

Article

Measuring Product Similarity with Hesitant Fuzzy Set for Recommendation

Chunsheng Cui ¹, Jielu Li ¹ and Zhenchun Zang ^{2,*}

¹ College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046, China; cuichunsheng@huel.edu.cn (C.C.); 20100732@huel.edu.cn (J.L.)

² School of Mathematics and Statistics, Zhoukou Normal University, Zhoukou 466001, China

* Correspondence: 20100712@huel.edu.cn

Abstract: The processing of a sparse matrix is a hot topic in the recommendation system. This paper applies the method of hesitant fuzzy set to study the sparse matrix processing problem. Based on the uncertain factors in the recommendation process, this paper applies hesitant fuzzy set theory to characterize the historical ratings embedded in the recommendation system and studies the data processing problem of the sparse matrix under the condition of a hesitant fuzzy set. The key is to transform the similarity problem of products in the sparse matrix into the similarity problem of two hesitant fuzzy sets by data conversion, data processing, and data complement. This paper further considers the influence of the difference of user ratings on the recommendation results and obtains a user's recommendation list. On the one hand, the proposed method effectively solves the matrix in the recommendation system; on the other hand, it provides a feasible method for calculating similarity in the recommendation system.



Citation: Cui, C.; Li, J.; Zang, Z. Measuring Product Similarity with Hesitant Fuzzy Set for Recommendation. *Mathematics* **2021**, *9*, 2657. <https://doi.org/10.3390/math9212657>

Academic Editors: Zhen-Song Chen, Witold Pedrycz, Lesheng Jin, Rosa M. Rodriguez, Luis Martínez López and Michael Gregory Voskoglou

Received: 24 August 2021
Accepted: 15 October 2021
Published: 20 October 2021

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



Copyright: © 2021 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/>).

Keywords: hesitant fuzzy set; recommendation system; sparse matrix; similarity

1. Introduction

In the era of Big Data, information overload has become a common phenomenon that plagues people. As an effective way to solve information overload, the recommendation system emphasizes discovering users' hobbies and guiding them to discover their information needs. A sound recommendation system can provide users with personalized services and can establish close relationships with them, making users rely on recommendations. Studies have indicated that the e-commerce recommendation system plays a vital role in consumer decision making and enterprise product sales [1,2].

Recommendation systems utilize the input of recommendations based on the information of users and products, the recommendation algorithm, and the recommendation output. Among them, the input of recommendations is the "source" of the whole system, as it largely determines the user experience and influences the system's recommendation quality. The recommendation algorithm is the "core" of the entire system, reflecting its intelligence. The mainstream recommendation algorithms include content-based recommendations [3,4], collaborative filtering recommendations [5], knowledge-based recommendations [6], and hybrid recommendations [7]. Finally, the recommendation output is the "face" of the system, reflecting its quality.

Irrespective of the recommendation algorithm, obtaining input data is the primary step of recommendations. Two kinds of input data exist in the system: Implicit browsing input and explicit rating input. The former requires mining users' browsing time, browsing paths, browsing behavior, and other implicit information, which has a high degree of uncertainty and passive features; hence, they are extensively studied by scholars. Meanwhile, the latter is evaluated by historical ratings provided by users, constrained by the number of data and the efficiency of the data. Typically, after purchasing a product, the user provides a product score to express their satisfaction and preferences. For instance, the Movielens,

Taobao, and JingDong scoring systems are adopted at level 5. However, the sparsity of the scoring matrix renders inconvenience to the improvement of the recommendation quality.

In the explicit scoring input, the sparsity of the scoring matrix brings much inconvenience to the accurate implementation of recommendations. In e-commerce websites, the number of user-rated products accounts for only a small part of the total number of products, leading to high data sparsity. Consequently, the sparsity of user rating data leads to significant errors in the similarity calculation of users and products, as the accuracy of user score prediction decreases sharply. Based on this, many scholars have proposed many methods to improve the sparse matrix. For example, Zhang et al. [8] used a random algorithm to deal with cold start problems when the data were sparse, and when the data reached a certain level, the hybrid algorithm was applied to an incremental recommendation. Li et al. [9] proposed a hybrid recommendation algorithm based on content and user collaborative filtering. Taking advantage of collaborative filtering, when the number of users and the evaluation level are large, the user scoring data matrix becomes relatively dense to reduce the sparsity of the matrix and to enable a more accurate collaborative filtering. Zhang et al. [10] proposed a new depth variational matrix factorization (DVMF) recommendation method for large-scale sparse datasets, and obtained the potential characteristics of users and items respectively through the depth nonlinear structure. Based on the potential factors combined with the matrix factorization method, an optimization method of the DVMF algorithm has been proposed. Wang et al. [11] put forward a recommendation model of prior relations for low-rank sparse matrix factorization, which predicted users' ratings of the item by learning the sum of the low-rank and sparse matrices, then effectively reducing sparsity and cold start problems using prior information. Liu et al. [12] proposed a new item recommendation algorithm based on a pattern recognition and statistical model to analyze and predict user behaviors. This algorithm can be applied to sparse user behavior datasets, avoiding the problems faced by collaborative filtering algorithms when the datasets are sparse. Huang et al. [13] devised a new CDCF algorithm, the low-rank sparse cross-domain (LSCD) recommendation algorithm, to extract potential feature matrix of users and items for each domain, instead of decompressing the matrix of each domain into three low-dimensional matrices by three factors to solve the problem of sparse data. To sum up, there are three primary approaches to improving the sparse matrix: (1) Increase data while maintaining an unchanged scale [14,15]; (2) reduce scale and data [16,17]; and (3) use neural networks and deep learning methods to predict user ratings for improving the sparse matrix [18]. These methods have three drawbacks in sparse matrix processing. First, in the process of complementing data, various complement strategies increase the uncertainty of information. Second, it may lose part of the helpful data information in the process of dimension reduction. Finally, it is relatively challenging to realize the deep mining of user and resource information. In this case, exploring the new sparse matrix solution is one of the core elements of e-commerce recommendation input research.

As an extension of fuzzy sets [19], hesitant fuzzy sets have been applied to many decision-making [20,21] problems: Mardani et al. [22] extended a new fuzzy approach under the hesitant fuzzy set (HFS) approach using stepwise weight assessment ratio analysis (SWARA) and the weighted aggregated sum product assessment (WASPAS) method to evaluate and rank the critical challenges of DT intervention to control the COVID-19 outbreak. Colak et al. [23] proposed an integrated MCDM model consisting of the Delphi, analytic hierarchy process (AHP), and Vİsekriterijumska Optimizcija I Kompromisno Resenje (VIKOR) methods to evaluate EST alternatives for Turkey under a hesitant fuzzy environment. Sahu et al. [24] found that the hesitant fuzzy-based symmetrical technique of the analytic hierarchy process (AHP) and the technique for order of preference by similarity to the ideal solution (TOPSIS) is an effective methodology for evaluating web applications' durability. Pratibha et al. [25] proposed a novel framework based on the COPRAS (complex proportional assessment) method and the SWARA (stepwise weight assessment ratio analysis) approach to evaluate and select the desirable sustainable supplier within the

HFSs context. Although hesitant fuzzy sets have been used in many fields, they are rarely used in the field of recommendation systems.

Considering the above ideas and the application of fuzzy tools to recommendations [26,27], this paper discusses how to make full use of known data to obtain high-quality recommendation outcomes without losing information in a sparse matrix. Its primary contributions are as follows: (1) The similarity between hesitant fuzzy theory and recommendation system is discussed; (2) the hesitant fuzzy set theory is applied to describe the embedded historical ratings in the recommendation system. Such an idea can ensure the similarity relationships between products without losing the sparse matrix rating information, provide a new way for sparse matrix processing and similarity discussion in the recommendation system, and find a new field for the research of hesitant fuzzy sets.

This paper is structured as follows. Section 2 mainly analyzes the characteristics of sparse matrix data and explores a suitable processing approach according to those characteristics. Section 3 introduces the hesitant fuzzy set theory to realize seamless docking between the fuzzy set and the recommendation system. Section 4 architects a measurement model of product similarity in the electronic commerce recommendation system with the thought of hesitant fuzzy sets. Section 5 extracts the data from MovieLens to conduct empirical research. Section 6 compares the results based on user recommendation with the results based on product similarity in the preceding part to verify the effectiveness of the proposed method, while Section 7 provides the conclusions and prospects.

2. Feature Description of Sparse Matrix in the E-commerce Recommendation System

The sparse matrix in the e-commerce recommendation system is represented by the sparsity of the user's rating matrix. The basis of e-commerce recommendation utilizes the rating matrix to dig the similarity of users or products and then to produce high-quality recommendation strategies. Typically, it can attain two different recommendation strategies with the aid of the user's rating matrix; one is collaborative filtering recommendation based on users, and the other is collaborative filtering recommendation based on products. The former applies the rating matrix to obtain the recommended user's nearest neighbor, and the latter helps recommend products to find its similar products. The essence of the two recommendations is seeking similarity, but the former studies the similarity of users, while the latter studies the similarity of products.

In the e-commerce system, the user rating matrix indicates the feature of sparsity, but also shows the 4V feature of Big Data. The formation of these data is accompanied by many uncertain factors, such as the attitude of users when providing rating data, which affects the reliability and authenticity of the data. The personality characteristics of users have a certain impact on the display of data; users may hesitate between different rating levels when rating. Based on the differences in the environment, situation, and time, the rating results may even be contradictory. For the same product, different users may separately use five or four points to express their satisfaction, while the same user in different situations can express their satisfaction or dissatisfaction with four points. Therefore, when the rating data are uncertain, the recommendation system should retain existing data as much as possible to avoid information loss.

Taking product ratings as an example, irrespective of the number of ratings, all data should be used as far as possible in later studies. If a product has multiple ratings, it can be considered that the evaluation of the product is hesitating among multiple values, and each value can reflect the actual attributes of the product from one angle or side. In the product-based collaborative filtering recommendation algorithm, it is necessary to describe the similarity between products. Typically, this similarity calculation needs to ensure that two products have the same evaluation number, while the massive product evaluation number in the network cannot reach an agreement. Fortunately, the hesitant fuzzy set [28] adequately expresses the thought of "dithering" in the progress of ratings, and it provides a convenient means to solve this problem.

There is much correspondence between hesitant fuzzy set theory and the e-commerce recommendation system, as shown in Table 1.

Table 1. Similarity between hesitant fuzzy set theory and the recommendation system.

	Hesitant Fuzzy Set	Recommendation System
Concept description	Describe a thing with a set of membership degrees	Describe a product with a set of history ratings
Characteristic	Use a set of numbers to express the degree of uncertainty of the object to be evaluated	Use a set of historical ratings to express the uncertainty of the user’s attitude
Data form	The difference in membership stems from the different attitudes of the raters	The difference in ratings stems from the user’s satisfaction with the product
Key element	Membership functions	User ratings
Numerical comparison	The degree of membership can be compared; the same degree of membership means the same attitude of users	User ratings can be compared horizontally; the same user rating represents the same level of user satisfaction

Notably, the similarity between products or between users can be studied based on the rating matrix. Under the condition of a hesitant fuzzy set, the essence of the two is the same. If a product is defined as an element in the hesitant fuzzy set, we can discuss the similarity between the products and then produce the collaborative filtering recommendation algorithm based on products. Meanwhile, if a user is defined as an element in the hesitant fuzzy set, we can explore the similarity between the users and then produce the collaborative filtering recommendation algorithm based on users. Without losing generality, this paper explores the former case.

3. The Introduction of a Hesitant Fuzzy Set

The hesitant fuzzy set was put forward by Spanish scholar Torra [29] in 2010, and it is a further extension of the fuzzy set theory. The idea of a hesitant fuzzy set is that people hover among multiple possible values when deciding the membership of an element belonging to a certain set, and then the multiple values are listed as membership. Thus, the hesitant fuzzy set can more carefully describe the uncertain characteristics of a decision-maker’s understanding of things.

3.1. Basic Definition

Definition 1 [29]. Let X be a reference set. Hesitant fuzzy set (HFS) A is a set of different numbers of membership functions $h_A(x)$ on X valued on $[0, 1]$.

To be easily understood, Xu and Xia [30] expressed HFS as a mathematical symbol:

$$A = \{ \langle x, h_A(x) \rangle \mid x \in X \}, \tag{1}$$

where $h_A(x) = \{ \gamma \mid \gamma \in h_A(x) \}$ is a set of some different values in $[0, 1]$, γ represents the possible membership degree of the element $x \in X$ to A , and $h_A(x)$ is called a hesitant fuzzy element (HFE) [30], which is a basic unit of HFS.

Example 1 [28]. Let $X = \{x_1, x_2, x_3\}$ be a fixed set, $h_A(x_1) = \{0.2, 0.4, 0.5\}$, $h_A(x_2) = \{0.3, 0.4\}$ and $h_A(x_3) = \{0.3, 0.2, 0.5, 0.6\}$ be the HFEs of $x_i (i = 1, 2, 3)$ to a set A respectively. Then A can be considered as a HFS:

$$A = \{ \langle x_1, \{0.2, 0.4, 0.5\} \rangle, \langle x_2, \{0.3, 0.4\} \rangle, \langle x_3, \{0.3, 0.2, 0.5, 0.6\} \rangle \}$$

Meanwhile, Torra [26] provided some special HFEs for x in X :

Empty set: $h = \{0\}$, denoted O^* for simplification.

Full set: $h = \{1\}$, denoted as I^* .

Complete ignorance for a $x \in X$ (all are possible): $h = [0, 1]$.

Nonsense set: $h(x) = \phi$.

In the hesitation fuzzy set theory, given a reference set, the membership function does not provide only one value, but rather a set of them, which provides a way of modeling hesitation. In the e-commerce recommendation system, a large number of users will provide a large number of ratings, and each rating actually has a certain degree of hesitation. Then, the rating for a product is actually a set of multiple user ratings, that is, the set of multiple fuzzy numbers, which just constitutes a hesitant fuzzy set. Hence, we can interpret the different users' ratings for a given item as the hesitation about the item.

3.2. Similarity

Similarity measures are fundamentally important in a variety of scientific fields, including decision making, pattern recognition, machine learning, and market prediction, and lots of studies have been conducted regarding this issue of hesitant fuzzy sets. Xu and Xia [31] originally developed a series of distance measures for hesitant fuzzy sets based on the proposed corresponding similarity measures.

Definition 2 [32]. Let A_1 and A_2 be two HFSs on X , then the distance between A_1 and A_2 is defined as $d(A_1, A_2)$, which satisfies the following properties:

- (1) $0 \leq d(A_1, A_2) \leq 1$.
- (2) $d(A_1, A_2) = 0$ if and only if $A_1 = A_2$.
- (3) $d(A_1, A_2) = d(A_2, A_1)$.

Definition 3 [32]. Let A_1 and A_2 be two HFSs on X , then the similarity between A_1 and A_2 is defined as $S(A_1, A_2)$, which satisfies the following properties:

- (1) $0 \leq S(A_1, A_2) \leq 1$.
- (2) $S(A_1, A_2) = 1$ if and only if $A_1 = A_2$.
- (3) $S(A_1, A_2) = S(A_2, A_1)$.

In fact, the calculation of the similarity of hesitant fuzzy sets has a precondition that the number of elements in two sets is equal. However, this is hard to guarantee in reality, which is also a manifestation of the diversity of hesitant fuzzy sets. Thus, the number of elements should be complemented before similarity calculation. There are many approaches to complement elements in hesitant fuzzy sets, such as the mean value approach, the modal number approach, and so on, but how to ensure the quality of the complement is a key problem in this paper.

By analyzing Definitions 2 and 3, it is noted that $S(A_1, A_2) = 1 - d(A_1, A_2)$. Therefore, the distance measurement formula can obtain the measure of similarity in the hesitant fuzzy set. The shorter the distance between the two sets, the higher the similarity between them. The similarity measurement of hesitant fuzzy values is similar to that of hesitant fuzzy sets. In fact, it only needs to measure the distance of each membership function between two hesitant fuzzy values. Xu et al. [33] provided the similarity calculation formula for two hesitant fuzzy values, h_1 and h_2 , based on the distance measurement formula:

$$S(h_1, h_2) = \frac{\sum_{i=1}^l (h_1^{\sigma(i)} h_2^{\sigma(i)})}{\left(\sum_{i=1}^l (h_1^{\sigma(i)})^2 \cdot \sum_{i=1}^l (h_2^{\sigma(i)})^2 \right)^{1/2}}, \tag{2}$$

where $h_1^{\sigma(i)}$ and $h_2^{\sigma(i)}$ are the i th largest values in h_1 and h_2 .

Obviously, the similarity measure of the hesitant fuzzy value adopts the idea of the Pearson’s similarity measure. This measurement method effectively avoids the uncertain information generated during the measurement of two hesitant fuzzy values.

4. Construction of a Similarity Model of Sparse Matrix Products

4.1. Affiliation of User Ratings

In an e-commerce recommendation system, there may be some differences in the rating results provided by users with different personality characteristics. For instance, some users think that four points is already high, while others think this is poor. Therefore, it is necessary to preprocess the user’s rating to eliminate differences and uncertainties in the rating.

Here, R_{ij} indicates the rating of $User_i$ on $Item_j$, and $\bar{R}_i = \frac{\sum_{j=1}^m R_{ij}}{m}$ is the average rating of all products rated by $User_i$ in the system. Then, the membership degree γ_{ij} of R_{ij} is defined as:

$$\gamma_{ij} = \frac{R_{ij} - \min R}{\max R - \min R}, \gamma_{ij} \in [0, 1], \tag{3}$$

where R represents the rating system in the recommendation system. Generally, Taobao and JingDong adopt a five-point system, so $\max R = 5$ and $\min R = 0$. Therefore, the membership degree of a five-point system is $\gamma_{ij} = \frac{R_{ij}}{5}$. There are also three-point and ten-point evaluation methods in evaluating enterprise after-sales services. In fact, a user’s rating of the product is a typical expression of the degree of membership, which intuitively expresses their satisfaction.

4.2. Product Rating Representation

Based on the above discussion, all of the ratings obtained by $Item_j$ in the system can be expressed as a hesitant fuzzy set:

$$h(Item_j) = \{ \gamma_{ij}, \gamma_{2j}, \dots, \gamma_{l_j j} \}, \tag{4}$$

where l_j represents the total number of ratings that product $Item_j$ receives. Obviously, l_j is an uncertain value, reflecting the sparse degree of the rating matrix.

In the collaborative filtering recommendation, the recommendation among the products is converted into a recommendation among the hesitant fuzzy set, and the product similarity is converted into similarity among the hesitant fuzzy set.

4.3. Horizontal Comparison of Products

Considering the objective existence of rating matrix, the number of elements in $h(Item_j)$ and $h(Item_k)$ is usually $l_j \neq l_k$, but when calculating the similarity, two hesitant fuzzy sets usually need to have the same hesitant fuzzy elements. How to fill in the short hesitation fuzzy sets has become one of the urgent problems to be solved. In order to solve the problem and calculate effectively, the following agreement is made.

The elements in $h(Item_j)$ and $h(Item_k)$ are arranged in ascending order. If and only if $\gamma_{ij} = \gamma_{ik} (i = 1, 2, \dots, l)$, $h(Item_j) = h(Item_k)$, here, γ_{ij}, γ_{ik} represents the i -th element, which is ordered by the ascending order in hesitant fuzzy set of $h(Item_j)$ and $h(Item_k)$. Elements are added into the hesitant fuzzy set, which has fewer elements until $l = \max\{l_1, l_2, \dots, l_j, l_k, \dots\}$.

In the personalized recommendation system, the ultimate purpose is to recommend the product $Item_j$ to $User_i$; thus, the preference of $User_i$ determines how to add elements. There are two strategies for adding elements in the hesitant fuzzy set; one is the forward strategy, and the other is the backward strategy:

- (1) Forward strategy: If the average of all ratings $R_{ij} (j = 1, 2, \dots)$ of historical products from $User_i$, which is expressed as \bar{R}_i , meet $\bar{R}_i \leq 4$, indicating that $User_i$ is a pessimistic

- user, $(l - l_j)\gamma_{l,j}$ is added before the first element of $h(Item_j)$. Hence, $\gamma_{1j} = \gamma_{2j} = \dots = \gamma_{l-l_j,j} = \gamma_{1j}$.
- (2) Backward strategy: If the average of all ratings $R_{ij}(j = 1, 2, \dots)$ of historical products from $User_i$, which is expressed as \bar{R}_i , meet $\bar{R}_i > 4$, indicating that $User_i$ is an optimistic user, $(l - l_j)\gamma_{l,j}$ is added after the last element of $h(Item_j)$. Hence, $\gamma_{l+1,j} = \gamma_{l+2,j} = \dots = \gamma_{l,j} = \gamma_{l,j}$.

4.4. Similarity Calculation of Products

Based on the above methods, we can guarantee that all the products in the system have the same number of ratings, from the perspective of hesitant fuzzy set, indicating that the amount of membership in each hesitant fuzzy set is the same. Furthermore, we can calculate the degree of similarity of two products by determining the degree of similarity of two hesitant fuzzy sets.

The common similarity calculation methods are Cosine similarity, Pearson’s correlation coefficient, and the Jaccard coefficient. Here, according to the similarity formula of hesitant fuzzy sets, which is proposed by Xu [31], we can obtain the similarity between two products, $h(Item_j)$ and $h(Item_k)$, in the recommendation system:

$$S(h(Item_j), h(Item_k)) = \frac{\sum_{i=1}^l \gamma_{ij} \cdot \gamma_{ik}}{\left(\sum_{i=1}^l (\gamma_{ij})^2 \cdot \sum_{i=1}^l (\gamma_{ik})^2\right)^{1/2}}, \tag{5}$$

4.5. Algorithm Implementation of Product Recommendation

On the basis of the previous known product similarity, we can make a recommendation based on the product. That is to say, for users with a high score of product $Item_j$, we can recommend the first few products with high similarity to product $Item_j$, so as to realize the recommendation of $User_i$.

Without available product similarity, we can make a recommendation based on users. The similarity of $User_i$ and $User_x$ can be calculated by the following formula:

$$S(User_i, User_x) = \frac{Item_{(User_i \cap User_x)}}{Item_{All}}, \tag{6}$$

where $Item_{(User_i \cap User_x)}$ represents the quantity of the same product purchased by $User_i$ and $User_x$, and $Item_{All}$ represents the quantity of all products. Obviously, the larger the $S(User_i, User_x)$, the higher the similarity between $User_i$ and $User_x$. We can recommend the first few products that similar users have purchased and scored higher to $User_i$, so as to realize the recommendation for $User_i$.

5. Case Application

This paper selected the ml-latest-small dataset from Movielens, which contains 100,836 ratings of 9742 movies by 610 users. Considering the complexity and repetition of the computational process, this paper extracted 10 movies randomly whose MovieID = 260, 293, 316, 349, 457, 527, 661, 736, 1222, and 2502 from 9742 movies of “ratings.dat” in the ml-latest-small dataset. The rating data of each movie were obtained, as shown in Table 2.

As Movielens applies the five-point evaluation rules, we transformed the score value into membership according to $\gamma_{ij} = \frac{R_{ij}}{5}, \gamma_{ij} \in [0, 1]$, as shown in Table 3.

Table 2. Movie rating sheet.

Movie ID	260	293	316	349	457	527	647	736	1222	2502
User ₁	5.0		3.0	4.0	5.0	5.0		3.0	5.0	5.0
User ₆		3.0	5.0	5.0	5.0	3.0	4.0	5.0		
User ₁₇	5.0	3.5				4.5			4.5	4.5
User ₂₈	4.0	4.5	4.0	3.0	3.0	3.0	2.5	2.5	4.0	2.0
User ₄₂	5.0		2.0	4.0	4.0	4.0	2.0		5.0	5.0
User ₅₇	5.0		3.0	4.0	5.0	5.0		2.0	4.0	
User ₆₄	3.5	4.0	3.0			5.0	3.5	3.5	4.0	4.5
User ₆₈	5.0	4.0	3.5		4.5	4.0	3.5	4.0	2.5	5.0
User ₈₄	4.0			4.0	4.0	5.0	4.0	3.0		
User ₉₁	4.5	4.0	4.0	3.0	4.0			2.5		4.5

Table 3. The membership of the products score.

Movie ID	260	293	316	349	457	527	647	736	1222	2502
User ₁	1		0.6	0.8	1	1		0.6	1	1
User ₆		0.6	1	1	1	0.6	0.8	1		
User ₁₇	1	0.7				0.9			0.9	0.9
User ₂₈	0.8	0.9	0.8	0.6	0.6	0.6	0.5	0.5	0.8	0.4
User ₄₂	1		0.4	0.8	0.8	0.8	0.4		1	1
User ₅₇	1		0.6	0.8	1	1		0.4	0.8	
User ₆₄	0.7	0.8	0.6			1	0.7	0.7	0.8	0.9
User ₆₈	1	0.8	0.7		0.9	0.8	0.7	0.8	0.5	1
User ₈₄	0.8			0.8	0.8	1	0.8	0.6		
User ₉₁	0.9	0.8	0.8	0.6	0.8			0.5		0.9

We can obtain the data description of each product and the number of scores by further applying the hesitant fuzzy set:

$$\begin{aligned}
 h(Item_{260}) &= \{1, 1, 0.8, 1, 1, 0.7, 1, 0.8, 0.9\}, l_{260} = 9 \\
 h(Item_{293}) &= \{0.6, 0.7, 0.9, 0.8, 0.8, 0.8\}, l_{293} = 6 \\
 h(Item_{316}) &= \{0.6, 1, 0.8, 0.4, 0.6, 0.6, 0.7, 0.8\}, l_{316} = 8 \\
 h(Item_{349}) &= \{0.8, 1, 0.6, 0.8, 0.8, 0.8, 0.6\}, l_{349} = 7 \\
 h(Item_{457}) &= \{1, 1, 0.6, 0.8, 1, 0.9, 0.8, 0.8\}, l_{457} = 8 \\
 h(Item_{527}) &= \{1, 0.6, 0.9, 0.6, 0.8, 1, 1, 0.8, 1\}, l_{527} = 9 \\
 h(Item_{647}) &= \{0.8, 0.5, 0.4, 0.7, 0.7, 0.8\}, l_{647} = 6 \\
 h(Item_{736}) &= \{0.6, 1, 0.5, 0.4, 0.7, 0.8, 0.6, 0.5\}, l_{736} = 8 \\
 h(Item_{1222}) &= \{1, 0.9, 0.8, 1, 0.8, 0.8, 0.5\}, l_{1222} = 7 \\
 h(Item_{2502}) &= \{1, 0.9, 0.4, 1, 0.9, 1, 0.9\}, l_{2502} = 7
 \end{aligned}$$

Obviously, $l = \max\{l_1, l_2, \dots, l_j, l_k, \dots\} = 9$.

We selected one active user, $User_{58}$, as the object to be evaluated from Movielens in order to implement recommendations. The user scored a total of 112 movies, and the average score of these movies was 3.90 points. Thus, we can consider the user as pessimistic and we can obtain the hesitant fuzzy set that increases the data after sorting in ascending order.

$$\begin{aligned}
 h(Item_{260}) &= \{0.7, 0.8, 0.8, 0.9, 1.0, 1.0, 1.0, 1.0, 1.0\} \\
 h(Item_{293}) &= \{0.6, 0.6, 0.6, 0.6, 0.7, 0.8, 0.8, 0.8, 0.9\} \\
 h(Item_{316}) &= \{0.4, 0.4, 0.6, 0.6, 0.6, 0.7, 0.8, 0.8, 1.0\} \\
 h(Item_{349}) &= \{0.6, 0.6, 0.6, 0.6, 0.8, 0.8, 0.8, 0.8, 1.0\} \\
 h(Item_{457}) &= \{0.6, 0.6, 0.8, 0.8, 0.8, 0.9, 1.0, 1.0, 1.0\} \\
 h(Item_{527}) &= \{0.6, 0.6, 0.8, 0.8, 0.9, 1.0, 1.0, 1.0, 1.0\} \\
 h(Item_{647}) &= \{0.4, 0.4, 0.4, 0.4, 0.5, 0.7, 0.7, 0.8, 0.8\} \\
 h(Item_{736}) &= \{0.4, 0.4, 0.5, 0.5, 0.6, 0.6, 0.7, 0.8, 1.0\} \\
 h(Item_{1222}) &= \{0.5, 0.5, 0.5, 0.8, 0.8, 0.8, 0.9, 1.0, 1.0\} \\
 h(Item_{2502}) &= \{0.4, 0.4, 0.4, 0.9, 0.9, 0.9, 1.0, 1.0, 1.0\}
 \end{aligned}$$

Applying the similarity Equation (5), we can determine the similarity between any two products.

$$\begin{aligned}
 &S(h(Item_{260}), h(Item_{293})) \\
 &= \frac{0.7 \times 0.6 + 0.8 \times 0.6 \times 2 + 0.9 \times 0.6 + 1.0 \times 0.7 + 1.0 \times 0.8 \times 3 + 1.0 \times 0.9}{\sqrt{(0.7^2 + 0.8^2 \times 2 + 0.9^2 + 1.0^2 \times 5) \cdot (0.6^2 \times 4 + 0.7^2 + 0.8^2 \times 3 + 0.9^2)}} \\
 &\approx 0.996080
 \end{aligned}$$

In the same way, the similarity between other movies can be calculated, as shown in Table 4.

Table 4. The similarity between products.

Movie ID	260	293	316	349	457	527	647	736	1222	2502
260	1	0.996080	0.981830	0.994104	0.995393	0.996563	0.978223	0.799065	0.988359	0.975629
293		1	0.989067	0.998548	0.995909	0.995856	0.989216	0.986584	0.989434	0.973720
316			1	0.989887	0.993023	0.990814	0.990538	0.996927	0.999930	0.980996
349				1	0.993149	0.993961	0.987747	0.989801	0.989588	0.974983
457					1	0.998860	0.987245	0.986571	0.992389	0.979900
527						1	0.986705	0.984144	0.991993	0.953648
647							1	0.991155	0.989228	0.978055
736								1	0.990056	0.976978
1222									1	0.994563
2502										1

*Item*₃₁₆ and *Item*₁₂₂₂ have the highest similarity in Table 3. After the data addition process, it can be considered that there is a certain degree of substitution between products. Moreover, there is a high similarity between *Item*₄₅₇ and *Item*₅₂₇, as well as *Item*₂₉₃ and *Item*₃₄₉. *Item*₂₆₀ and *Item*₇₃₆ have the lowest similarity and the worst substitution.

From the Movielens dataset, it can be seen that the five-point movies watched by *User*₅₈ are Movie ID = 293, 457, and 527. The top five movies similar to these movies that have not been watched by *User*₅₈ as recommended movies: Movie ID = 316, 1222, and 260.

6. Algorithm Verification

To verify the effectiveness of the proposed algorithm, we used user-based recommendation to recommend movies for *User*₅₈. From the selected part of the Movielens score table, we know *Item*_{All} = 10. According to Equation (6), we can calculate the similarity between *User*₅₈ and *User*_{*i*} (*i* = 1, 6, 17, 28, 42, 57, 64, 68, 84, 91), as shown in the following Table 5.

Table 5. The similarity between *User*₅₈ and the other users.

User _{<i>i</i>}	1	6	17	28	42	57	64	68	84	91
User ₅₈	0.4	0.5	0.2	0.5	0.3	0.4	0.4	0.4	0.4	0.4

It is concluded that the largest number of users who have seen the same movie with $User_{58}$ are $User_6$ and $User_{28}$, then found out the movies that $User_6$ and $User_8$ scored higher and that $User_{58}$ did not watch: Movie ID = 316, 1222, and 647.

It can be seen from the results based on the user and product recommendations that the method proposed in this paper is highly effective.

In order to verify the practical effect of the recommendation algorithm based on hesitant fuzzy sets, we randomly extracted ten groups of data from Movielens to express the calculation process. The verification of other data can be obtained through similar calculations. The calculation of all data can be obtained by computer, but the data conversion process cannot be seen in the programming process, so this paper only used part of the data calculation.

7. Conclusions and Prospect

The steps of the method proposed in this paper are as follows: First, the rating matrix of the data for processing is obtained in the recommendation system, which is often a sparse matrix. Then, the sparse matrix in the recommendation system is supplemented by the forward or backward strategies. Consequently, the similarity between the products is calculated using the supplemented sparse matrix and the idea of hesitant fuzzy sets, while the product-based recommendation is obtained according to the similarity. Finally, the algorithm is verified based on user recommendation, and the results indicate that the proposed method is very effective. The main innovations of this paper include two aspects. On the one hand, it proposes making up the sparse matrix through the forward or the backward strategies. It ensures that the similarity between the products is obtained without losing the rating information of the sparse matrix. The effective information of the rating matrix is maximized, thus providing a new approach for the sparse matrix processing and similarity discussion in the recommendation system. On the other hand, beginning from the inherent uncertainty in the recommendation system, the hesitant fuzzy set can solve the recommendation quality problem with the help of the processing tool of uncertain problems, which undoubtedly finds a new field for the study of hesitant fuzzy set.

Nevertheless, there are still some problems in the research of this paper. First, this paper attempted to use hesitant fuzzy set theory to solve the complex problems in the recommendation system. Although there was a good docking between the two, the recommendation quality was compared only to a simple user-based recommendation, and the results may be inaccurate. Determining other more complex and accurate methods for comparison is one of the focuses of the authors' subsequent work. Second, because the calculation of the whole dataset needs to be realized by computer programs, and this process cannot describe in detail how to use hesitant fuzzy sets for data conversion, this paper no longer shows the program code.

Author Contributions: Formal analysis, C.C.; methodology, Z.Z.; writing—original draft preparation, J.L. and C.C.; writing—review and editing, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This study uses the Movielens dataset, <http://files.grouplens.org/datasets/movielens/> (accessed on 10 October 2021).

Acknowledgments: This study was supported by the 2020 Henan University Philosophy and Social Sciences Applied Research Major Project Plan (no. 2020-YYZD-02), the Humanities and Social Science Research General Project of Henan Provincial Department of Education in 2021 (no. 2021-ZZJH-020), the 2020 Henan Province Philosophy and Social Science Planning Project (no. 2020BJJ041), and the 2021 Key Scientific Research Projects of Colleges and Universities in Henan Province (no. 21A520021).

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

References

1. Kim, Y.S. Text recommender system using user's usage patterns. *Ind. Manag. Data Syst.* **2011**, *111*, 282–297. [[CrossRef](#)]
2. Bobadilla, J.; Ortega, F.; Hernando, A.; Gutiérrez, A. Recommender systems survey. *Knowl.-Based Syst.* **2013**, *46*, 109–132. [[CrossRef](#)]

3. Zenebe, A.; Norcio, A.F. Representation, similarity measures and aggregation methods using fuzzy sets for content-based recommender systems. *Fuzzy Sets Syst.* **2009**, *160*, 76–94. [[CrossRef](#)]
4. Pérez-Almaguer, Y.; Yera, R.; Alzahrani, A.A.; Martínez, L. Content-based group recommender systems: A general taxonomy and further improvements. *Expert Syst. Appl.* **2021**, *184*, 115444. [[CrossRef](#)]
5. Shen, J.; Zhou, T.; Chen, L. Collaborative filtering-based recommendation system for big data. *Int. J. Comput. Sci. Eng.* **2020**, *21*, 219–225. [[CrossRef](#)]
6. Tarus, J.K.; Niu, Z.; Mustafa, G. Knowledge-based recommendation: A review of ontology-based recommender systems for e-learning. *Artif. Intell. Rev.* **2018**, *50*, 21–48. [[CrossRef](#)]
7. Li, H.; Li, H.; Zhang, Z.; Cheng, J. Intelligent learning system based on personalized recommendation technology. *Neural Comput. Appl.* **2019**, *31*, 4455–4462. [[CrossRef](#)]
8. Zhang, H.-R.; Min, F.; He, X.; Xu, Y.-Y. A hybrid recommender system based on user-recommender interaction. *Math. Probl. Eng.* **2015**, *2015*, 145636. [[CrossRef](#)]
9. Li, L.; Zhang, Z.; Zhang, S. Hybrid Algorithm Based on Content and Collaborative Filtering in Recommendation System Optimization and Simulation. *Sci. Program.* **2021**, *2021*, 7427409.
10. Zhang, W.; Zhang, X.; Wang, H.; Chen, D. A deep variational matrix factorization method for recommendation on large scale sparse dataset. *Neurocomputing* **2019**, *334*, 206–218. [[CrossRef](#)]
11. Wang, J.; Zhu, L.; Dai, T.; Xu, Q.; Gao, T. Low-rank and sparse matrix factorization with prior relations for recommender systems. *Appl. Intell.* **2021**, *51*, 3435–3449. [[CrossRef](#)]
12. Liu, D.; Lu, Y.X.; Zhang, Y.S.; Guo, L.-Y. A new item recommend algorithm of sparse data set based on user behavior analyzing. In Proceedings of the 12th International Conference on Signal Processing (ICSP), Hangzhou, China, 19–23 October 2014; pp. 1377–1380.
13. Huang, L.; Zhao, Z.L.; Wang, C.D.; Huang, D.; Chao, H.Y. Lscd: Low-rank and sparse cross-domain recommendation. *Neurocomputing* **2019**, *366*, 86–96. [[CrossRef](#)]
14. Vozalis, E.; Margaritis, K.G. Analysis of recommender systems' algorithms. In Proceedings of the 6th Hellenic European Research Conference on Computer Mathematics and its Application, Athens, Greece, 25–27 September 2003.
15. Chen, Z.S.; Chin, K.S.; Li, Y.L.; Yang, Y. Proportional hesitant fuzzy linguistic term set for multiple criteria group decision making. *Inf. Sci. Int. J.* **2016**, *357*, 61–87. [[CrossRef](#)]
16. Linden, G.; Smith, B.; York, J. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Intemct. Comput.* **2003**, *7*, 76–80. [[CrossRef](#)]
17. Guo, Y.; Yin, C.; Li, M.; Ren, X.; Liu, P. Mobile e-commerce recommendation system based on multi-source information fusion for sustainable e-business. *Sustainability* **2018**, *10*, 147. [[CrossRef](#)]
18. Wei, J.; He, J.; Chen, K.; Zhou, Y.; Tang, Z. Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Syst. Appl.* **2017**, *69*, 29–39. [[CrossRef](#)]
19. Rodríguez, R.M.; Martínez, L.; Herrera, F.; Torra, V. A review of hesitant fuzzy sets: Quantitative and qualitative extensions. In *Fuzzy Logic in Its 50th Year*; Springer: New York, NY, USA, 2016; Volume 341, pp. 109–128.
20. Rodríguez, R.M.; Bedregal, B.; Bustince, H.; Dong, Y.C.; Farhadinia, B.; Kahraman, C.; Martínez, L.; Torra, V.; Xu, Y.J.; Xu, Z.S.; et al. A position and perspective analysis of hesitant fuzzy sets on information fusion in decision making. Towards high quality progress. *Inf. Fusion* **2016**, *29*, 89–97. [[CrossRef](#)]
21. Rodriguez, R.M.; Martinez, L.; Herrera, F. Hesitant fuzzy linguistic term sets for decision making. *IEEE Trans. Fuzzy Syst.* **2011**, *20*, 109–119. [[CrossRef](#)]
22. Mardani, A.; Saraji, M.K.; Mishra, A.R.; Rani, P. A novel extended approach under hesitant fuzzy sets to design a framework for assessing the key challenges of digital health interventions adoption during the COVID-19 outbreak. *Appl. Soft Comput.* **2020**, *96*, 1–42. [[CrossRef](#)]
23. Çolak, M.; Kaya, İ. Multi-criteria evaluation of energy storage technologies based on hesitant fuzzy information: A case study for Turkey. *J. Energy Storage* **2020**, *28*, 1–14. [[CrossRef](#)]
24. Sahu, K.; Alzahrani, F.A.; Srivastava, R.K.; Kumar, R. Hesitant fuzzy sets based symmetrical model of decision-making for estimating the durability of Web application. *Symmetry* **2020**, *12*, 1770. [[CrossRef](#)]
25. Rani, P.; Mishra, A.R.; Krishankumar, R.; Mardani, A.; Cavallaro, F.; Soundarapandian Ravichandran, K.; Balasubramanian, K. Hesitant fuzzy SWARA-complex proportional assessment approach for sustainable supplier selection (HF-SWARA-COPRAS). *Symmetry* **2020**, *12*, 1152. [[CrossRef](#)]
26. Yera, R.; Martinez, L. Fuzzy tools in recommender systems: A survey. *Int. J. Comput. Intell. Syst.* **2017**, *10*, 776–803. [[CrossRef](#)]
27. Castro, J.; Barranco, M.J.; Rodríguez, R.M.; Martínez, L. Group recommendations based on hesitant fuzzy sets. *Int. J. Intell. Syst.* **2018**, *33*, 2058–2077. [[CrossRef](#)]
28. Chen, N.; Xu, Z.; Xia, M. Correlation coefficients of hesitant fuzzy sets and their applications to clustering analysis. *Appl. Math. Model.* **2013**, *37*, 2197–2211. [[CrossRef](#)]
29. Torra, V. Hesitant fuzzy sets. *Int. J. Intell. Syst.* **2010**, *25*, 529–539. [[CrossRef](#)]
30. Xu, Z.; Xia, M. Induced generalized intuitionistic fuzzy operators. *Knowl.-Based Syst.* **2011**, *24*, 197–209. [[CrossRef](#)]
31. Xu, Z.; Cai, X. Uncertain power average operators for aggregating interval fuzzy preference relations. *Group Decis. Negotiat.* **2011**, *21*, 381–397. [[CrossRef](#)]

-
32. Xu, Z.; Xia, M. Distance and similarity measures for hesitant fuzzy sets. *Inf. Sci.* **2011**, *181*, 2128. [[CrossRef](#)]
 33. Xu, Z.; Xia, M. On distance and correlation measures for hesitant fuzzy sets. *Int. J. Intell. Syst.* **2011**, *26*, 410–425. [[CrossRef](#)]