

Received January 29, 2020, accepted February 19, 2020, date of publication March 13, 2020, date of current version March 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2980624

# Improvement of Collaborative Filtering Recommendation Algorithm Based on Intuitionistic Fuzzy Reasoning Under Missing Data

YANJU ZHANG<sup>1,2</sup>, YUE WANG<sup>1</sup>, AND SHIQIN WANG<sup>1</sup>

<sup>1</sup>School of Business Administration, Liaoning Technical University, Huludao 125105, China

<sup>2</sup>Academy of Management Science and Engineering, Liaoning Technical University, Huludao 125105, China

Corresponding author: Yanju Zhang (zhangyanju@lntu.edu.cn)

This work was supported in part by the Provincial Department of Education Project of Liaoning (Organization Optimization of Urban Space System Based on Check-in Data of Residents Mobile Network), China, under Grant LJ2017QW010, and in part by the Provincial Social Science Planning Fund of Liaoning (Research on Urban Space-time Elasticity Based on Data Signed by Resident Mobile Network), China, under Grant L18CJY003.

**ABSTRACT** Commodity recommendation plays an essential role in the marketing field in the Internet era, and collaborative filtering, as a powerful technique of commodity recommendation, has been widely concerned in both academic studies and practical applications. Existing research on collaborative filtering often uses methods such as genetic algorithm and neural network to solve the sparsity and cold start problems while ignoring the fuzziness of users' ratings on goods or services. To solve the problems, we propose a recommendation algorithm (IFR-CF) based on intuitionistic fuzzy reasoning and collaborative filtering. In this algorithm, the characteristic coefficient in intuitionistic fuzzy reasoning is used to replace the traditional similarity coefficient to determine neighbor set, and the finite prior ordering method is used to replace traditional algorithm to recommend commodity. Two groups of data are extracted from Movielens and Jester datasets for experiments, and the MAE value generated by the recommended items is taken as the metric to verify the algorithm performance. Experimental results show that compared with the traditional algorithms, our algorithm achieves lower MAE value and higher recommendation accuracy. Meanwhile, the intuitionistic index of fuzzy set is taken into account in the calculation of the hesitation coefficient, which provides a novel solution to the problem of missing scoring data of users.

**INDEX TERMS** Collaborative filtering, intuitionistic fuzzy reasoning, intuitive index, similarity.

## I. INTRODUCTION

WITH the development and popularization of e-commerce, online shopping has become a typical behavior of the public. When users surf the Internet, they often receive recommendations for products. The recommendation of these products will undoubtedly provide a powerful boost for merchants to attract customers. Internet service platforms have been working hard to offer product recommendations to users more accurately so that products recommended to users are closer to their interests.

The associate editor coordinating the review of this manuscript and approving it for publication was Xin Zhang<sup>1</sup>.

In the product recommendation research, the nearest neighbor collaborative filtering recommendation algorithm is widely used. The main idea of the algorithm is to discover users' preferences by mining their historical behavior records, group users according to different preferences and recommend similar products [1]. So far, collaborative filtering recommendation algorithms are continually being improved. To alleviate the impact of data sparseness and cold start on the recommendation, Wu *et al.* [2] propose to combine the limited Boltzmann machine model and trust information to improve the performance of recommendation, where the trust information is the degree of trust between the target user and other users. In the recommendation process, the accuracy of the recommendation is improved

by considering the trust degree of the target user to the recommendation opinion. Mohammadpour *et al.* [3] combine genetic algorithm and gravity simulation local search algorithm to enhance clustering effects in collaborative filtering, reduce mean absolute error (MAE) and root mean square error (RMSE), thus make coverage criteria better. Laishram *et al.* [4] use an evolutionary algorithm to obtain the subsets of highly correlated terms, and the local least squares method is used to analyze the user-subgroup of highly relevant items to calculate the missing scores. Genetic algorithm, as an essential technology in modern intelligent computing, is also introduced into the collaborative filtering recommendation algorithm to improve the quality of clustering and the accuracy of the recommendation [5]–[7]. Zhang *et al.* [8] propose a new collaborative filtering recommendation algorithm that combines the technique of time window and rating prediction to estimate the preferences of the users without any rating items. Besides, the singular value and the trust factor are also considered in the calculation of the collaborative filtering algorithm [9], [10]. To meet the individual needs of users, the personalized recommendation algorithm combining probabilistic semantic clustering analysis and collaborative filtering is used to recommend the most relevant items to users [11]. From the perspective of user reviews, some scholars generate user representations of the overall sentiment about item characteristics from their comments, analyze the user sentiment to explore their interests, and complete the recommendations [12]. Although the above methods improve the accuracy of recommendation, there still exist uncertainties in practical problems, including the inability to accurately determine user interests and the inability to accurately describe in the recommendation, which needs further research to solve them.

In this paper, intuitionistic fuzzy reasoning is introduced into the collaborative filtering recommendation algorithm, and the commodity recommendation problem based on similar users is studied from the perspective of intuitionistic fuzzy reasoning. In our solution, user interests are used as fuzzy sets, and rating values on commodities are converted into the membership degree of interest fuzzy sets. The membership degree, non-membership degree and intuition index in intuitionistic fuzzy reasoning are used to determine the neighbor users and then recommend the products for them.

The contributions and advantages of our work are as follows:

- Intuitionistic fuzzy reasoning, as an extension of fuzzy reasoning, is introduced to solve the problem of commodity recommendation. It cannot only describe the fuzzy concept of “Both this and that”, but also describe the fuzzy concept of “Neither this nor that”, which makes the mathematical description more consistent with the nature of fuzzy objects in the real world.
- Feature coefficients in intuitionistic fuzzy reasoning are used to replace the conventional similarity coefficients to determine the neighbor set, and the finite relationship order method is used to replace the traditional

recommendation algorithm for product recommendation, which improves the accuracy of the recommendation results.

- During the calculation process, the intuition index of intuitionistic fuzzy set is taken into account by the hesitation coefficient, and a new method is proposed to deal with the problem of missing user scoring data from the perspective of intuitionistic fuzzy reasoning.

## II. RELATED WORK

The recommendation system is the technique to provide personalized services based on historical data of users, and has been applied to various fields to offer suitable services for different types of users. However, in the current network environment, the privacy of users is better protected, and many users are reluctant to disclose their personal information. Therefore, the recommendation system is usually faced with highly sparse data sets, which increases the difficulty of recommendation. To solve the problem of data sparseness and cold start in the recommendation system, Iwanaga *et al.* [13] use a shaped constraint optimization model to estimate the probability of item selection based on the recently visited web pages and corresponding frequencies of each user. Huang *et al.* [14] propose a new low-rank sparse cross-domain recommendation algorithm to improve the recommendation performance between related domains through knowledge transfer, and also propose a solution to the data sparsity problem to improve the recommendation quality. Batmaz and Kaleli [15] introduce a collaborative filtering system based on multiple criteria to improve the personalization of the system, and extract the nonlinear relationship between users and items by deep learning. It makes the recommendation algorithm more personalized and provides more accurate product recommendations for different users. Yu *et al.* [16] construct a Contextual-boosted Deep Neural Collaborative filtering (CDNC) model for item recommendation, which uses item introductions and user ratings to alleviate the cold start problem of item recommendation. Deng *et al.* [17] design a novel deep hybrid recommendation framework, Neural variational collaborative filtering (NVCF), which incorporates the profile information of user and item in the generation process to alleviate rating sparsity, and better latent user/item representations are obtained. Xiao and Shen [18] propose a novel deep generative model, namely Neural Variational Matrix Factorization, which also integrates profile information of users and items, and ratiocinates the complex nonlinear representations of users and items through the neural network.

In terms of data sparseness, many new theories and models have been introduced into the traditional algorithm to improve the accuracy of recommendation. Due to the user ratings for products or services are fuzzy, fuzzy mathematics is introduced into this study to deal with such problems. Through intuitionistic fuzzy reasoning, the similarity between users is better measured, and the fuzzy problem can be described quantitatively. Moreover, through the

introduction of the hesitation coefficient, the missing data in scoring matrix is filled, and the sparsity of data is reduced.

### III. TRADITIONAL COLLABORATIVE FILTERING ALGORITHM

The collaborative filtering recommendation algorithm is the earliest and well-applied recommendation algorithm, which is used primarily for preference prediction and item recommendation. By mining a specified user’s historical behavior data, the algorithm analyzes the user’s interest, finds other users with similar interest in the user set, synthesizes the evaluation of these related users on certain items, forms the system’s preference prediction for the items, and finally recommends items with similar interest for the user [19], [20].

In the collaborative filtering process, users’ preferences should be collected first, then similar users should be found, and items finally should be recommended to the target user. When searching for similar users, the Pearson correlation coefficient is generally used to calculate the similarity [21] as Equation 1.

$$p(x, y) = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{(n-1)S_x S_y}$$

$$\frac{\sum x_i y_i - n \bar{x} \bar{y}}{(n-1)S_x S_y} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad (1)$$

$x$  and  $y$  are two variables, and they are used to represent two users (i.e., two fuzzy sets) in this paper,  $n$  is the number of products being rated, and  $x_i$  and  $y_i$  are two users’ ratings for the product  $i$ .

The similarity calculation finds the neighbor users, and then the item recommendation is performed for the target user according to the preference of his neighbor users. In the recommendation process, the interest degree of the user can be calculated by Equation 2

$$s(u, i) = \sum_{v=1}^n p(u, i) \times \gamma(v, i) \quad (2)$$

$v$  is the neighbor user of user  $u$ ,  $n$  is the number of neighbor users,  $p(u, i)$  is the similarity between user  $u$  and user  $v$ , and  $\gamma(v, i)$  is the user  $v$ ’s interest in the product  $i$ .

### IV. INTUITIONISTIC FUZZY REASONING

Intuitionistic fuzzy reasoning is a thinking process and method to infer a new proposition according to known propositions and given rules. There are many inference methods, including intuitionistic fuzzy implicit reasoning, conditional reasoning, multiple reasoning, multidimensional reasoning, and multiple multidimensional reasoning.

Let  $Q$  be a given universe, then an intuitionistic fuzzy set  $A$  on  $Q$  is  $A = \{(q, \mu_A(q), \gamma_A(q)) | q \in Q\}$  where  $0 \leq \mu_A(q) \leq 1$  and  $0 \leq \gamma_A(q) \leq 1$  are the membership function  $\mu_A(q)$  and the non-membership function  $\gamma_A(q)$  of the intuitionistic fuzzy set, respectively, and all  $q \in Q$ ,  $0 \leq \mu_A(q) + \gamma_A(q) \leq 1$  is true [22].

When  $A \in IFS(Q)$  and  $A$  contains the membership function  $\mu_A(q)$ , the non-membership function  $\gamma_A(q)$  and the intuition index  $\pi_A(q)$  [23], The relationship between them satisfies the following constraints:

$$\mu_A(q) + \gamma_A(q) + \pi_A(q) = 1, \quad (3)$$

then  $\pi_A(q) = 1 - \mu_A(q) - \gamma_A(q)$  is the intuitionistic index of the intuitionistic fuzzy set  $A$ , which is used to measure the hesitation degree of  $q$  to  $A$ . From the above definition, the range of hesitation is  $0 \leq \pi_A(q) \leq 1$ .

Let  $A_{ij}, B_i \in [0, 1]$  be intuitionistic fuzzy reasoning, and  $A_{ij}$  be valued in domain  $X$ ,  $B_i$  in domain  $Y$ ,  $i, j = 1, 2, \dots, n$ , then the form of intuitionistic fuzzy multi-dimensional reasoning is recorded as

$$[(A_{11} \times A_{12} \times \dots \times A_{1n} \rightarrow B_1), (A_{21} \times A_{22} \times \dots \times A_{2n} \rightarrow B_2), \dots, (A_{n1} \times A_{n2} \times \dots \times A_{nn} \rightarrow B_n)].$$

Intuitionistic fuzzy reasoning is carried out according to specific rules, when all conditions of rule  $A_{ij}$  are satisfied,  $B_i$  can be inferred, and the final result can be obtained gradually. In this paper, we use the method of direct fuzzy reasoning to fuzzify the data and calculate the similarity between fuzzy sets. Finally, according to the preference of the neighbor users, we recommend products in that the target user is most likely interested.

### V. COLLABORATIVE FILTERING ALGORITHM BASED ON INTUITIONISTIC FUZZY REASONING

In this section, intuitionistic fuzzy reasoning is introduced into collaborative filtering recommendation algorithm, and we use this new algorithm to recommend products for users based upon their interests and preferences. The basic ideas are as follows:

- Fuzz the relevant data and calculate the corresponding membership degree and non-affiliation degree;
- Use the similarity formulation of intuitionistic fuzzy rough set to calculate the similarity between fuzzy sets, and find the neighbor users;
- According to the preference of the neighbor users, the product recommendation is executed to the target user.

#### A. DATA FUZZIFICATION

In the current network environment, after using or accepting a certain product or service, the user is often required to evaluate the product or service, and then reflect his satisfaction, interest, or other information. Due to the wide variety of products in the network and millions scale of users, a huge data network has been formed. In the real-life e-commerce platform, goods can often be divided into several categories. For example, in the movie recommendation system, movie categories can be divided into comedy, science fiction, action, etc. Without loss of generality, suppose that  $n$  types of products are divided into  $K$  categories. As shown in Table 1,  $r_{ij}$  is the rating score of user  $i$  on product  $j$ , and  $1, 2, \dots, n$  is

TABLE 1. Evaluation data based on product classification.

	$C_1$			$C_2$		$C_3$		...		$C_K$	
	1	2	...	i	...	j	...	...	...	n	
1	$r_{11}$	$r_{12}$	...	$r_{1i}$	...	$r_{1j}$	...	...	...	$r_{1n}$	
2	$r_{21}$	$r_{22}$	...	$r_{2i}$	...	$r_{2j}$	...	...	...	$r_{2n}$	
...	...	...	...	...	...	...	...	...	...	...	
m	$r_{m1}$	$r_{m2}$	...	$r_{mi}$	...	$r_{mj}$	...	...	...	$r_{mn}$	

TABLE 2. Data for intuitionistic fuzzy reasoning.

	$A_1$	$A_2$	$A_3$	...	$A_K$
1	$(z_{11}, \dots, z_{1i-1})$	$(z_{1i}, \dots, z_{1j-1})$	$(z_{1j}, \dots, z_{1t})$	...	$(z_{1s}, \dots, z_{1n})$
2	$(z_{21}, \dots, z_{2i-1})$	$(z_{2i}, \dots, z_{2j-1})$	$(z_{2j}, \dots, z_{2t})$	...	$(z_{2s}, \dots, z_{2n})$
...	...	...	...	...	...
m	$(z_{m1}, \dots, z_{mi-1})$	$(z_{mi}, \dots, z_{mj-1})$	$(z_{mj}, \dots, z_{mt})$	...	$(z_{ms}, \dots, z_{mn})$

the product sequence, 1, 2, ..., m is the user sequence. The maximum value of  $r_{ij}$  is  $r_{max}$ , and the unrated values are 0. According to the normalization operation,  $r_{ij}$  is normalized to the value among the [0, 1]. Suppose  $z_{ij}$  is the normalized value of the user's rating, which is a fuzzy number that indicates the interest degree of user  $i$  in product  $j$ . Consequently,  $C_i$  can be converted into the fuzzy set  $A_i$  as shown in Table 2. It is worth mentioning that in solving practical problems, it is necessary to divide or define corresponding levels according to the significance of rating scores.

**B. FIND NEIGHBOR USERS**

Through the similarity calculation between the target user and other users, the neighbor user is found among many users, and then the product recommendation for the target user is performed through the preference of the neighbor users. However, in the network filled with a large amount of data, there always exists missing data, which results in incomplete and inaccurate data, and further affects the similarity calculation. In order to reduce this negative impact, this paper adopts the similarity measurement method related to intuitionistic fuzzy sets for calculation.

*Definition 1:* (Distance between intuitionistic fuzzy sets) Let  $A$  and  $B$  be the intuitionistic fuzzy subsets of the given domain  $X$ ,  $d$  is a mapping and satisfies that  $d : IFS(X) \times IFS(X) \rightarrow [0, 1]$ . If  $d(A, B)$  satisfies the property (DP1) – (DP4), it is called the distance between intuitionistic fuzzy sets  $A$  and  $B$  [24].

- (DP1):  $0 \leq d(A, B) \leq 1$
- (DP2):  $d(A, B) = 0$  If and only if  $A = B$
- (DP3):  $d(A, B) = d(B, A)$
- (DP4): If  $A, B, C \in X, A \subseteq B \subseteq C$ , then  $d(A, C) \geq d(A, B), d(A, C) \geq d(B, C)$

*Definition 2:* Let  $A$  and  $B$  be the intuitionistic fuzzy sets on domain  $X = x$ , Where  $A = \{(x, \mu_{Ax}, \gamma_{Ax})\}, B = \{(x, \mu_{Bx}, \gamma_{Bx})\}$  corresponds to the interval  $[\mu_{Ax}, 1 - \mu_{Ax}]$ ,

$[\mu_{Bx}, 1 - \mu_{Bx}]$ . The distance between  $A$  and  $B$  can be obtained by using Hausdorff measure [25].

$$d_H(A, B) = \max\{|\mu_{Ax} - \mu_{Bx}|, |1 - \mu_{Ax} - (1 - \mu_{Bx})|\} = \max\{|\mu_{Ax} - \mu_{Bx}|, |\gamma_{Ax} - \gamma_{Bx}|\} \tag{4}$$

*Definition 3:* (Similarity between intuitionistic fuzzy sets) Let  $A$  and  $B$  be the intuitionistic fuzzy subsets of the given domain  $S$ ,  $d$  is a mapping and satisfies that  $S : IFS(X) \times IFS(X) \rightarrow [0, 1]$ . If  $S(A, B)$  satisfies the property (SP1) – (SP4), it is called the similarity between intuitionistic fuzzy sets  $A$  and  $B$  [26].

- (SP1):  $0 \leq S(A, B) \leq 1$
- (SP2):  $S(A, B) = 1$  If and only if  $A = B$
- (SP3):  $S(A, B) = S(B, A)$
- (SP4): If  $A, B, C \in X, A \subseteq B \subseteq C$ , then  $S(A, C) \geq S(A, B), S(A, C) \geq S(B, C)$

*Theorem 1:* If  $d(A, B)$  is the distance between intuitionistic fuzzy sets  $A$  and  $B$  satisfying definition 1, then  $S(A, B) = 1 - d(A, B)$  is the similarity between intuitionistic fuzzy sets  $A$  and  $B$ .

*Proof 1:* The definition of the distance between intuitionistic fuzzy sets and the similarity of intuitionistic fuzzy sets shows that the relationship between  $S(A, B)$  and  $d(A, B)$  changes in the opposite direction.  $S(A, B)$  increases with the decrease of  $d(A, B)$ , and  $S(A, B) \leq 1$ , so the similarity between the intuitionistic fuzzy sets  $A$  and  $B$  is  $S(A, B) = 1 - d(A, B)$ . The theorem is proven.

Under discrete data, the standardized Hamming distance is

$$l_M(A, B) = \frac{1}{2n} \times \sum_{i=1}^n (|\mu_A(x_i) - \mu_B(x_i)| + |\gamma_A(x_i) - \gamma_B(x_i)| + \rho \times |\pi_A(x_i) - \pi_B(x_i)|). \tag{5}$$

So the similarity calculation formula of intuitionistic fuzzy sets  $A$  and  $B$  is

$$S(A, B) = 1 - l_M(A, B) = 1 - \frac{1}{2n} \times \sum_{i=1}^n (|\mu_A(x_i) - \mu_B(x_i)| + |\gamma_A(x_i) - \gamma_B(x_i)| + \rho \times |\pi_A(x_i) - \pi_B(x_i)|). \tag{6}$$

In the real scenario, considering that user has no evaluation or lack of data in the evaluation of the product, it is generally considered that the evaluation may have either a high score or a low one, so it is considered reasonable to assume that the probability of high scores and low scores is half in the absence of other prior knowledge, that is,  $\rho = \frac{1}{2}$  is generally reasonable.

For the calculation of the intuitionistic index  $\pi_A(x_i)$ , the method of the intuitionistic fuzzy rough set is introduced. Since the intuitionistic fuzzy set is a special state of the intuitionistic fuzzy rough set [27], it is reasonable to use this method for the approximation calculation of intuitionistic fuzzy set with missing data (i.e., data with a score of 0).

Let  $A, B$  be two intuitionistic fuzzy rough sets on the non-empty universe  $X = \{x_1, x_2, \dots, x_n\}$ ,  $RS_A(x) = \langle \mu_{A-}(x), \mu_{A+}(x), \gamma_{A-}(x), \gamma_{A+}(x) \rangle$  be the intuitionistic fuzzy

TABLE 3. The interest similarity under different score gaps.

Score gap	Interest similarity	
	Low score	High score
0	0.5	0.5
1	0.4	0.6
2	0.3	0.7
3	0.2	0.8
4	0.1	0.9
5	0	1

rough value of  $X$  in  $A$ , and  $RS_B(x) = \langle \mu_{B^-}(x), \mu_{B^+}(x), \gamma_{B^-}(x), \gamma_{B^+}(x) \rangle$  be the intuitionistic fuzzy rough value of  $X$  in  $B$ , where  $\mu_{A^-}(x)$  is the lower approximate membership degree of  $A$  and  $\mu_{A^+}(x)$  is the upper one,  $\gamma_{A^-}(x)$  is the lower approximate non-membership degree of  $A$  and  $\gamma_{A^+}(x)$  is the upper one, and the mathematical expression of  $B$  is similar to  $A$ . The calculation method of the upper approximate membership degree  $\mu_{A^+}(x)$  and the lower approximate membership degree  $\mu_{A^-}(x)$  is as follows:

$$\mu_{A^+}(x) = \frac{\sum_{z \in A} z + a \times \frac{\bar{\alpha}}{r_{max} - r_{min}}}{n}; \tag{7}$$

$$\mu_{A^-}(x) = \frac{\sum_{z \in A} z + a \times \frac{\bar{\beta}}{r_{max} - r_{min}}}{n}, \tag{8}$$

where  $z$  is the fuzzy data in set  $A$ ,  $n$  is the number of fuzzy data in set  $A$ , and  $a$  is the number of missing data in set  $A$ ,  $\bar{\alpha}$  is the average of the amount of the scores above the median in the scoring range,  $\bar{\beta}$  is the average of the sum of the scores below the median in the scoring range.  $r_{max}$  is the maximum value in the scores, and  $r_{min}$  is the minimum value in the scores.

The calculation method of  $\pi_A(x)$  is shown in Equation 9

$$\begin{aligned} \pi_A(x) &= \mu_{A^+}(x) - \mu_{A^-}(x) = \gamma_{A^-}(x) - \gamma_{A^+}(x) \\ &= 1 - \mu_{A^+}(x) - \gamma_{A^+}(x). \end{aligned} \tag{9}$$

C. PRODUCT RECOMMENDATION FOR TARGET CUSTOMERS

There may be many neighborhood users, and each neighborhood user will score the predicted objects. We can use the priority ordering method to determine the degree of interest of the neighbor users to the predicted objects, determine the priority order of these objects, and finally make recommendations. Suppose the products to be recommended are  $c_1, c_2, \dots, c_n$ , the total is  $n$ , and their priority matrix  $C = (c_{ij})_{n \times m}$ . In matrix  $C$ , let  $c_{ij}$  be the degree of similarity between the user's interest in the product  $u_i$  and the interest in the product  $u_j$ . Use the following steps to determine  $c_{ij}$  ( $c_{ij} \in [0, 1], i \neq j, i, j = 1, 2, \dots, n$ ):

- When two products are compared, users rate the two products differently, which is  $c_{ij} + c_{ji} = 1$ ;
- When comparing the same product,  $c_{ii} = 0, i = 1, 2, \dots, n$ ;
- When two products have the same rating,  $c_{ij} = c_{ji} = 0.5$  is recorded.

Since it is the comparison between the scores of commodities, the similarity of interest values between different scores can be determined by Table 3.

Algorithm 1 Collaborative Filtering Recommendation Algorithm Based on Intuitionistic Fuzzy Reasoning(IFR-CF)

Require: User set  $\{u_1, \dots, u_m\}$ , Commodity set  $\{c_1, \dots, c_n\}$ , Score matrix  $r(m \times n)$ , Category  $K$ , Regulatory factor  $\rho$ , Threshold  $\delta$ , Score levels  $\{r_{min}, r_{min} + 1, \dots, r_{max}\}$

Ensure: Recommended commodity set  $R$

$U \leftarrow \emptyset, T \leftarrow \emptyset, R \leftarrow \emptyset$

/\*normalization\*/

for all commodity  $j \leftarrow 1$  to  $n$  do

$r_{max} \leftarrow \max\{r_{1j}, \dots, r_{mj}\}$

$r_{min} \leftarrow \min\{r_{1j}, \dots, r_{mj}\}$

for all user  $i \leftarrow 1$  to  $m$  do

$z_{ij} \leftarrow (r_{ij} - r_{min}) / (r_{max} - r_{min})$

end for

end for

/\*membership calculation\*/

for category  $k \leftarrow 1$  to  $K$  do

$C_k \leftarrow \{c | c \in \{c_1, \dots, c_n\} \text{ and } c \text{ belongs to category } k\}$

for all user  $i \leftarrow 1$  to  $m$  do

$A_{ik} \leftarrow \{z_{ij} | c_j \in C_k\}$

$\mu_{A_{ik}} \leftarrow \frac{\sum_{z_{ij} \in A_{ik}} z_{ij}}{|A_{ik}|}$

if all  $z_{ij} \in A_{ik} \neq 0$  then

$\pi_{A_{ik}} \leftarrow 0$

$\gamma_{A_{ik}} \leftarrow 1 - \mu_{A_{ik}}$

else

$a \leftarrow$  the number of zero elements in  $A_{ik}$

$\bar{\alpha} \leftarrow ((r_{min} + r_{max}) / 2] + r_{max}) / 2$

$\bar{\beta} \leftarrow (r_{min} + [(r_{min} + r_{max}) / 2] - 1) / 2$

$\mu_{A_{ik}}^+ \leftarrow \frac{\sum_{z_{ij} \in A_{ik}} z_{ij} + a \times \frac{\bar{\alpha}}{r_{max} - r_{min}}}{|A_{ik}|}$

$\mu_{A_{ik}}^- \leftarrow \frac{\sum_{z_{ij} \in A_{ik}} z_{ij} + a \times \frac{\bar{\beta}}{r_{max} - r_{min}}}{|A_{ik}|}$

$\pi_{A_{ik}} \leftarrow \mu_{A_{ik}}^+ - \mu_{A_{ik}}^-$

$\gamma_{A_{ik}} \leftarrow 1 - \mu_{A_{ik}} - \pi_{A_{ik}}$

end if

end for

end for

/\*Select similar users for target user  $j^*$ \*/

for all candidate user  $i$  with similar interests to user  $j$  do

$S(i, j) = 1 - \frac{1}{2k} \times \sum_{k=1}^K (|\mu_{A_{ik}} - \mu_{A_{jk}}| + |\gamma_{A_{ik}} - \gamma_{A_{jk}}| + \rho \cdot |\pi_{A_{ik}} - \pi_{A_{jk}}|)$

if  $S(i, j) > \delta$  then

Insert user  $i$  into  $U$

end if

end for

/\*Making commodity recommendations\*/

for all user  $i \in U$  do

Calculate  $n \times n$  priority relation matrix  $P_i$

end for

$P \leftarrow \frac{\sum P_i}{|U|}$

```

Insert all elements of matrix  $P$  into  $T$ 
repeat
     $\lambda \leftarrow \max_{e \in T} e$ 
    for all  $i \leftarrow 1$  to  $n$  do
        for all  $j \leftarrow 1$  to  $n$  and  $j \neq i$  do
            if  $P_{ij} < \lambda$  then
                break
            end if
        end for
    end for
    if  $j > n$  then
         $R \leftarrow i$ 
    end if
end for
Remove all elements whose values equal to  $\lambda$  in  $T$ 
until  $T = \emptyset$  or  $R \neq \emptyset$ 
return  $R$ 
    
```

Algorithm 1 demonstrates the collaborative filtering recommendation algorithm based on intuitionistic fuzzy reasoning. First, the raw scoring data is normalized (lines 2-9). In the score data to be analyzed, we find the maximum (line 4) and minimum (line 5) value of all the scores, then normalize the score data for all users and get the normalized results (lines 6-8). Second, the membership and non-membership of users for different product categories are calculated (lines 10-29). For each category  $k$ , the membership of user  $i$  is calculated using the normalized results of his rating data for products (line 15). If all normalized data for the category  $k$  are not zero (line 16), the values of the intuition index (line 17) and the non-membership (line 18) are obtained. Otherwise, the upper membership degree (line 23) and the lower membership degree (line 24) of the user  $i$  are calculated to obtain the intuitive index (line 25) and the non-membership degree (line 26). Third, we find the target user's neighbors (lines 30-36). Calculate the similarity between each user and the target user (line 32). When the calculated similarity value is higher than the specific parameter value  $\delta$ , the user  $i$  is the neighbor of the target user  $j$  (lines 33-35). Fourth, recommend commodities for the target user (lines 37-57). Here, the priority relation ordering method is used. Refer to Table 3, the priority relation matrix is constructed according to the grading relation of each neighbor user to the predicted products (lines 38-40). For all matrices, calculate the mean value of matrix elements to obtain the final priority relation matrix  $P$ , and then insert all elements of  $P$  into the set  $T$  (lines 41-42). The elements in  $T$  are set as threshold  $\lambda$  one by one in descending order, and  $\lambda$  is compared with the elements of  $P$  in turn. If there exists a row  $i$  whose all elements except the diagonal are not less than  $\lambda$ , the operation is completed, and the commodities corresponding to the  $i$ -th row are recommended to the target user (lines 43-57).

#### D. EXAMPLES

In order to clearly illustrate our algorithm, we take users' movie ratings as an example, apply the algorithm to find neighbor customers through the similarity calculation, judge

TABLE 4. Customer's data on film evaluation.

	Comedy			Action			Science fiction			Forecast object		
	a	b	c	d	e	f	g	h	i	j	k	
Customer 1	3	4	2	4	0	3	3	0	3	4	5	
Customer 2	2	0	3	3	5	3	0	4	2	3	3	
Customer 3	5	4	1	0	4	2	2	0	2	3	5	
Customer 4	4	3	0	3	5	2	2	4	?	?	?	

TABLE 5. Data for intuitionistic fuzzy inference.

	$D_1$	$D_2$	$D_3$
Customer 1	(0.60, 0.40)	(0.47, 0.37)	(0.30, 0.45)
Customer 2	(0.33, 0.50)	(0.73, 0.27)	(0.40, 0.35)
Customer 3	(0.67, 0.33)	(0.40, 0.43)	(0.20, 0.55)
Customer 4	(0.47, 0.37)	(0.67, 0.33)	(0.60, 0.40)

their preferences for movies according to the neighbor users' ratings of the movie  $i, j, k$ , and based on this judgment, recommend the movie they are most interested in to the target user.

In the example, missing data is recorded as 0 point. Missing data may be any of the ratings {1, 2, 3, 4, 5}, and the probability of each score is the same. The treatment of missing data in this paper does not calculate specific prediction values but predicts the membership of the fuzzy set by the possible selection value of the missing data. It is generally considered that a score of 3 or above is of interest, and a score of less than 3 is not of interest. The upper membership degree and lower membership degree of each fuzzy set are calculated based on the intuitionistic fuzzy rough set, and then the intuition index is calculated. Take the data set [2,0,3] corresponding to customer 2 as an example, where 0 is the missing data, and score 4 is the average score in that the user is interested. The normalization method is used to calculate the upper membership of the fuzzy set is 0.6; The score 1.5 is the average score when the user is not interested. The normalization method is used to calculate the lower membership of the fuzzy set at 0.43, and then the intuition index is 0.17. In the case of missing data, the membership of the fuzzy set is 0.33, the intuition index is 0.17, and the non-membership degree of the fuzzy set is 0.5. Therefore, intuitionistic fuzzy reasoning is performed on the set of data [2,0,3], and the result is (0.33, 0.5). The same processing is performed on the data in Table 4 to obtain the intuitionistic fuzzy inference data shown in Table 5.

Through the intuitionistic fuzzy processing of missing data, the intuition index is taken into account, which reduces the possibility of data result distortion, makes the data better reflect the actual situation, and lays a good data foundation for the calculation of similarity between users. This method increases the accuracy of the recommended results.

We use Equation 6 to calculate the similarity between users and get  $S(1, 4) = 0.77, S(2, 4) = 0.83, S(3, 4) = 0.77$  respectively. Assume that  $\delta = 0.8$ , that is, the similarity value greater than 0.8 is the neighbor customer, so customer 2 is the neighbor of customer 4.

When recommending a movie to a target customer, the priority ordering method in intuitionistic fuzzy reasoning is used to determine the priority of the predicted movie, and the priority relation matrix of  $i, j, k$  is obtained.

$$C = \begin{pmatrix} 0 & 0.4 & 0.4 \\ 0.6 & 0 & 0.5 \\ 0.6 & 0.5 & 0 \end{pmatrix}$$

After determining the priority of the three movies, the truncation matrix is obtained by taking the threshold of 0.6 and 0.5 respectively.

$$C_{0.6} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad C_{0.5} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

When  $\lambda = 0.5$ , except the diagonal, the values in the second and third rows of the matrix are 1, which means that users have a higher similarity of interest in the two movies of  $j$  and  $k$ , and the interest similarity values are above 0.5.

**VI. EXPERIMENT**

To measure the recommendation effect of our algorithm, we compare it with traditional recommendation algorithms and other improved recommendation algorithms from the perspective of user interest. The algorithms to be compared together are as follows:

- Traditional collaborative filtering algorithm(ICF) [1]: The final recommendation is generated by user’s nearest neighbors. Firstly, the correlation between items is calculated. Then the score of the target user can be predicted by the scores of neighbor users, and the recommendation can be made;
- Collaborative filtering algorithm based on time factor(TFCF) [8]: By introducing time factor, the value of different data can be better divided. The more recent the evaluation, the more likely it is to reflect the current interest of the target users. Since users’ interests are not fixed, and they are only interested in a limited number of items at a certain time, the items that users prefer in a short period have a higher degree of similarity.
- Hybrid recommendation algorithm based on collaborative filtering and implicit semantic model(LFIRS-CF) [11]: By using the liner fusion method, the collaborative recommendation algorithm based on item similarity and the crypto-semantic model recommendation algorithm based on user interest migration are fused to improve the recommendation accuracy.
- Collaborative filtering recommendation algorithm based on fusion of user interests and rating differences (CF-UIRD) [12]: This collaborative filtering algorithm considers both the similarity measure of user interest change and the score difference similarity measure. By combining these two factors, the algorithm can describe user preference more accurately.

**TABLE 6. The characteristics of datasets (UN=User number, NOPD=Number of products, NOPI=Number of points, SR=Scoring range, S=Sparsity).**

Data set	UN	NOPD	NOPI	SR	S
Movielens	943	1682	100000	{1,2,3,4,5}	0.063
Jester	24983	100	1761439	[-10,10]	0.705

**A. DATASET**

Experimental data are extracted from the *Movielens* dataset and *Jester* dataset. The *Movielens* dataset is a data set on movie ratings, which contains 100,000 user ratings for 1682 movies with a rating range from 1 to 5. The *Jester* dataset includes 73,421 users’ rating data on 100 jokes, totaling about 4.1 million records, and each rating with an integer value range from -10 to 10. In our experiment, some records are extracted from the two datasets, and Table 6 describes some basic characteristics of the data.

**B. MEASUREMENT METRICS**

We use the mean absolute error (MAE) as a measure which evaluates the accuracy of the algorithms by calculating the average error between the predicted preferences and users’ actual preference scores. When the MAE value is smaller, the error is smaller, so that the recommendation accuracy is higher.

$$MAE(X, h) = \frac{1}{m} \sum_{i=1}^m |h(x_i) - y_i| \tag{10}$$

$X$  represents user set,  $m$  represents the number of users,  $h(x_i)$  represents the predicted score of the calculation system, and  $y_i$  represents the actual score of the user.

**C. ANALYSIS OF EXPERIMENTAL RESULTS**

We experimentally vary the number of nearest neighbors to observe the MAE values of different algorithms and verify their accuracy. In the first group of experiments, 500 users were selected from the two data sets to test the accuracy of the algorithms. The comparison experiments are shown in Figure1 and Figure2.

It can be observed from Figures 1 and 2 that as the number of nearest neighbors increases, the MAE value of each algorithm shows a decreasing trend, which indicates that the recommended accuracy is getting higher and higher. In addition, compared with the traditional algorithm (ICF). The MAE values of other algorithms are lower, which indicates that the traditional algorithm is improved by introducing time factor, hidden semantic model, user interest and scoring differences, and intuitive fuzzy reasoning, thus increases the accuracy of the traditional collaborative filtering algorithm. At the same time, it can be found in the value comparison that the collaborative filtering recommendation algorithm (IFR-CF) under intuitionistic fuzzy reasoning has a lower average absolute error value and higher recommendation accuracy than other algorithms, which indicates that the introduction

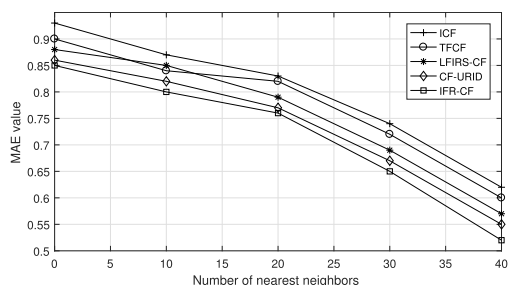


FIGURE 1. Comparison of MAE values of various algorithms for *MovieLens* dataset with 500 users.

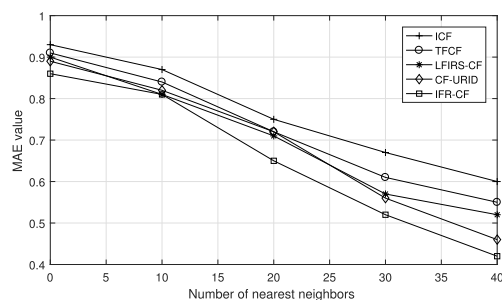


FIGURE 4. Comparison of MAE values of various algorithms for *Jester* dataset with 1000 users.

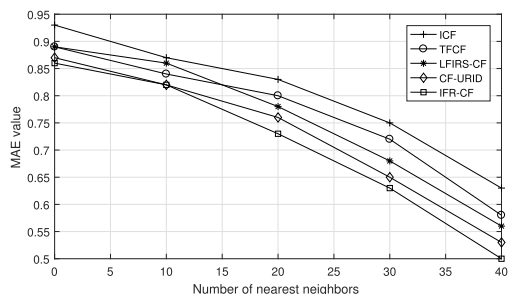


FIGURE 2. Comparison of MAE values of various algorithms for *Jester* dataset with 500 users.

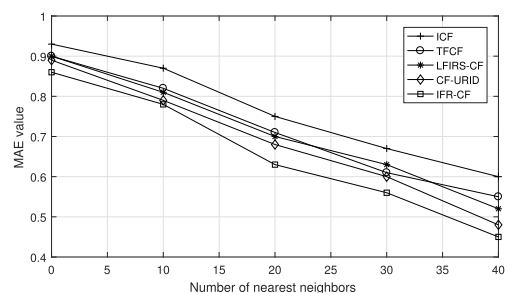


FIGURE 3. Comparison of MAE values of various algorithms for *MovieLens* dataset with 1000 users.

of intuitionistic fuzzy reasoning into collaborative filtering recommendation algorithm is reasonable and achieves better improvement effect.

In the second group of experiments, 1000 users were selected from the two datasets to check the accuracy of the algorithm. The comparison experiments are shown in Figures 3 and 4.

Through the comparison, the nearest neighbors selected are more representative when the amount of users increases, so the overall accuracy is improved compared to the previous group of users; In addition, the MAE values of the collaborative filtering recommendation algorithm under intuitionistic fuzzy reasoning in the two data sets are both lowest. When the number of nearest neighbors is 40, the MAE values calculated in the *MovieLens* data set and the *Jester* data set are 0.42 and 0.45, respectively, and there is a certain improvement in the recommended accuracy, the accuracy is 15% and 18% higher than the traditional algorithm.

In general, the advantages of the collaborative algorithm based on intuitionistic fuzzy reasoning are mainly reflected in the following two points: First, from the perspective of the description of the problem, the reference of intuitionistic fuzzy reasoning can make those unclear boundary problems be described using mathematical language, and break the limitations of the traditional collaborative filtering algorithm in semantic description, so that more objective problems can be quantitatively solved. Second, from the perspective of the algorithm theory, the recommendation algorithm under intuitionistic fuzzy reasoning defines three parameters: membership degree, non-membership degree and intuition index, and these parameters are used to the calculation. The missing scores are predicted and supplemented by using the intuition index, which makes the calculation result more accurate and reduces the calculation amount of recommended products in the traditional algorithm.

### VII. CONCLUSION

This paper combines the intuitionistic fuzzy inference method with collaborative filtering recommendation algorithm, introduces non-membership degree function and intuition index into the calculation of collaborative filtering recommendation algorithm, and draws the following conclusions:

- The introduction of intuitionistic fuzzy reasoning breaks the limitation of the traditional collaborative filtering recommendation algorithm in terms of scalability and accuracy. It describes the fuzzy problem with unclear boundaries in reality and expands the scope of research.
- This method not only considers the problem more comprehensively but also introduces the membership degree and non-membership degree parameters, improves the traditional similarity algorithm, replaces the similarity coefficient with the characteristic coefficient of intuitionistic fuzzy reasoning, which makes the final recommendation result more accurate, more consistent with the actual situation hidden in raw data.
- When calculating the intuition index, this method reasonably uses the hesitation parameter to interpret the missing data, which provides a solution to the problem of missing scoring data.

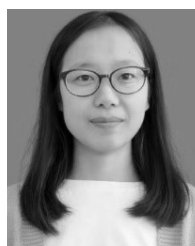


## REFERENCES

- [1] A. Sattar, M. A. Ghazanfar, and M. Iqbal, "Building accurate and practical recommender system algorithms using machine learning classifier and collaborative filtering," *Arabian J. Sci. Eng.*, vol. 42, no. 8, pp. 3229–3247, Aug. 2017.
- [2] X. Wu, X. Yuan, C. Duan, and J. Wu, "A novel collaborative filtering algorithm of machine learning by integrating restricted Boltzmann machine and trust information," *Neural Comput. Appl.*, vol. 31, no. 9, pp. 4685–4692, Sep. 2019.
- [3] T. Mohammadpour, A. M. Bidgoli, R. Enayatifar, and H. H. S. Javadi, "Efficient clustering in collaborative filtering recommender system: Hybrid method based on genetic algorithm and gravitational emulation local search algorithm," *Genomics*, vol. 111, no. 6, pp. 1902–1912, Dec. 2019.
- [4] A. Laishram, V. Padmanabhan, and R. P. Lal, "Analysis of similarity measures in user-item subgroup based collaborative filtering via genetic algorithm," *Int. J. Inf. Technol.*, vol. 10, no. 4, pp. 523–527, Dec. 2018.
- [5] V. Gnanaprasagam and I. Egambaram, "Dynamic Web service recommendations based on model based collaborative filtering methods and genetic algorithm," in *Proc. 4th Int. Conf. Adv. Recent Technol. Commun. Comput. (ARTCom)*, Bengaluru, India, 2012, pp. 224–228.
- [6] D. Mustafi and G. Sahoo, "A hybrid approach using genetic algorithm and the differential evolution heuristic for enhanced initialization of the  $k$ -means algorithm with applications in text clustering," *Soft Comput.*, vol. 23, no. 15, pp. 6361–6378, Aug. 2019.
- [7] A. Khan, Z. ur Rehman, M. A. Jaffar, J. Ullah, A. Din, A. Ali, and N. Ullah, "Color image segmentation using genetic algorithm with aggregation-based clustering validity index (CVI)," *Signal, Image Video Process.*, vol. 13, no. 5, pp. 833–841, Jul. 2019.
- [8] P. Zhang, Z. Zhang, T. Tian, and Y. Wang, "Collaborative filtering recommendation algorithm integrating time windows and rating predictions," *Int. J. Speech Technol.*, vol. 49, no. 8, pp. 3146–3157, Aug. 2019.
- [9] J. Wang, X. Li, W. Wu, and Y. Liu, "An algorithm of collaborative filtering based on svd and trust factors," *J. Chin. Comput. Syst.*, vol. 6, p. 22, 2017.
- [10] L. Jiang, Y. Cheng, L. Yang, J. Li, H. Yan, and X. Wang, "A trust-based collaborative filtering algorithm for E-commerce recommendation system," *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 8, pp. 3023–3034, 2019.
- [11] C. Xu, L. Xu, Y. Lu, H. Xu, and Z. Zhu, "E-government recommendation algorithm based on probabilistic semantic cluster analysis in combination of improved collaborative filtering in big-data environment of government affairs," *Pers. Ubiquitous Comput.*, vol. 23, nos. 3–4, pp. 475–485, Jul. 2019.
- [12] R. M. D'Addio, M. A. Domingues, and M. G. Manzato, "Exploiting feature extraction techniques on users' reviews for movies recommendation," *J. Brazilian Comput. Soc.*, vol. 23, no. 1, pp. 1–16, Dec. 2017.
- [13] J. Iwanaga, N. Nishimura, N. Sukegawa, and Y. Takano, "Improving collaborative filtering recommendations by estimating user preferences from clickstream data," *Electron. Commerce Res. Appl.*, vol. 37, Sep. 2019, Art. no. 100877.
- [14] L. Huang, Z.-L. Zhao, C.-D. Wang, D. Huang, and H.-Y. Chao, "LSCD: Low-rank and sparse cross-domain recommendation," *Neurocomputing*, vol. 366, pp. 86–96, Nov. 2019.
- [15] Z. Batmaz and C. Kaleli, "AE-MCCF: An autoencoder-based multi-criteria recommendation algorithm," *Arabian J. Sci. Eng.*, vol. 44, no. 11, pp. 9235–9247, Nov. 2019.
- [16] S. Yu, M. Yang, Q. Qu, and Y. Shen, "Contextual-boosted deep neural collaborative filtering model for interpretable recommendation," *Expert Syst. Appl.*, vol. 136, pp. 365–375, Dec. 2019.
- [17] X. Deng, F. Zhuang, and Z. Zhu, "Neural variational collaborative filtering with side information for top-K recommendation," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 11, pp. 3273–3284, Nov. 2019.
- [18] T. Xiao and H. Shen, "Neural variational matrix factorization for collaborative filtering in recommendation systems," *Int. J. Speech Technol.*, vol. 49, no. 10, pp. 3558–3569, Oct. 2019.
- [19] K. Liao, N. Sun, and J. OuYang, "Improved recommendation sorting of collaborative filtering algorithm," in *Proc. 2nd Int. Conf. Autom., Mech. Control Comput. Eng. (AMCCE)*, Beijing, China: Atlantis Press, 2017.
- [20] K. Xue and J. Wang, "Collaborative filtering recommendation algorithm for user interest and relationship based on score matrix," in *Proc. Int. Conf. Math., Modelling, Simulation Algorithms (MMSA)*, Atlantis Press, Jan. 2018, pp. 228–232.
- [21] Y. Ju, L. Bailin, and Z. Zhixiang, "A collaborative filtering recommendation algorithm with improved similarity calculation," in *Proc. 2nd Int. Conf. Sensor Netw. Comput. Eng. (ICSNCE)*, Xi'an, China: Atlantis Press, 2018, vol. 3, no. 1, pp. 97–100.
- [22] M. Esmaili and E. Eslami, "Intuitionistic fuzzy reasoning using the method of optimizing the similarity of truth tables," *Soft Comput.*, vol. 23, no. 18, pp. 8765–8774, Sep. 2019.
- [23] A. Asim, R. Nasar, and T. Rashid, "Correlation coefficient of intuitionistic hesitant fuzzy sets based on informational energy and their applications to clustering analysis," *Soft Comput.*, vol. 23, no. 20, pp. 10393–10406, Oct. 2019.
- [24] L. Zhang, J. Zhan, Z. Xu, and J. C. R. Alcantud, "Covering-based general multigranulation intuitionistic fuzzy rough sets and corresponding applications to multi-attribute group decision-making," *Inf. Sci.*, vol. 494, pp. 114–140, Aug. 2019.
- [25] E. Szmidt and J. Kacprzyk, "Distances between intuitionistic fuzzy sets," *Fuzzy Sets Syst.*, vol. 114, no. 3, pp. 505–518, Sep. 2000.
- [26] P. Grzegorzewski, "Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on the Hausdorff metric," *Fuzzy Sets Syst.*, vol. 148, no. 2, pp. 319–328, Dec. 2004.
- [27] S.-M. Chen, "Measures of similarity between vague sets," *Fuzzy Sets Syst.*, vol. 74, no. 2, pp. 217–223, Sep. 1995.



**YANJU ZHANG** received the Ph.D. degree from Liaoning Technical University, Huludao, China. She is currently a Lecturer with the School of Business Administration, Liaoning Technical University. Her current research interests include intelligent decision and recommendation systems.



**YUE WANG** is currently pursuing the master's degree with the School of Business Administration, Liaoning Technical University, Huludao, China. Her research interests include business management and recommendation systems.



**SHIQIN WANG** is currently pursuing the master's degree with the School of Business Administration, Liaoning Technical University, Huludao, China. His research interests include task assignment in spatial crowdsourcing and recommendation systems.