

Hybrid Recommendation Research Based on Ontology Semantics and Demographic Information

Jianhong Jiang^{1,a}, Qianqian Li^{1,b}

¹*School of Business, Guilin University of Electronic Technology, Guilin 541004, China*
a. *jhome@guet.edu.cn*, b. *monicalee0224@foxmail.com*

Keywords: ontology, semantic relevance, hybrid recommendation, demographic information, neural network.

Abstract: To reduce the influence of data sparseness and cold start problem on recommendation, an improved hybrid recommendation algorithm based on ontology semantics and demographic information is introduced. The ontology structure is constructed by the hybrid recommendation model based on the item description. The item similarity results is strengthened by analyzing the semantic relevance of the ontology. Then combining with the recommendation method based on demographic information, and the two recommended algorithms are trained by BP neural network to achieve the optional weight for mixed recommendation. Finally, the Guilin tourism destination domain ontology is established and the web crawler is used to climb the data of user platform to train and test the recommendation algorithm. The mean absolute error (MAE) and the prediction coverage (COV) value are used as evaluation index to compared with other recommendation algorithms. The results illustrate that after adding demographic information to the recommendation based on the ontology similarity recommendation algorithm, the recommendation accuracy is improved and recommendation effect is more stable.

1. Introduction

With the advancement of the mobile Internet, massive and even Galaxy-level data are constantly emerging, and it is difficult for people to obtain valuable information accurately and efficiently. In order to help users to effectively filter information, search engines and indexes do not solve the problem of information overload when users' demands are not clear or cannot be clearly expressed, so the recommendation system has become a research hotspot for scholars at home and abroad. The mainstream recommendation systems can be divided into three categories: content-based recommendation system, collaborative filtering recommendation system and hybrid recommendation system. The content-based recommendation system recommends items having similarity characteristics on the content to the user by analyzing the characteristics of the item that the user likes. However,

limited by the feature extraction method, the content-based recommendation system has poor applicability to multimedia data, and has the disadvantages of cold start, poor algorithm scalability and poor recommendation result diversity. The collaborative filtering recommendation system recommends users' preferences or items with high similarity by calculating the similarity between users or items. In addition to the disadvantages of cold start and poor scalability, it also requires large amount of user's historical data, which exists data sparse problem easily. Therefore, for improving the recommendation accuracy, some scholars have proposed a hybrid recommendation algorithm, which combines various methods by weighting and other methods to achieve common recommendation. Zhang J used item domain features to construct user preferences and combined collaborative filtering for personalized recommendation. Liu proposed a model that combines the interest community with the trust neighbors. Compared with the traditional recommendation algorithm, the model also improves the diversity and accuracy of the recommendation. Literature eleven combines collaborative filtering based on user interest characteristics with content-based recommendation methods, and content-based recommendation methods will carry auxiliary recommend when facing data sparseness problem. However, the above algorithms do not consider the influence of context semantics on the recommendation effect. The integration of semantic technology into the recommendation system has become a research hotspot of scholars at home and abroad in recent years. In the literature, Ai Danxiang modeled the context semantic information association and automatically generated inference rules to achieve catering service recommendation. Literature combines semantic association with situational awareness to improve the phenomenon that recommendation results are too specialized. Based on this, this paper proposes a hybrid recommendation algorithm used ontology semantics to calculate the similarity of the item and combined demographic information.

2. Recommendation Algorithm Based On Ontology Semantics

At the ECAI 2006 Recommender Systems Workshop in Trento, Italy, Dr. Louis A pointed out that Since the traditional recommendation algorithm has some advantages, such as not considering the composition of context semantics on the application scenario, lead to these algorithms have serious shortcomings in real-time, robustness and recommendation quality. Therefore, the concept of semantic recommendation algorithm is proposed and the core of the idea is to incorporate semantic knowledge into the recommendation process to make up the shortcomings of traditional algorithms. Subsequently, many domestic and foreign scholars began to integrate semantic technology with the recommendation system. This part will introduce ontology-based semantic reasoning technology to improve the recommendation effect on some data sparse users by strengthening the node similarity results.

3. Defining Ontology

Since the ontology definition is proposed, it is gradually applied to other specific fields. By describing the concept and relationship of field, the ontology be regarded as a combination of tree structure and semantic information, and has a good logical hierarchy. Perez et al. define the primitives of ontology modeling as classes, relations, functions, axioms, and examples. In this paper, the structure of the ontology is defined as a five-dimensional tuple $O(\text{Ontology}) = \langle C, I, P, A, R \rangle$:

(1) C represents concept set. The concept in C represents an item of the same kind in the real world. It is an abstract template, not a single individual, but a collection of all the items that satisfy the concept description. When you define a concept, you need to define its name and give a brief description.

(2) I represent instance set. An instance is a specific object created according to the corresponding concept. The instance is unique and independent of each other.

(3) P represents relationship set. Relationship reflect the interrelatedness of different concepts, divided into two parts: the definition domain and the value domain, which limit and constrain the use of the relationship. When the value range is a specific numerical value, the relationship degenerates into an attribute, so the attribute can be considered as a relationship that takes a value. There are various relationships, but research is usually more concerned general ones, such as describing the relationship of concept subordination.

(4) A represents attribute set. There are two parts, one part is the attribute of the concept, and the other part is the attribute of the relationship. The attributes of a concept reflect the characteristics of a concept. Different concepts are distinguished by attributes. Usually a concept has multiple attributes. The attribute of a relationship is a further constraint on the relationship domain and the value domain and the nature of the attribute affects the value range of the relationship.

(5) R represents inference rule set. Inference rules are rules and constraints defined on the attributes and concepts. For example, RIC is the mapping relationship between instance set I and concept set C when $I_i \in I, C_i \in C$ is defined.

Association refers to the relationship edge between two ontology instances that are linked to each other by a relationship attribute. Malak Al-Hassan proposed that association has three characteristics: (1) autocorrelation, and (2) reversibility. For example, if instance I_x is associated with I_y through the relationship attribute op , which is $I_x \xrightarrow{op} I_y$, there is an association between I_y and I_x ,

$I_y \xrightarrow{op^{-1}} I_x$, and (3)transitive, if $I_x \xrightarrow{op_1} I_y$ and $I_y \xrightarrow{op_2} I_z$, then $I_x \xrightarrow{op_3} I_z$. The associated instances form

an associated network through direct or indirect connections. The associated network structure of the instance I_x is defined as a four-dimensional tuple AN (associate network)= $\langle I_1, I_2, CL, OP \rangle$, where

$I_1, I_2 \subset I$, CL is the degree of association, indicating the distance from I_{xy}^k to the root node,

satisfying $k \in [1, N], i \in [1, N_{op_x}^k], j \in [1, N_{ins,i}^k]$, the associated edge attribute set

$OP = \{op_i^k | k \in [1, N], i \in [1, N_{op_x}^k]\}$, N represents the maximum hierarchical distance of the associated

network, $N_{op_x}^k$ indicating the value of the associated edge attribute of the Kth layer.

The following figures (Fig. 1, Fig. 2) are the associated network of the instance I_x and the instance

I_y . In Figure 1, the value of N is 4, I_{xy}^k represents the jth instance in the kth layer under the connection

of the relationship attribute $op_{x_i}^k$, for example, $I_{x_{12}}^3$ represents the second instance of the

$op_{x_1}^3$ connection in the layer 3 of the I_x association network. Figure 2 is the same.

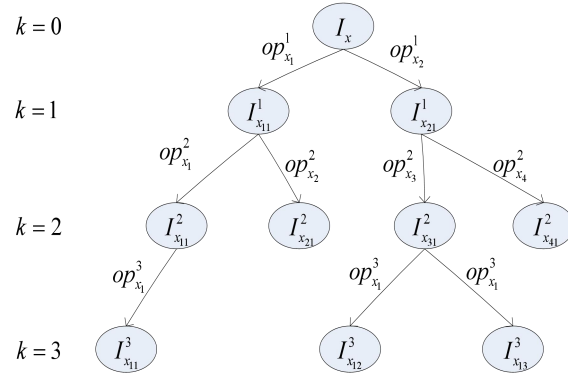


Figure 1: The association network of example I_x .

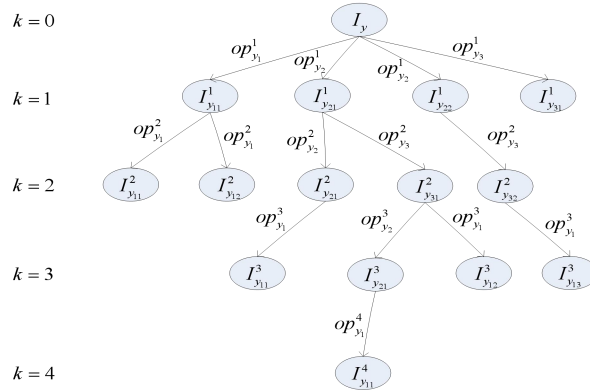


Figure 2: The association network of example I_y .

4. Building the Common Associated Pair Set

In the associated network of two instances (I_x, I_y), the instance satisfying the following conditions are added to the common associated pair set CAPS (I_x, I_y): (1) the two instances of an associated pair are from two associated networks; (2) the two elements in each pair are from the same level of association, which has the same k values; (3) the two elements in each pair are associated to each other by the same associated attributes and direct predecessor nodes, and then enter their respective associated networks; (4) the direct predecessor nodes of the two elements in each pair are also a pair in the set of shared association pairs. The m th associated pair of the set CAPS can be represented as $\text{Pair}_m = (I_{x_{w_i}}^k, I_{y_{w_j}}^k)$, and the joint connection attribute connecting element and the predecessor nodes is $op_w^k = op_{x_w}^k = op_{y_w}^k$.

For example, in the above Fig. 1 and Fig. 2, if $op_{x_1}^1 = op_{y_1}^1$ was given, it can be concluded that $(I_{x_{11}}^1, I_{y_{11}}^1)$ is an element of the set CAPS (I_x, I_y), which satisfies $I_{x_{11}}^1$ and $I_{y_{11}}^1$ exist respectively the associated network of I_x and I_y , and the association attribute and the association degree hierarchy are the same, and the predecessor node is the root node which exists in the collection.

5. Calculating the Weight of the Ontology Association Edge

In the hierarchical structure of the ontology tree, there are various relationships among the instances, and different relationships reflect the degree of association of different intensities. In the original

computing to ontology concept similarity, the weights of the associated edges are not weighted, and the default weight is 1, which means that the degrees of association of concepts connected by one associated edge are the same. Combined with the previous concept similarity research, this is obviously does not match the actual results. The weight of the associated edge shows the association strength between the two concepts that it connects. There are many factors that can affect the association strength, including the depth of the concept, the breadth of the concept's associated network, and the associated density of the associated edges. Therefore, different associated edges should have different weights.

The associated edge weight W_m indicates the association strength of the joint associated pair $(I_{x_{wi}}^k, I_{y_{wj}}^k)$ in the descendant instances of the target instance I_x and I_y ontology tree, and also indicates the influence degree of the descendant instance relevance degree on the similarity of the target instances I_x and I_y . The specific formula is as follows:

$$W(I_{x_{wi}}^k, I_{y_{wj}}^k) = \begin{cases} \frac{1}{2^k} & \text{Root node} \\ \frac{1}{3} \left(\frac{1}{2^k} \times \frac{1}{l} \times \frac{1}{C} \right) & \text{Not root node} \end{cases} \quad (1)$$

Where l is the breadth of the instance's associated network and C is the associated density of the associated edge, the l and C are computed as follows:

$$l(I_{x_{wi}}^k, I_{y_{wj}}^k) = \prod_{q=0}^{q=k-1} \left(\left| R \left(\sup Ass_{I_{x_{wi}}^k}^q, op_w^{q+1} \right) \right| \times \left| R \left(\sup Ass_{I_{y_{wj}}^k}^q, op_w^{q+1} \right) \right| \right) \quad (2)$$

Equation (2) shows the relationship between l and the predecessor node. $(I_{x_{wi}}^k, I_{y_{wj}}^k)$ is a pair of common associated pairs of the k layer in the CAPS (I_x, I_y) . $\left| R \left(\sup Ass_{I_{x_{wi}}^k}^q, op_w^{q+1} \right) \right|$ represents the breadth of association network of the same attribute of each layer of the direct predecessor node of $I_{x_{wi}}^k$, that is, the number of the same attribute associated edge of each layer of the immediate predecessor node of $I_{x_{wi}}^k$, which is $\left(\sup Ass_{I_{x_{wi}}^k}^q \right) \rightarrow op_w^{q+1} (I_{x_{wi}}^k)$, and then calculating the number of object attributes op_w^{q+1} is n by counting $\sup Ass_{I_{x_{wi}}^k}^q \cdot \left| R \left(\sup Ass_{I_{y_{wj}}^k}^q, op_w^{q+1} \right) \right|$ is the same.

Calculating $l(I_{x_{31}}^2, I_{y_{31}}^2)$ with Fig.1 and Fig.2 as examples, if instance $I_{x_{31}}^2$ has direct predecessor node set $\{I_x, I_{x_{21}}^1\}$, instance $I_{y_{31}}^2$ has direct predecessor node set $\{I_y, I_{y_{21}}^1\}$, then

$$\begin{aligned} l(I_{x_{31}}^2, I_{y_{31}}^2) &= \prod_{q=0}^{q=1} \left(\left| R \left(\sup Ass_{I_{x_{31}}^k}^q, op_w^{q+1} \right) \right| \times \left| R \left(\sup Ass_{I_{y_{31}}^k}^q, op_w^{q+1} \right) \right| \right) \\ &= \left(\left| R \left(I_{x_{21}}^1, op_3^2 \right) \right| \times \left| R \left(I_{y_{21}}^1, op_3^2 \right) \right| \right) \times \left(\left| R \left(I_x, op_2^1 \right) \right| \times \left| R \left(I_y, op_2^1 \right) \right| \right) \\ &= (1 \times 1) \times (1 \times 2) \\ &= 2 \end{aligned} \quad (3)$$

$$C(I_{x_{wi}}^k, I_{y_{wj}}^k) = \prod_{q=0}^{q=k-1} N(I_{x_{wi}}^q, I_{y_{wj}}^q) \quad (4)$$

Where C is the associated density of the associated edges, and equation (4) indicates the relationship between C and the predecessor nodes and the shared relationship attributes. $N(I_{x_{wi}}^q, I_{y_{wj}}^q)$ indicates the number of common relationship attributes of the k=q layer predecessor association pair. Calculating $C(I_{x_{wi}}^k, I_{y_{wj}}^k)$ with Fig.1 and Fig.2 as examples, then the predecessor associated pair is $\{(I_x, I_y), (I_{x_{21}}^1, I_{y_{21}}^1)\}$, and then

$$\begin{aligned} C(I_{x_{31}}^2, I_{y_{31}}^2) &= \prod_{q=0}^{q=1} N(I_{x_{wi}}^q, I_{y_{wj}}^q) \\ &= N(I_{x_{21}}^1, I_{y_{21}}^1) \times N(I_x, I_y) \\ &= 1 \times 2 \\ &= 2 \end{aligned} \quad (5)$$

From the above-mentioned associated edge weight calculation formula (1),

$$W(I_{x_{31}}^2, I_{y_{31}}^2) = \frac{1}{3} \left(\frac{1}{2^k} \times \frac{1}{l} \times \frac{1}{C} \right) = \frac{1}{3} \left(\frac{1}{2^2} \times \frac{1}{2} \times \frac{1}{2} \right) = \frac{1}{48} \quad (6)$$

can be obtained.

6. Item Similarity Based on Ontology Semantics

In the process of solving similarity, there are many influencing factors. This paper mainly chooses two aspects to measure the similarity, that is, the total amount of information between nodes and the shortest distance between nodes.

In the measurement method of measuring the similarity by the amount of information of the node, it is usually defined from two aspects, the amount of shared information or the amount of different information. The more shared information, the higher the similarity between the two nodes. And the more the amount of different information, the lower the similarity between the two nodes. The most representative one calculation method is proposed by Lin D. It is considered that the similarity of the two instances can be represented by the proportion of the node information of the nearest common ancestor of the two instances in the sum of the information amounts of the two concepts. But since the amount of information in the root node is 0, which is not considered, the similarity comparison between the root nodes cannot be performed. Therefore, the constant a is added to avoid this situation and the improved formula is as follows:

$$Sim_{IC}(I_x, I_y) = \frac{2 \times IC(NCA(I_x, I_y)) + a}{IC(I_x) + IC(I_y) + a} \quad (7)$$

Where NCA (Ix, Iy) is the nearest common ancestor of instance Ix and instance Iy. IC(Ix) represents the amount of information contained in instance Ix, and currently using the method proposed by Resnik to calculate the information amounts IC contained in the proposed nodes:

$$IC(I_x) = -\log P, \quad IC(I_x) = -\log P \quad (8)$$

$$P = \frac{n(I_x)}{N} \quad (9)$$

Where P indicates the probability that instance Ix appears in the ontology, n(Ix) indicates the number of subinstances included in instance Ix, and N is the total number of instances in the ontology.

The semantic similarity is usually measured by the shortest distance between nodes, and the depth of node is used to reflect the degree of abstraction of nodes. Therefore, the degree of common abstraction is represented by the depth of the nearest ancestor node, and the shortest distance between the two nodes represents the degree of difference abstraction. The specific formula is as follows:

$$Sim_{SD}(I_x, I_y) = \frac{d(NCA(I_x, I_y))}{d(I_x) + d(I_y) - d(NCA(I_x, I_y))} \quad (10)$$

Where $d(NCA(I_x, I_y))$, $d(I_x)$, $d(I_y)$ are respectively the node depth of the nearest common ancestor, Ix, Iy. From the above equations (7) (8) (9) (10), the item similarity based on the ontology semantics can be obtained as:

$$Sim(I_x, I_y) = \sum_{m=1}^M F_m \times (Sim_{IC}(Pair_m) + Sim_{SD}(Pair_m)) \quad (11)$$

At this point, the predicted recommendation value is:

$$P_1 = \bar{r}_x + \frac{\sum (r_{u,y} - \bar{r}_y) \times Sim(I_x, I_y)}{\sum Sim(I_x, I_y)} \quad (12)$$

Where \bar{r}_x is the average score of the item x, $r_{u,y}$ is the score of the user u for the item y, and \bar{r}_y is the average score of the item y, and $Sim(I_x, I_y)$ is the similarity between the item x and the item y.

7. Similarity Algorithm Based on Demographic Information

In order to solve the cold start problem in the recommendation system, demographic information is introduced to assist to calculate the similarity between users. When lacking of the user's empirical data, the recommendation is based on the similarity between users. In the demographic information, the age type data is a numeric attribute, while the occupational, gender type data are classification attributes, so the demographic information is data of a mixed attribute. Therefore, the user's demographic information cannot be measured by a uniform similarity measure. The overlap measure proposed by Charu C. Aggarwal is applied to the similarity calculation of occupational, gender and other classification attributes. The formula is as follows:

$$Sim(u_i, v_i) = \begin{cases} 1 & u_i = v_i \\ 0 & u_i \neq v_i \end{cases} \quad (13)$$

The data similarity of the classification attribute generally only has two possible values, "identical" or "not the same", that is, the value is 0 or 1. For the numerical attribute of age class, there is a big gap

between the data values of different individuals. The usual practice is to map it to a smaller interval by functions, and the paper select the negative exponential decay function to map:

$$Sim(u_i, v_i) = \exp(-\eta dis^\alpha(u_i, v_i)) \quad (14)$$

Where η is the scale factor, $\eta \in (0, +\infty)$, and α is a positive parameter. Among them, the value of η and α is based on the experimental conclusions proposed by M.Y.H. Al-Shamri in his graduation thesis: when $\eta = 3.8, \alpha = 2$, the function maps to the range $[0, 1]$. It can be concluded from equations (13) and (14) that under the joint action of N kinds of classification attributes and numerical attributes, the user similarity calculation formula is:

$$Sim(U_x, V_y) = \sum_{i=1}^N \lambda_i \cdot sim(u_i, v_i), \text{ and satisfying } \sum_{i=1}^N \lambda_i = 1 \quad (15)$$

At this point, the predicted recommendation value is:

$$P_2 = \bar{r}_u + \frac{\sum (r_{v,x} - \bar{r}_v) \times Sim(u, v)}{\sum Sim(u, v)} \quad (16)$$

Where \bar{r}_u is the average score of the target user u and when the target user is a new user, there is no rating information, the \bar{r}_u is 0. $r_{v,x}$ is the score of user v for the item x, \bar{r}_v indicates the average score of the user v, $Sim(u, v)$ indicates the similarity of user u and user v.

8. Hybrid Recommendation Algorithm Based on Ontology Semantics and Demographic Information

BP neural network is a kind of convergence data prediction method, which can be used to determine the optimal weight of User-based and Item-based hybrid recommendation. Liu Haiou proposed to use the ant colony neural network to train the weights of the mixed recommendation, and realized the hybrid recommendation of collaborative filtering. Other scholars used BP neural network to predict the user's preference for resource categories, the user group's satisfaction for a certain product and the overall priority of the e-commerce page, which respectively achieve accurate personalized information recommendation, strengthen the recommendation of new products and recommend the required e-commerce web pages to users quickly and accurately. Therefore, in order to improve the recommendation effect, this paper introduces BP neural network to train optimal hybrid recommendation weight during the recommendation process, and realizes the combination of item-based recommendation and demographic information to complete the common recommendation.

The recommendation matrix P is composed of the above recommendation result of the item-based recommendation and demographic information:

$$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} P_1(u_1, x_1), \dots, P_1(u_n, x_m) \\ P_2(u_1, x_1), \dots, P_2(u_n, x_m) \end{bmatrix} \quad (17)$$

The more commonly used activation functions of BP neural network are sigmoid function, tanh function, relu function and so on. In this paper, we choose the monotonically increasing and continuous

sigmoid function $f(x) = (1 + e^{-x})^{-1}$, and map the real field to the [0,1] space smoothly, which satisfies the requirements of the positive probability of the article. The matlab R2014a software is used to perform neural network weight training of the acquired data on the local PC to reduce the calculation amount and calculation time. After several iterations, the best mixed recommendation is:

$$P_{con} = \left(1 + e^{-\left(W_2 \times \left(\frac{1 - e^{-2(W_1 \times P + B_1)}}{1 + e^{-2(W_1 \times P + B_1)}} \right) + B_2 \right)} \right)^{-1} \quad (18)$$

Where W1 and W2 are the weight vectors between the input layer and the hidden layer, the hidden layer and the output layer respectively, B1 and B2 are the offset vectors of the hidden layer and the output layer respectively, and P is the recommendation matrix. Calculating the Pcon and finally providing the top N best recommendation values for the target users.

9. Experiment Analysis

Taking Guilin City as an example, this paper creates an example of the tourism destination ontology by Protégé software with the seven-step method, and OWL is used as the ontology description language. Some examples are shown in Table 1. Using the java crawler to obtain the related information of Guilin on the Mafengwo, Ctrip, and Qunar website, which includes more than 20,000 pieces of score information, and users' gender, the temporary residence area, rating and so on that can partially reflect the demographic information of users.

Table 1: Partial instance of Guilin tourism destination ontology.

Natural attractions				Humanities attractions	
Geography	Biological	Water	Meteorological	Historical site	Buildings
Longji Rice Terraces,	Big Banyan	Yulong	Seven	Xingping Old	Yangshuo West Street,
Elephant Trunk Hill,	Tree,	River,	Stars with	Town,	Arcadia, Le Mandi Theme
Yangshuo ten li gallery,	Butterfly	Two	the Moon,	Princess	Park, Strange Building
Yinzi Cave, Reed Flute	Spring,	Rivers and	Moon	Jingjiang	visual Are Museum
Cave, Nine horses	Marine	Four	Viewing	Residence,	
mountain scenic spot, Sun	Ginkgo Tree,	Lakes,	Road,	Daxu Old	
Moon Pagoda, Cat	Peach Garden	Gudong	Wind-rain	Town, Totem	
mountain	of Daling	Waterfall,	Bridge	Trail, Former	
	Mountain	Longmen		Residence of	
		Waterfall		Xu Beihong,	
				Qixia Temple	

According to the uniqueness, legality, and integrity, a total of 20,000 pieces of data are obtained after data cleaning. The data capacity of each data set after cleaning is shown in Table 2. User rating data is divided into 5 levels from 1 to 5 stars. Through the collation of each data set, the number of 3, 4, and 5 stars is 80.34% of the total number of reviews, indicating that most users are Without special circumstances, it is not inclined to give a low score evaluation.

Table 2: The information of data sets.

Data sets	Amount of data before cleaning	Amount of data after cleaning
Mafengwo	11258	10000
Ctrip	5246	5000
Qunar	5309	5000
Total	21813	20000

The data storage format is as shown in Equation (17) and Table 3, which are the user-item scoring matrix and the user demographic information table. The rating information of the user U_n for the item P_m is g_{nm} , which is stored in the user-item scoring matrix and the demographic information of the user U_n is placed in the user demographic information table.

$$\begin{matrix}
 & P_1 & P_2 & \cdots & P_m \\
 U_1 & \left[\begin{matrix} g_{11} & g_{12} & \cdots & g_{1m} \\
 U_2 & g_{21} & g_{22} & \cdots & g_{2m} \\
 \vdots & \vdots & \vdots & \cdots & \vdots \\
 U_n & g_{n1} & g_{n2} & \cdots & g_{nm} \end{matrix} \right.
 \end{matrix} \quad (19)$$

Table 3: User demographic information table.

User	Sex	Local	Level
U_1	F	Guilin, Guangxi	12
...
U_n	M	Shenzhen, Guangdong	10

The data in each dataset, which is obtained by the crawler, is randomly divided according to the ratio of 8:2, 80% of them is used as the training set and 20% is used as the test set to verify the hybrid recommendation algorithm proposed in this paper. Experimental data were evaluated using Mean Absolute Error and Coverage. The specific formula is as follows:

$$MAE = \frac{1}{n} \sum_{i \in E} |p_i - r_i| \quad (20)$$

Where p_i is the predicted value and r_i is the actual value. The absolute value of the difference between the predicted value and the actual value is used to calculate the average value, which reflect the error between the predicted value and the actual value and the smaller the value, the better the performance.

$$Cov = \frac{n}{N} \quad (21)$$

The forecast coverage is mainly to reflect whether the recommendation system can predict the scoring effect of the item. Where n is the number of predicted items and N is the capacity of the test set. The larger the coverage, the better the performance.

The algorithm proposed in this paper (paper) is compared with the traditional demographic-based recommendation method (DB), the ontology similarity recommendation method (OB), and the item-based collaborative filtering recommendation method (Item-CF), and the experimental results under different data sparse conditions are shown in Tables 4, 5, and 6, the overall changes are shown in Fig. 4 and Fig. 5.

Table 4: Comparison of the first result.

K Value	MAE				Cov			
	DB	OB	Item-CF	Paper	DB	OB	Item-CF	Paper
K=10	0.8515	0.8001	0.7968	0.7950	0.5883	0.7142	0.6411	0.7172
K=20	0.8668	0.7824	0.8075	0.7738	0.6715	0.7969	0.7832	0.8203
K=30	0.8647	0.7780	0.8086	0.7667	0.7136	0.8712	0.8517	0.8990
K=40	0.8436	0.7621	0.7925	0.7519	0.7927	0.9079	0.8924	0.9324
K=50	0.8423	0.7613	0.7901	0.7523	0.8934	0.9405	0.9179	0.9602

Table 5: Comparison of the second result.

K Value	MAE				Cov			
	DB	OB	Item-CF	Paper	DB	OB	Item-CF	Paper
K=10	0.8438	0.8108	0.8015	0.7935	0.6656	0.7089	0.6695	0.7201
K=20	0.8366	0.7834	0.7911	0.7648	0.7318	0.7864	0.8034	0.8198
K=30	0.8424	0.7645	0.7864	0.7518	0.7534	0.8633	0.8658	0.8869
K=40	0.8339	0.7579	0.7861	0.7479	0.8418	0.9054	0.8972	0.9267
K=50	0.8218	0.7504	0.7809	0.7453	0.9021	0.9399	0.9214	0.9579

Table 6: Comparison of the third result.

K Value	MAE				Cov			
	DB	OB	Item-CF	Paper	DB	OB	Item-CF	Paper
K=10	0.8935	0.8101	0.7935	0.7891	0.5960	0.7032	0.6775	0.7249
K=20	0.8216	0.7815	0.7824	0.7668	0.6715	0.7756	0.8077	0.8173
K=30	0.8102	0.7659	0.7769	0.7546	0.7413	0.8613	0.8660	0.8821
K=40	0.8056	0.7584	0.7713	0.7503	0.8322	0.9025	0.8972	0.9234
K=50	0.7995	0.7536	0.7702	0.7471	0.8319	0.9357	0.9187	0.9446

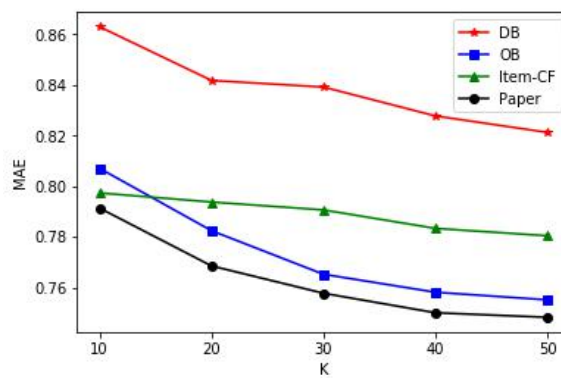


Figure 4: MAE values of the four methods.

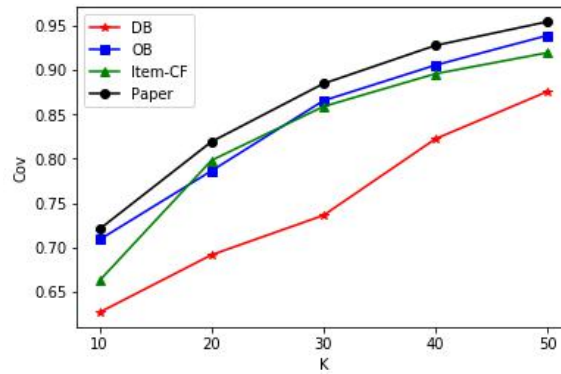


Figure 5: Cov values of the four methods.

From the vertical comparison of each method, we can see that with the increase of the neighboring number K value, the MAE value of all the algorithms almost show the decreasing trend, which is consistent with the normal phenomenon manifestation, and the difference is the change scale and scope of each algorithm. It can be seen from the experimental data that although the overall MAE value of the demographic-based recommendation method (DB) shows a downward trend, there are still several fluctuation points, which indicates that the method has poor stability and low fault tolerance. The item-based collaborative filtering recommendation method (Item-CF) and the ontology similarity recommendation method (OB) are the more commonly used recommendation methods. And in this experiment, their MAE values show an overall downward trend. The algorithm proposed in this paper (Paper) improves the calculation of similarity by the integration of ontology semantics and demographic information., its MAE value shows an overall decline and the experimental data has almost no fluctuation points. Therefore, in terms of algorithm stability, the demographic-based recommendation method (DB) performs the worst, and the ontology similarity recommendation method (OB) and the proposed algorithm perform well. From Fig. 5, the Cov values of the four algorithms all show a gradual increase with the increase of the nearest neighbor K value, and the coverage can reach a higher level when $K=50$.

Horizontally, in terms of MAE values, the four methods all show an overall downward trend, but the decline rate of the item-based collaborative filtering recommendation method (Item-CF), the ontology similarity recommendation method (OB) and the proposed algorithm (paper) tend to be flat, indicating that the degree of influence on the MAE value of the recommended method begins to decrease when the neighbor number K increases to a certain value. But the MAE value of the demographic-based recommendation method (DB) is much higher than the other three methods, which indicates that the recommended accuracy is very low when using this method alone. In terms of Cov value, the four methods all show an upward trend, and the rate tends to be flat. The recommendation method based on ontology similarity (OB) and the algorithm proposed in this paper (paper) have always performed well.

In summary, the demographics-based recommendation method alone has low accuracy and poor stability, but the proposed algorithm has good stability and excellent recommended accuracy after combining with the recommendation method based on ontology similarity that always has excellent performance.

10. Conclusions

Using the recommendation method based on ontology semantic association to strengthen item similarity is an important way to solve the problem of data sparseness. This paper combines the

demographic-based recommendation method and uses the weight training of neural network to realize the common recommendation of the two recommendation methods. The superiority of this hybrid recommendation has been verified initially and alleviate the data sparseness and cold start problems of the traditional recommendation system to some extent. The hybrid recommendation based on ontology semantic association and demographic information is only a simple attempt and verification. However, due to the gradient of BP neural network, local optimization is easy to occur during training. So how to improve BP neural network to overcome the local optimal problem and improve the recommendation accuracy, how to combine the item similarity algorithm based on ontology semantic association with the content-based collaborative filtering algorithm is the direction to be discussed later. The experimental data of this paper is limited and comes from the network, and no specific empirical activities have been carried out, so the next step is to collect more realistic data to improve the recommendation effect.

Acknowledgments

Funded projects: Special Research Program of National Natural Science Foundation of China (No. 71940008); Humanities and Social Sciences Fund Youth Project of the Ministry of Education (No.17YJCZH074); Guilin University of Electronic and Technology Graduate Education Innovation Program (No.2016YJCX57).

References

- [1] Jinyun Zhu, Jian Chen, Dezheng Wang. *A guide for big data architects*[M]. Beijing: Journal of Tsinghua University, 2016.06:5-6.
- [2] Ye Tian, Zhongming Zhu, Shudong Liu. *Review of Recommendation System Based on Linked Data* [J]. *New Technology of Library and Information Service*, 2013(10):1-7.
- [3] Lei Li, Li Zheng, Fan Yang, et al. *Modeling and broadening temporal user interest in personalized news recommendation*[J]. *Expert Systems with Applications*,2014,41(7):3168-3177.
- [4] G Adomavicius, A Tuzhilin. *Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions*[J]*IEEE Transactions on Knowledge and Data Engineering*,2005, 6 (17):734-749.
- [5] Hailing Xu, Xiao Wu, Xiaodong Li, et al. *Comparison Study of Internet Recommendation System*[J]. *Journal of Software*,2009,20(2):350-362.
- [6] Zhenhua Huang, Jiawen Zhang, Bo Zhang, et al. *Survey of Semantics-Based Recommendation Algorithms*[J]. *Acta Electronica Sinica*,2016,44(9):2262-2275.
- [7] Zheng Z, Ma H, Lyu MR, et al. *QoS-aware web service recommendation by collaborative filtering*[J]. *IEEE Transactions on Services Computing*,2011,4(2):140-152.
- [8] Junpeng Guo, Yuanyuan Li, Jing Ning, et al. *A Hybrid Recommendation Algorithm Based on Symbolic Modal-valued Data Analysis*[J]. *Systems Engineering*,2014.32(4):111-116.
- [9] Zhang J, Peng Q, Sun S, et al. *Collaborative filtering recommendation algorithm based on user preference derived from item domain features*[J]. *Physical A Statistical Mechanics and Its Applications*, 2014.396(2):66-76.
- [10] Qihua Liu. *Hybrid Recommended Model Based on Communities of Interests and Trust Neighbors* [J]. *Information Science*, 2016,34(2):65-69.
- [11] Liang T, Liang Y, Fan J, et al. *A hybrid recommendation model based on estimation of distribution algorithms*[J]. *Journal of Computational Information Systems*,2014.10(2):781-788.
- [12] Danxiang Ai, Yufeng Zhang, Gaoyong Liu, et al. *Contextual Semantics Modeling and Rule-based Reasoning for Mobile Business Catering Recommendation*[J]. *Information Studies: Theory & Application*,2016,39(2):82-88.
- [13] Fenglin Li, Dexin Chen, Shaoxing Liang. *Research on Personalized Recommendation Method Based on Semantic Association and Context Awareness*[J]. *Journal*.

- [14] Loizou A, Dasmahapatra S. *Recommender systems for the semantic web*[A]. *Proc of Recommender Systems Workshop*[C], Trento: ECAI,2006.1-5.
- [15] Perez A G, Benjamins V R. *Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods*[A]. In: *Proceedings of IJCAI-99 Workshop on Ontologies and Problem Solving Methods (KRR5)*[C].Stockholm, Sweden, 1999:1-15.
- [16] Malak Al-Hassan, Haiyan Lu, Jie Lu. *A semantic enhanced hybrid recommendation approach: A case study of e-Government tourism service recommendation system*[J]. *Decision Support Systems*, 2015,(72):97-109.
- [17] Lin D. *An information-theoretic definition of similarity*[A]. In: *Proceedings of the 15th International Conference on Machine Learning*[C]. Madison, USA: Morgan Kaufmann, 1998:296–304.
- [18] Resnik P. *Using information content to evaluate semantic similarity in a taxonomy*[A]. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*[C]. Montreal, Canada: Morgan Kaufmann, 1995:448–453.
- [19] Wenqing Li, Sun Xin, Changyou Zhang, Ye FENG. *A Semantic Similarity Measure between Ontological Concepts*[J]. *Acta Automatica Sinica*,2012,32(2):229-235.
- [20] Charu C. Aggarwal. *Data Mining: The Textbook*[M]. New York:Springer,2015:74-75.
- [21] Mohammad Yahya H, Al-Shamri. *User profiling approaches for demographic recommender systems* [J]. *Knowledge-Based Systems*,2016(100):175-187.
- [22] M.Y.H.Al-Shamri, *Fuzzy and genetic algorithm approaches to recommender systems*[D], India, New Delhi: Jawaharlal Nehru University, 2008.
- [23] Haiou Liu. *Context-aware Hybrid Recommendation for QoS Preference Prediction of Mobile Service in Cloud Environment*[J]. *Journal of Intelligence*,2016,35(4):183-189+194.
- [24] Puxiong Zhou, Bingrong Zhang, Longwen Zhao. *Research on Contextual Information Recommendation Service Based on BP Neural Network*[J]. *Information Science*,2016,34(03):71-75.
- [25] Jie Wu, Feng Feng. *A Collaborative filtering algorithm incorporating users preference and recommending new products first*[J]. *Computer Applications and Software*,2014,31(10):285-287+328.
- [26] Feiyu Wang, Hua Wen, Hongmei Wang. *E-commerce webpage recommendation scheme base on semantic mining and neural networks*[J]. *Control Engineering of China*,2018,25(01):149-153.