregation size (bold font is our method).

s	2	4	8	16
AUC	0.5738	0.5981	0.6237	0.6094

Table 5. AUC results of KCN4CERP with different iteration numbers (bold font is our method).

H	1	2	3	4
AUC	0.5847	0.6237	0.5783	0.5526

Table 6. AUC results of KCN4CERP with different embedding dimensions (bold font is our method).

d	4	8	16	32
AUC	0.6082	0.6237	0.5918	0.5738

From the above discussion, in order to make full use of geographical adjacency, we set iteration number H to 2, dimension of embedding d to 8, neighbor aggregation time k to 2 and neighbor aggregation size s to 8, which are very important for recommendation performance.

It can be seen from the above discussion that the KGCN4CEPR can improve the recommendation AUC of countryside ecological patterns, and the established data can be applied to solve the problems of countryside recommendation. Specifically, when “Hancunhe Countryside” is input as a countryside entity, our method KGCN4CEPR can obtain the best matching recommendation output *industrial development pattern* (see Section 3.2.2 for detail). Moreover, parameters are also important when solving geographical recommendation problems. The construction of method and the selection of parameters could solve the countryside patterns recommendation problems well.

5.2. Discussion on KCN4CERP’s Application in Reality

In this section, we discuss on how our method can be applied in reality. Please see the following for the detail.

Take “Hancunhe Countryside” as an example to illustrate the recommendation of KGCN4CEPR. In general, the knowledge graph and pattern–countryside interaction matrix of a specified countryside are enacted as input data of the KGCN4CEPR method, and correspondingly obtain a suitable ecological pattern as the recommendation output. Specifically, the geographical related entities and properties of “Hancunhe Countryside” are firstly obtained by using the countryside knowledge graph, such as “Fangshan District”, “Road length”, etc., which makes up for the sparsity of data through the rich semantic information. Then, the geographical similarity of “Hancunhe Countryside” is mined through the

convolutional network, as shown in Section 3.2. Furthermore, each predicted value of ten patterns for “Hancunhe Countryside” is calculated, and the minimum value of the difference between the predicted value and the true value is eventually selected. Finally, we obtain the most suitable pattern, *industrial development pattern*, for “Hancunhe Countryside” with the minimum value selected. Similarly, the other countryside except “Hancunhe Countryside” also can be recommended. The recommendation results can help local people reasonably carry out sustainable development.

6. Conclusions

In this study, we focused on countryside ecological patterns recommendation and made two main optimizations: (1) we establish a domain knowledge graph by mining geographical features, which fully reveals the rich semantic correlation between countryside ecological patterns so that it can effectively cope with the data-sparsity problem. (2) We design a convolutional network, which sufficiently explores the adjacency of neighborhood countryside in geographical space and captures the neighborhood information of countryside ecological patterns, successfully addressing the problem of “cold start”. In terms of evaluation metrics *AUC*, *F1*, *ACC* and *Recall*, the experimental results demonstrate that the above two optimizations are effective, and the proposed method is significantly superior to existing methods.

Compared with the recommendation systems of other application scenarios, such as movie and music, countryside ecological patterns recommendation has certain specificity and challenges. There is still some room for improvement on recommendation effect. Therefore, our future work plans to further solve the problem of countryside data sparsity, such as data supplementation via web crawlers and data augmentation by using an adversarial generation network. We will also expand our method into many more applications related to the geography field on a larger spatial scale, such as districts, cities and even provinces.

Author Contributions: Conceptualization, Zhiqiang Zou and Shu Wang; methodology, Zhiqiang Zou and Xuhui Zeng; software, Zhiqiang Zou, Xuhui Zeng and Mengfei Xu; validation, Zhiqiang Zou, Xuhui Zeng, Shu Wang and Mengfei Xu; formal analysis, Zhiqiang Zou and Xuhui Zeng; investigation, Zhiqiang Zou, Xuhui Zeng and Mengfei Xu; resources, Yunqiang Zhu; data curation, Xuhui Zeng; writing—original draft preparation, Zhiqiang Zou, Xuhui Zeng and Yunqiang Zhu; writing—review and editing, Zhiqiang Zou, Shu Wang and Yunqiang Zhu; visualization, Xuhui Zeng and Mengfei Xu; supervision, Shu Wang; project administration, Zhiqiang Zou and Shu Wang; funding acquisition, Zhiqiang Zou and Shu Wang. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Strategic Priority Research Program of the Chinese Academy of Science (grant number XDA23100100), the Chinese Scholarship Council (grant number 202008320044), and the National Natural Science Foundation of China (grant number 42050101).

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: [<https://data.cnki.net/yearbook/Single/N2021110004>, accessed on 15 March 2021].

Acknowledgments: The authors would like to thank the Institute of Geographical Sciences and Natural Resources Research and Nanjing University of Posts and Telecommunications. Furthermore, we thank the editors and reviewers for their valuable comments.

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

References

1. Adomavicius, G.; Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 734–749. [[CrossRef](#)]
2. Chicaiza, J.; Valdiviezo-Diaz, P. A comprehensive survey of knowledge graph-based recommender systems: Technologies, development, and contributions. *Information* **2021**, *12*, 232. [[CrossRef](#)]
3. Wang, H.; Zhang, F.; Xie, X.; Guo, M. DKN: Deep knowledge-aware network for news recommendation. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; pp. 1835–1844.

4. Koren, Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, NV, USA, 24–27 August 2008; pp. 426–434.
5. Yang, D.; Wang, Z.; Jiang, J.; Xiao, Y. Knowledge embedding towards the recommendation with sparse user-item interactions. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Vancouver, BC, Canada, 27–30 August 2019; pp. 325–332.
6. Gu, Y.; Wu, Y.; Liu, J.; Xu, M.; Zuo, T. Ecological civilization and government administrative system reform in China. *Resour. Conserv. Recycl.* **2020**, *155*, 104654. [[CrossRef](#)]
7. Jiang, B.; Bai, Y.; Wong, C.P.; Xu, X.; Alatalo, J.M. China's ecological civilization program—Implementing ecological redline policy. *Land Use Policy* **2019**, *81*, 111–114. [[CrossRef](#)]
8. Du, W.; Yan, H.; Feng, Z.; Yang, Y.; Liu, F. The supply-consumption relationship of ecological resources under ecological civilization construction in China. *Resour. Conserv. Recycl.* **2021**, *172*, 105679. [[CrossRef](#)]
9. Wang, H.; Zhang, F.; Wang, J.; Zhao, M.; Li, W.; Xie, X.; Guo, M. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 417–426.
10. Guo, Q.; Zhuang, F.; Qin, C.; Zhu, H.; Xie, X.; Xiong, H.; He, Q. A survey on knowledge graph-based recommender systems. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 3549–3568. [[CrossRef](#)]
11. Balabanović, M.; Shoham, Y. Fab: Content-based, collaborative recommendation. *Commun. ACM* **1997**, *40*, 66–72. [[CrossRef](#)]
12. Pazzani, M.J.; Billsus, D. Content-based recommendation systems. In *The Adaptive Web*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 325–341.
13. De Campos, L.M.; Fernández-Luna, J.M.; Huete, J.F. Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks. *Int. J. Approx. Reason.* **2010**, *51*, 785–799. [[CrossRef](#)]
14. Konstas, I.; Stathopoulos, V.; Jose, J.M. On social networks and collaborative recommendation. In Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Boston MA USA, 19–23 July 2009; pp. 195–202.
15. Zhou, K.; Zhao, W.X.; Bian, S.; Zhou, Y.; Wen, J.R.; Yu, J. Improving conversational recommender systems via knowledge graph based semantic fusion. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, CA, USA, 6–10 July 2020; pp. 1006–1014.
16. Ji, S.; Pan, S.; Cambria, E.; Martinen, P.; Philip, S.Y. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 494–514. [[CrossRef](#)] [[PubMed](#)]
17. Xu, M.; Wang, S.; Song, C.; Zhu, A.; Zhu, Y.; Zou, Z. The Recommendation of the Rural Ecological Civilization Pattern Based on Geographic Data Argumentation. *Appl. Sci.* **2022**, *12*, 8024. [[CrossRef](#)]
18. Wang, H.; Zhang, F.; Zhao, M.; Li, W.; Xie, X.; Guo, M. Multi-task feature learning for knowledge graph enhanced recommendation. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 2000–2010.
19. Wang, H.; Zhao, M.; Xie, X.; Li, W.; Guo, M. Knowledge graph convolutional networks for recommender systems. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 3307–3313.
20. Tobler, W.R. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **1970**, *46* (Suppl. S1), 234–240. [[CrossRef](#)]
21. Zhu, A.X.; Lu, G.; Liu, J.; Qin, C.Z.; Zhou, C. Spatial prediction based on Third Law of Geography. *Ann. GIS* **2018**, *24*, 225–240. [[CrossRef](#)]
22. Yang, Y.J.; Xu, B.; Hu, J.W.; Tong, M.H.; Zhang, P.; Zheng, L. Accurate and efficient method for constructing domain knowledge graph. *Ruan Jian Xue Bao/J. Softw.* **2018**, *29*, 2931–2947.
23. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [[CrossRef](#)]
24. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
25. Chen, X.; Liu, Y.; Li, F.; Li, X.; Jia, X. Remote sensing image recommendation based on spatial-temporal embedding topic model. *Comput. Geosci.* **2021**, *157*, 104935. [[CrossRef](#)]