applied sciences

MDPI

Article

RSII: A Recommendation Algorithm That Simulates the Generation of Target Review Semantics and Fuses ID Information

Qiulin Ren [1,2], Jiwei Qin [1,2,*], Jianjie Shao [1,2] and Xiaoyuan Song [1,2]

[1] School of Information Science and Engineering, Xinjiang University, Urumqi 830046, China
[2] Key Laboratory of Signal Detection and Processing, Xinjiang Uygur Autonomous Region, Xinjiang University, Urumqi 830046, China
* Correspondence: jwqin_xju@163.com

Abstract: The target review has been proven to be able to predict the target user's rating of the target item. However, in practice, it is difficult to obtain the target review promptly. In addition, the target review and the rating may sometimes be inconsistent (such as preference reviews and low ratings). There is currently a lack of research on the above issues. Therefore, this paper proposed a Recommendation algorithm that Simulates the generation of target review semantics and fuses the ID Information (RSII). Specifically, based on the characteristics of the target review available during the model training, this paper designed a teacher module and a review semantics learning module. The teacher module learned the semantics of the target review and guided the review semantics learning model to learn these semantics. Then, this study used the fusion module to dynamically fuse the target review semantics and the ID information, enriching the representation of predictive features, thereby, alleviating the problem of inconsistency between the target review and the rating. Finally, the RSII model was extensively tested on three public datasets. The results showed that compared with seven of the latest and most advanced models, the RSII model improved the MSE metric by 8.81% and the MAE metric by 10.29%.

Keywords: recommendation system; deep learning; review; representation learning; rate prediction

check for updates

Citation: Ren, Q.; Qin, J.; Shao, J.; Song, X. RSII: A Recommendation Algorithm That Simulates the Generation of Target Review Semantics and Fuses ID Information. *Appl. Sci.* **2023**, *13*, 3942. https://doi.org/10.3390/app13063942

Academic Editor: Yu-Dong Zhang

Received: 3 March 2023
Revised: 16 March 2023
Accepted: 17 March 2023
Published: 20 March 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the rapid development of Internet technology, the Internet has brought people much information while also causing the problem of information overload. This makes it difficult for people to quickly and accurately select the parts they are interested in from massive amounts of information. Collaborative filtering (CF) [1–3] technology is one of the methods used to solve this problem. However, CF cannot accurately predict the user ratings under data sparsity. To alleviate this problem, researchers use reviews to predict user ratings. For example, Zhang et al. [4] proposed the DeepCoNN model, as shown in Figure 1. First, CNN [5,6] is used to learn the user and item representations from reviews, then FM [7,8] is used to predict the ratings, resulting in excellent performance and triggering a research boom in using reviews to predict ratings.

The performance of the DeepCoNN model depends on the target reviews. The target review is the review that the target user writes on the target item when the recommendation system predicts the target user's rating for the target item. However, in real life, there are two issues with target reviews, as shown in Figure 2. The two problems are: (1) inconsistency between the target review and the rating; (2) difficulty in obtaining the target review promptly.

In terms of the difficulty in obtaining target reviews, Rose et al. [9] pointed out that target reviews should not be present during prediction, but they can be used for model training and learning. They proposed the TransNets [9] model, which consists of a source network and a target network. The source network learns to predict features from the

user and item review documents. The target network takes target reviews as input, first learning the representation of predictive features, and then guiding the source network to learn predictive features. This allows the TransNets model to avoid using target reviews during the prediction process. Although the TransNets model alleviates the problem of the difficulty in obtaining target reviews promptly, the simple structure of the target network and the complexity of the task make it difficult to effectively guide the learning of the source network. In addition, the inconsistency between the target reviews and the ratings is not considered.

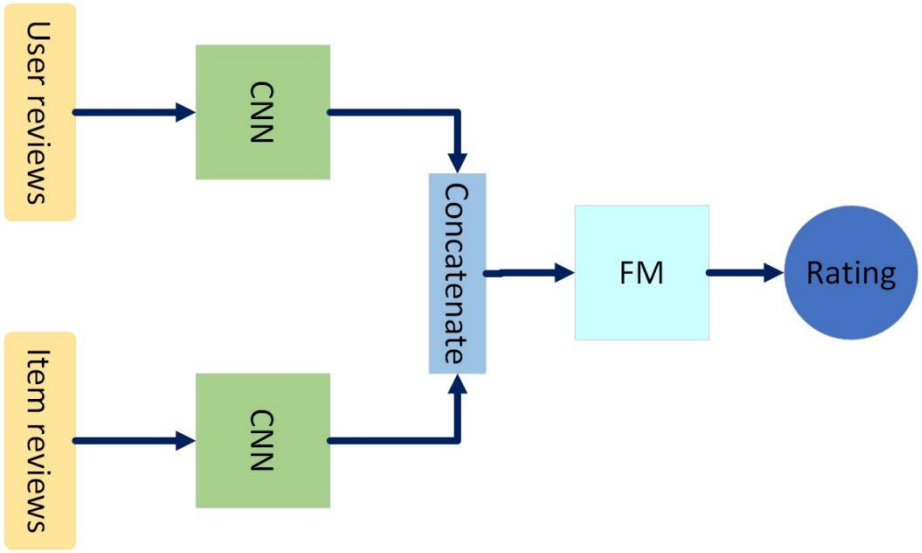


Figure 1. The architecture of the DeepCoNN model.

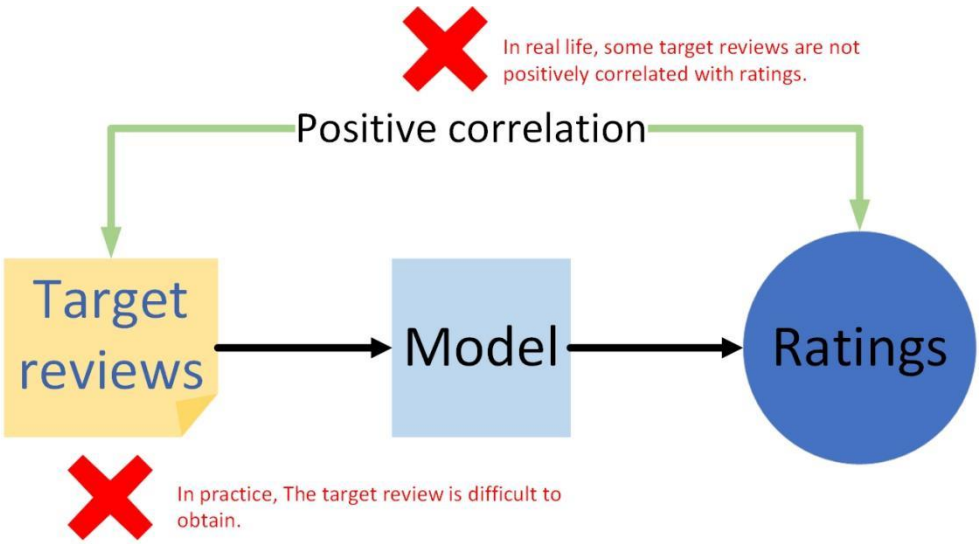


Figure 2. The problems encountered in practical life by recommendation models based on reviews.

In terms of inconsistent target reviews and ratings, Seo et al. proposed the D-Attn [10] model based on the DeepCoNN, which introduced local and global attention mechanisms to find more reviews and words related to the rating in user (item) reviews and increase their corresponding weights, thereby, improving the recommendation performance. Chen et al. subsequently proposed the NARRE [11] model based on the DeepCoNN, which introduced a review attention mechanism to remove reviews unrelated to the predicted rating and improve the recommendation performance. Subsequently, Tay et al. proposed the MPCN [12] model, which selected several relevant reviews first through a multi-pointer scheme and a joint attention mechanism, and then constructed a user–item matrix. They retained the most

relevant part of the rating through a hierarchical filtering approach, thereby improving the recommendation performance. Shao et al. proposed the MPCAR [13] model based on MPCN to suppress the selection of irrelevant reviews for the target item by introducing ID information to enrich the feature representation. However, they ignored the difficulty of obtaining target reviews and either linearly fused the ID information or concatenated multiple vectors to fuse the ID information, making it difficult to dynamically balance the information between the target reviews and the ID.

In order to solve the above problem, this paper proposed a Recommendation algorithm that Simulates the generation of target review semantics and fuses ID Information (RSII). Specifically, based on the characteristics of the target review available during the model training, this paper designed a teacher module and a review semantics learning module. The teacher module learned the semantics of the target review and guided the review semantics learning model to learn these semantics. Then, this study used the fusion module to dynamically fuse the target review semantics and the ID information, enriching the representation of the predictive features, thereby, alleviating the problem of inconsistency between the target review and the rating. The main contributions of this paper are as follows:

1. This paper presents a review semantic learning module and a teacher module, where the teacher module could effectively guide the review semantic learning module to learn the target review semantics, thus, alleviating the problem of difficulty in obtaining the target reviews promptly.
2. This paper proposes a fusion module that dynamically combines the ID information and the target review semantics to enrich the expression of predictive features, thereby alleviating the problem of inconsistency between the target reviews and ratings.
3. The RSII model was extensively tested on three public datasets. The results showed that compared with seven state-of-the-art models, the RSII model improved the MSE metric by 8.81% and the MAE metric by 10.29%.

2. Related Work

In this section, the paper reviews the work related to this study, including self-supervised learning and review-based recommendation algorithms.

2.1. Review-Based Recommendations

Due to the sparsity of data, collaborative filtering algorithms have difficulty providing accurate recommendation services. Researchers have proposed some methods to alleviate the problem of data sparsity [14–16], and recommendation algorithms based on reviews are one of them. The early work [17,18] typically incorporated review information using topic modeling. However, this method cannot effectively capture the deep features in review text.

Recently, with the rapid development of deep learning, an increasing number of researchers are introducing deep learning to process review information and provide personalized recommendations to users [19–21]. For example, Zheng et al. proposed the DeepCoNN model, which first uses two CNNs to extract the user and item review features from the user and item review documents, respectively. The user and item review features are stitched into a vector, and finally, the rating prediction is performed using FM. Rose et al. pointed out that obtaining the target reviews in practice is difficult, and the DeepCoNN model cannot solve this problem. Therefore, they proposed the TransNets model. The TransNets model consists of a source network and a target network. The source network learns to predict the features from the user and item review documents. The target network takes the target reviews as input, first learning the representation of predictive features, and then guiding the source network to learn predictive features. Seo et al. proposed the D-Attn model based on the DeepCoNN, which introduced local and global attention mechanisms to find more reviews and words related to the rating in user (item) reviews and increase their corresponding weights, thereby, improving the

recommendation performance. Chen et al. subsequently proposed the NARRE model based on the DeepCoNN, which introduced a review attention mechanism to remove reviews unrelated to the predicted rating and improve the recommendation performance. Since attention mechanisms are implemented differently, Wu et al. proposed the CARL [22] model, which learns the critical information of a review in the form of an attention matrix and fuses the interaction features by linear interpolation. Tay et al. proposed the MPCN model, which selected several relevant reviews first through a multi-pointer scheme and a joint attention mechanism, and then constructed a user–item matrix. They retained the most relevant part of the rating through a hierarchical filtering approach, thereby, improving the recommendation performance. Shao et al. proposed the MPCAR model based on MPCN to suppress the selection of irrelevant reviews for the target item by introducing the ID information to enrich the feature representation. However, the aforementioned studies focused on one of the two issues in Figure 2 and did not comprehensively investigate both issues.

2.2. Self-Teacher Learning

Self-teacher learning [23,24] is a novel machine learning paradigm that learns data representation from unlabeled data by setting up pre-training tasks. This approach splits the input data into multiple task views, thus enabling the model to learn a complete representation. It was first applied in images, where researchers created auxiliary supervision signals by rotating, cropping, and recoloring the images to enhance the model's learning expression. Subsequently, significant progress has been made in computer vision, audio processing, and natural language understanding. There have already been works combining self-teacher learning with recommendation algorithms. For example, Zhou et al. [25] used the correlation in the context information as the self-teacher signal in sequence recommendation to maximize the mutual information between the attribute, item, and sequence views. CCDR [26] designed a domain-specific contrastive learning and three domain-cross contrastive learning tasks to better perform representation learning. Unlike the aforementioned methods, this paper applied self-supervised learning to review-based recommendation algorithms. The reviews in the data were divided into target reviews and available reviews, and the mutual information between the target reviews and the available reviews was maximized. This enabled the teacher module to guide the semantic learning of the review module towards the target reviews, thereby enhancing the generation ability of the review module.

3. Preliminaries

In this section, we will provide a detailed introduction to the commonly used symbols and word embeddings in this paper.

3.1. Notation

This paper defines a user set $U = \{u_1, \cdots, u_i\}$, an item set $V = \{v_1, \cdots, v_j\}$, a review set $C = \{c_{1,1}, \cdots, c_{i,j}, \cdots, c_{n,m}\}$ and a rating set $R = \{r_{1,1}, \cdots, r_{i,j}, \cdots, r_{n,m}\}$, u_i, v_j, $c_{i,j}$ and $r_{i,j}$ denote the i-th user, the j-th item, the i-th user's review of the j-th item and the i-th user's rating of the j-th item. Let $\hat{r}_{i,j}$ be the predicted rating of the i-th user for the j-th item. n represents the number of users, and m represents the number of items. d represents the embedding size. In addition, we defined some key symbols used in the intermediate process of the model, as shown in Table 1.

Table 1. Notations and definitions.

Notations	Definitions
$r_{i,j}$	The i-th user's rating of the j-th item
$c_{i,j}$	The i-th user's review of the j-th item
$\hat{r}_{i,j}$	The predicted rating of the i-th user for the j-th item

Table 1. *Cont.*

Notations	Definitions
d	The embedding size
$c^1_{i,j}$	The quantized value of $c_{i,j}$
$c^2_{i,j}$	The semantics of $c^1_{i,j}$
$c^3_{i,j}$	The reconstruction vector of $c^1_{i,j}$
C^j_i	All reviews of the i-th user excluding $c_{i,j}$
C^j_I	All reviews on the j-th item excluding $c_{i,j}$
$\hat{c}_{i,j}$	The generated semantic features of $c_{i,j}$ reviews
$O_{i,j}$	The ID features of the i-th user and j-th item
$a_{i,j}$	The relative coefficient between $O_{i,j}$ and $\hat{c}_{i,j}$

3.2. Word Embedding

Recently, in NLP, preprocessing methods have been widely used. The most famous methods are word2vec and GloVe [27]. GloVe is a type of word embedding model that represents words as vectors in a high-dimensional space. It was developed by Stanford University researchers in 2014 and is based on the idea that words that appear in similar contexts tend to have similar meanings. GloVe uses a co-occurrence matrix to capture the statistical relationships between words. The matrix is constructed by counting the number of times each word appears in the context of every other word in a large corpus of text. The resulting matrix is then factorized using matrix factorization techniques to obtain a lower-dimensional representation of the words. In this paper, we chose glove.6B.50d. It represents mapping each word to a 50-dimensional space according to the rules of GloVe.

4. Proposed RSII Model

Figure 3 shows the overall architecture of the RSII, which consisted of three modules and one predictor. The details of each module will be explained in the following sections.

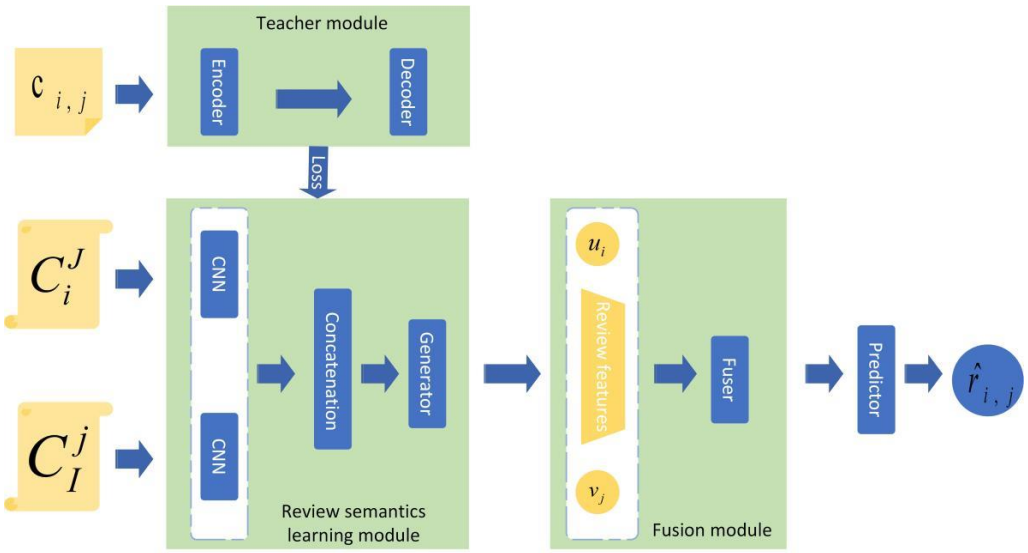


Figure 3. The framework of the RSII model.

4.1. The Teacher Module

In this section, we will discuss the working process and significance of the teacher module in detail. As is well known, reviews are real evaluations written by users themselves, expressing their true thoughts at a certain moment. Additionally, reviews have the following characteristics: (1) Subjectivity: reviews are the personal opinions and evaluations of users on a certain thing or event; thus, they are subjective. (2) Evaluative: reviews

usually contain evaluations on a certain thing or event, including its quality, advantages, disadvantages, etc. Therefore, it can be seen that target reviews are important for predicting the ratings between the target users and the target items. However, in practice, it is difficult to obtain the target reviews promptly. Therefore, it is necessary to generate target reviews indirectly for use in the recommendation process. The teacher module designed in this paper, combined with data division, could solve this problem.

In this paper, based on the data processing process of the TransNets model, the data was divided as follows: when the model predicted the rating of the i-th user on the j-th item, $c_{i,j}$ was a target review, C_i^J was all reviews of the i-th user excluding $c_{i,j}$, and C_I^j was all reviews on the j-th item excluding $c_{i,j}$. Although the target at test time, $c_{i,j}$ was available during training. As such, the teacher module was only used in model training stage.

$c_{i,j}$ first obtained quantized value $c_{i,j}^1$ after word embedding. $c_{i,j}^1$ was then fed into the encoder to get $c_{i,j}^2$. $c_{i,j}^2$ represents the semantics of $c_{i,j}^1$.

$$c_{i,j}^1 = G(c_{i,j}) \tag{1}$$

$$c_{i,j}^2 = CNN(c_{i,j}^1) \tag{2}$$

where G denotes word embedding and CNN denotes CNN text processing. Currently, there are widely popular methods in mining semantics, such as TextCNN [28], TextRNN, paragraph vector, and BERT [29]. With the current research on review-based recommendation algorithms, this paper took the mainstream TextCNN processing approach. This article took a vector P of review length Q as an example to elaborate the CNN used in this article. CNN is composed of convolutional, pooling, and fully connected layers. The convolutional layer consists of K neurons that generate new features by applying convolutional operators to P. The j-th convolution kernel $K_j \in \mathbb{R}^{d \times q}$ in the convolutional layer performs convolutional operation on a window of size q, as shown in the following formula.

$$p_j = actf\left(P * K_j + b_j\right) \tag{3}$$

where, $*$ represents the convolution operation, b_j is the bias term, and $actf$ is the activation function, which uses Rectified Linear Units. The formula for Rectified Linear Units as the activation function is as follows:

$$f(x) = max\{0, x\} \tag{4}$$

Then, the max-pooling operation was performed on the feature map, and this article used the maximum value as the corresponding feature h_j of the convolution kernel. The formula is as follows:

$$h_j = max\left\{p_1, p_2, \cdots, p_{Q-q+1}\right\} \tag{5}$$

Equations (3)–(5) describe the operation of a single convolution kernel. The CNN used in this article uses multiple convolution kernels, and its output is shown in Equation (6):

$$H = \{h_1, h_2, \cdots, p_{n_1}\} \tag{6}$$

Finally, the output of the max-pooling layer was connected to the fully connected layer to obtain the feature X. The formula is as follows:

$$X = f_{tanh}(H) \tag{7}$$

where, f_{tanh} represents the fully connected layer with a tanh activation function.

The most commonly used CNN hyperparameter selection in the current majority of review-based recommendation algorithms is as follows: the number of convolutional kernels is 100, the size of the convolutional kernels is 3.

Subsequently, $c_{i,j}^2$ through the decoder to get $c_{i,j}^3$. The formulation is as follows:

$$a_0 = f(c_{i,j}^2) \tag{8}$$

$$a_1 = UpSamle(a_0) \tag{9}$$

$$a_2 = \sigma(a_1) \tag{10}$$

$$c_{i,j}^3 = Cov_T(a_2) \tag{11}$$

where f denotes the fully connected layer, input size d, the output size is the length of each review times the number of convolutional kernels of the CNN, and tanh activation function. $UpSamle$ denotes the upsampling function [30], σ the Relu activation function, and Cov_T the deconvolution function [31]. The hyperparameters of its deconvolution function are the opposite of CNN.

In order to achieve a one-to-one mapping between the target review and the target review semantics, the teacher module adopted an encoding–decoding structure. Here, we minimized the error between $c_{i,j}^1$ and $c_{i,j}^3$. The formulation is as follows:

$$L_0 = \sum (c_{i,j}^1 - c_{i,j}^3)^2 \tag{12}$$

4.2. The Review Semantics Learning Module

In this subsection, we describe the structure and workflow of the review semantics learning module in detail. Due to the fact that the review semantic learning module does not include the target review in its input, the learned predictive features are difficult to effectively reflect the target user's opinion on the target item. Therefore, we urgently needed to generate the semantics of the target review in order to understand the target user's opinion on the target item. Combining the teacher and review semantic learning modules can achieve the above functions.

The review semantics learning module consisted of two CNNs and a generator. Firstly, C_i^J and C_I^j were quantified by word embeddings to get $\dot{C}_i^J$ and $\dot{C}_I^j$. Then, $\ddot{C}_i^J$ and $\ddot{C}_I^j$ were obtained by their respective CNNs. The formulation is as follows:

$$\dot{C}_i^J = G\left(C_i^J\right) \tag{13}$$

$$\dot{C}_I^j = G\left(C_I^j\right) \tag{14}$$

$$\ddot{C}_i^J = CNN\left(\dot{C}_i^J\right) \tag{15}$$

$$\ddot{C}_I^j = CNN\left(\dot{C}_I^j\right) \tag{16}$$

In order to better learn the feature representation, this paper concatenates $\ddot{C}_i$ and $\ddot{C}^j$ to form $\dot{c}_{i,j}$ through vector stitching. The formulation is as follows:

$$\dot{c}_{i,j} = \ddot{C}_i^J \oplus \ddot{C}_I^j \tag{17}$$

where $\oplus$ denotes concatenate.

Then, $\dot{c}_{i,j}$ goes through the generator to get the generated semantic features of reviews $\hat{c}_{i,j}$. The formulation is as follows:

$$h_0 = Tanh(W_0\dot{c}_{i,j} + b) \tag{18}$$

$$\hat{c}_{i,j} = Tanh(W_1 h_0 + b) \tag{19}$$

where W_0 and W_1 separately denote the weight matrix of the first fully connected layer and the second fully connected layer, b denotes the bias vector of the fully connected layer, *Tanh* denotes the tanh activation function. We set the input size of the first layer to be $2d$ and the output size to be d. We set the input size of the second layer to be d and the output size to be d.

In this paper, the teacher module was used to supervise the learning of the review semantics learning module. The error between $\hat{c}_{i,j}$ and $c^2_{i,j}$ was minimized in this paper. The formulation is as follows:

$$L_1 = \sum (\hat{c}_{i,j} - c^2_{i,j})^2 \tag{20}$$

4.3. The Fusion Module

By jointly training the teacher module and the review semantic learning module, it was possible to generate the target review semantics effectively and also reasonably alleviate the problem of difficult target review acquisition. However, some target reviews still showed low credibility, meaning that they were inconsistent with the rating. If only the features are extracted from the target reviews to predict the rating, it is difficult to achieve significant results. Therefore, this paper introduced user and item ID information to mine the missing information from the target reviews, dynamically integrating the ID information and the target review semantics to enrich the prediction features. Figure 4 shows the structure of the fusion module.

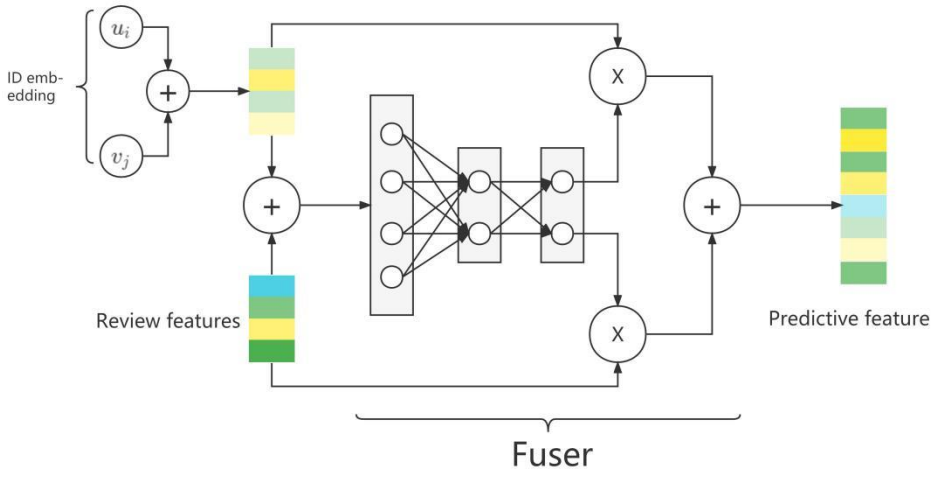


Figure 4. The structure of the fusion module.

First, this paper embedded the user and item IDs into a dense space to get u^1_i and v^1_j. In order to better capture the interaction information between the users and items, this paper mapped u^1_i and v^1_j into the same space to get the ID information $O_{i,j}$. The formulation is as follows:

$$u^1_i = W_i u_i \tag{21}$$

$$v^1_j = W_j v_j \tag{22}$$

$$O_{i,j} = u^1_i \oplus v^1_j \tag{23}$$

where W_i and W_m denote the learnable matrices, $W_i \in R^{n \times \frac{d}{2}}$, and $W_j \in R^{m \times \frac{d}{2}}$.

To compensate for the lack of review information, this article calculated the relative coefficient between the target review semantics and the ID information, and then combined them to enrich the prediction features of the target review semantics and ID information. Firstly, this paper put $O_{i,j}$ and $\hat{c}_{i,j}$ in the same space to get $\overline{T}_{i,j}$. Then put $\overline{T}_{i,j}$ into the heater to obtain the relative coefficient $a_{i,j}$ between $O_{i,j}$ and $\hat{c}_{i,j}$. Finally, it formed a brand-new

predictive feature, $T_{i,j}$. Additionally, the fuser was a neural network based on MLP. The formulation is as follows:

$$\overline{T}_{i,j} = O_{i,j} \oplus \hat{c}_{i,j} \tag{24}$$

$$a_{i,j} = soft(f_1(\sigma(f_0(\overline{T}_{i,j})))) \tag{25}$$

$$T_{i,j} = a_{i,j}[0] \times O_{i,j} \oplus a_{i,j}[1] \times \hat{c}_{i,j} \tag{26}$$

where f_0 represents a fully connected layer with inputs of size $2d$ and outputs of size 2. f_1 represents a fully connected layer with inputs of size 2 and outputs of size 2.

4.4. Predictor

In order to better extract feature information, this paper adopted the MLP [32] structure as the predictor of the model. Then, we obtained the prediction rate $\hat{r}_{i,j}$ and the regression loss function for prediction evaluation L_2. The formulation is as follows:

$$a_3 = \sigma(f_3(T_{i,j})) \tag{27}$$

$$a_4 = \sigma(f_4(a_3)) \tag{28}$$

$$\hat{r}_{i,j} = f_5(a_4) \tag{29}$$

$$L_2 = \sum (\hat{r}_{i,j} - r_{i,j})^2 \tag{30}$$

where the input size of f_3 is $2d$ and the output size is $\frac{d}{2}$. The input size of f_4 is $\frac{d}{2}$ and the output size is $\frac{d}{4}$. The input size of f_5 is $\frac{d}{4}$ and the output size is 1.

Finally, we used a step-by-step training approach with the GAN [33,34] network to train the RSII model. We first updated the teacher module using L_0, then updated the comment semantic learning module using L_1, and finally updated the review semantic learning module, fusion module, and predictor using L_2.

5. Experimental Section

5.1. Datasets and Evaluation Metric

The paper selected three publicly available 5-core datasets [35]: Digital Music, Musical Instruments, and Office Products, which are from different domains. The analysis of the datasets is shown in Table 2. DM means Digital Music, MI means Musical Instruments, and OP means Office Products. These three datasets contained honest user reviews from Amazon between May 1996 and July 2014. Each dataset included the users, the items, and the user reviews and ratings. Every user in these datasets had posted at least five reviews on the platform. In the experiments, the paper divided the datasets in a ratio of 8:1:1, where the validation set results were used to adjust the hyperparameters, and the results of the test set were used as the final experiments.

Table 2. Statistics of the datasets.

Statistics	MI	OP	DM
of users	1429	4905	5541
of items	900	2420	3568
of ratings	10,254	53,237	64,705
Sparsity	99.20%	99.51%	99.67%

As the rating prediction task s a regression task, we adopted the MSE and MAE as the evaluation metrics for the RSII model. The formulation is as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widetilde{y}_i)^2 \tag{31}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widetilde{y}_i| \tag{32}$$

5.2. Baseline Method

A comparison of our proposed RSII with the following currently most competitive review-based recommendation methods was conducted.

The DeepCoNN model extracts user preferences from the user review documents and project attributes from the project reviews, and then concatenates the user preferences and project attributes into prediction features using vector concatenation. Finally, the prediction features are fed into a factorization machine to obtain the final predicted rating. However, its performance depends on the target reviews.

The D-Attn model is based on the DeepCoNN model and introduces two attention mechanisms: a local attention mechanism and a global attention mechanism.

The NARRE model introduces an attention mechanism on the basis of the DeepCoNN to mitigate the impact of harmful comments.

The TransNets model consists of a target network and a source network. The target network learns the representation of target comments as predictive features and guides the source network to learn the predictive feature representation, thereby, alleviating the problem of difficult access to target comments.

The CARL model utilizes attention to extract features from comments, linearly combines them with features derived from the matrix factorization of the rating matrix, and finally predicts ratings.

The MPCN model is a hierarchical neural network that first selects a few comments most relevant to the rating from all reviews, and then construct a user–item matrix to improve the recommendation performance.

The MPCAR model is based on MPCN and introduces ID information to alleviate the conflicts between the selected comments and the ratings that are not related.

In our model, we set the maximum length of each comment to 80 and the minimum length to 2. The learning rate was 0.01, the batch size was 64, the decay factor of the learning rate was 0.99, and the L2 regularization hyperparameter was 0.000001.

5.3. Comparison of Experimental Results

The proposed RSII model was compared with eight contrast algorithms on three public datasets and the experimental results are shown in Table 3. The bold text in Table 2 represents the best experimental results. Δ represents the relative percentage improvement of the RSII model compared to the optimal benchmark model. The comparison results show that both indicators were excellent on all three datasets. This also verified that the proposed model was effective. Through the analysis of the experimental results, we can draw the following conclusions.

Table 3. Comparison of evaluation for different models.

Model	MI		OP		DM	
	MAE	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**
DeepCoNN	0.622	0.557	0.682	0.718	0.783	1.055
D-Attn	0.644	0.862	0.618	0.746	0.653	0.840
NARRE	0.640	0.839	0.658	0.797	0.683	0.876
CARL	0.709	0.880	0.624	0.742	0.681	0.884
TransNets	0.784	1.162	0.661	0.718	0.825	1.197
MPCN	0.571	0.739	0.728	0.749	0.764	1.017
MPCAR	0.670	0.858	0.615	0.682	0.660	0.857
OURS	**0.482** [1]	**0.539** [1]	**0.553** [1]	**0.575** [1]	**0.619** [1]	**0.777** [1]
Δ	15.59%	3.23%	10.08%	15.69%	5.21%	7.50%

[1] The bold numbers represent the best experimental results.

First, we found that the models with the ID information (NARRE, MPCAR, CARL and RSII) performed better than those without the ID information (DeepCoNN, D-Attn and MPCN). This showed that introducing the ID information could to some extent correct the prediction error caused by the reviews. Therefore, the ID information should be introduced when designing a recommender algorithm based on reviews.

Second, compared with all the other models, the experimental results showed that the RSII model performed the best. This indicated that the dynamic fusion of the target review semantics and the ID information could enrich the prediction features and improve the recommendation performance, also demonstrating the effectiveness of the RSII model.

Thirdly, compared with the MPCAR model, the RSII model was superior. Additionally, the application scope of RSII model was larger than that of the MPCAR model. This showed that as long as the model had the ability to generate target reviews semantically and dynamically fuse the review and the ID information, it could not only expand the application scope of the model, but it also had excellent recommendation performance.

5.4. Ablation Experiment

To explore the impact of the different modules in RSII, we removed the various modules to conduct ablative studies.

RSII_si: Remove the teacher module and the fusion module.
RSII_i: Remove the fusion module and retain the teacher module.

The results of the ablation experiment are shown in Table 4. We found that the RSII performed better than the RSII_i, and the RSII_i performed better than the RSII_si. This indicates that the teacher module and fusion module are useful. RSII_i had one more supervision module than RSII_si, which indicated that the teacher module could effectively guide the review semantics learning module to learn targeted review semantics. Additionally, experiments have shown that the semantics of the target reviews can effectively improve the recommendation performance. Generally speaking, the target reviews can best reflect the users' preferences for the target items. At the same time, we found that there was a huge performance improvement of the RSII compared with the RSII_i. This also indicated that the joint learning of the review semantic learning module and the teacher module could not alleviate the problem of inconsistency between the target review and rating. However, the missing information in the target review could be extracted from the user and the item IDs, and dynamically integrating the information from reviews and IDs could enrich the prediction features.

Table 4. Table of experimental results of different modules on the RSII model.

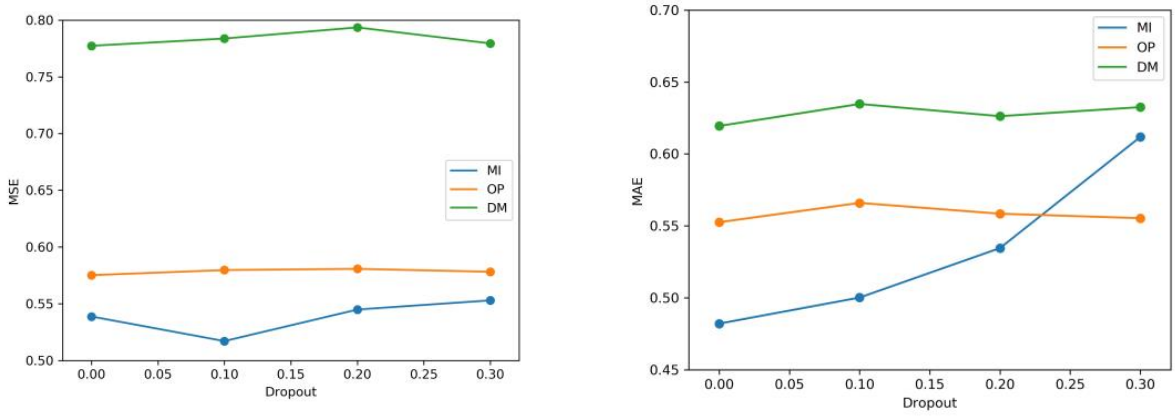
Model	MI		OP		DM	
	MAE	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**
RSII_si	0.6085	0.6528	0.7116	0.7074	0.8079	1.0939
RSII_i	0.5627	0.5390	0.6898	0.7009	0.7750	1.0962
RSII	**0.4820** [2]	**0.5387** [2]	**0.5525** [1]	**0.5751** [1]	**0.6194** [1]	**0.7773** [1]

[1] The bold numbers represent the best experimental results. [2] The bold numbers represent the best experimental results.

5.5. The Impact of Different Parameters on the Model

In this section, we used the control variable method to explore the effects of two hyperparameters (dropout and embedding size) on the model, because the model of neural networks would be affected by different parameters. Based on previous related research, we determined the range of values of different hyperparameters in this paper.

Dropout is used to solve the overfitting problem in machine learning. During the forward propagation process, it randomly stops some neurons from working with a certain probability. Different dropouts will have different effects on the model. Figure 5a,b shows the effects of different dropouts on the RSII model.

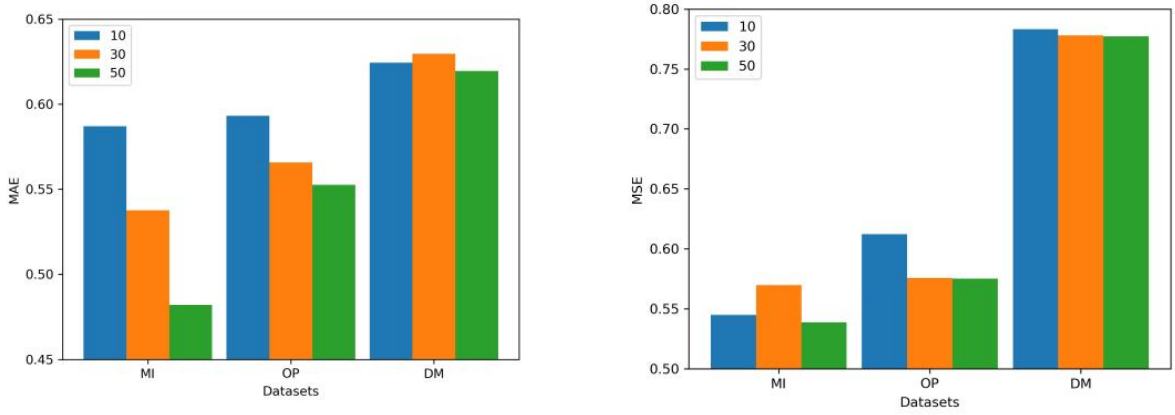


(**a**) (**b**)

Figure 5. Performance of the different dropouts of the RSII model. (**a**) The MSE of different dropouts; (**b**) The MAE of different dropouts.

We found that as the dropout increased, the recommendation performance decreased, particularly in terms of the MAE index, while the change of MSE was relatively small. MAE reflects the accuracy while MSE indicates the stability. Furthermore, when the RSII model performed best, most of the dropouts were 0. This indicated that the RSII model had low overfitting chance and higher accuracy under strong stability.

The experiments in this paper exploring the effect of d on performance were conducted, and the results are shown in Figure 6a,b. By observing Figure 6a,b, we found that for both MAE and MSE, the optimal RSII model was mostly at $d = 50$. Furthermore, the MSE did not change much with the variation of d. This indicated that the RSII model had good stability.



(**a**) (**b**)

Figure 6. Performance of the different embedding size of the RSII model. (**a**) The MAE of different embedding size; (**b**) The MSE of different embedding size.

6. Conclusions

In practice, it is difficult to obtain the target review promptly. In addition, the target review and the rating may sometimes be inconsistent (such as preference reviews and low ratings). There is currently a lack of research on the above issues. Therefore, this paper proposed a Recommendation algorithm that Simulates the generation of target review semantics and fuses ID Information (RSII). Specifically, based on the characteristics of the target review available during the model training, this paper designed a teacher module and a review semantics learning module. The teacher module learned the semantics of the target review and guided the review semantics learning model to learn these semantics. Then,

this paper used the fusion module to dynamically fuse the target review semantics and the ID information, enriching the representation of predictive features, thereby, alleviating the problem of inconsistency between the target review and the rating. Finally, the RSII model was extensively tested on three public datasets. The experiment showed that the model proposed in this article could effectively solve the above-mentioned problems.

Author Contributions: Conceptualization, Q.R.; methodology, Q.R.; software, Q.R.; validation, Q.R.; formal analysis, Q.R.; writing—original draft preparation, Q.R.; writing—review and editing, J.S.; supervision, X.S.; funding acquisition, J.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Science Fund for Outstanding Youth of Xinjiang Uygur Autonomous Region under Grant No. 2021D01E14.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: We evaluated our algorithm on four datasets. They all belong to the Amazon dataset. http://jmcauley.ucsd.edu/data/amazon/index_2014.html (accessed on 20 October 2022).

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

References

1. Srifi, M.; Oussous, A.; Ait Lahcen, A.; Mouline, S. Recommender Systems Based on Collaborative Filtering Using Review Texts—A Survey. *Information* **2020**, *11*, 317. [CrossRef]
2. Kim, T.-Y.; Ko, H.; Kim, S.-H.; Kim, H.-D. Modeling of Recommendation System Based on Emotional Information and Collaborative Filtering. *Sensors* **2021**, *21*, 1997. [CrossRef] [PubMed]
3. Margaris, D.; Vassilakis, C.; Spiliotopoulos, D. On Producing Accurate Rating Predictions in Sparse Collaborative Filtering Datasets. *Information* **2022**, *13*, 302. [CrossRef]
4. Zheng, L.; Noroozi, V.; Yu, P.S. Joint deep modeling of users and items using reviews for recommendation. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, Cambridge, UK, 6–10 February 2017; pp. 425–434.
5. Zhao, H.; Min, W.; Xu, J.; Han, Q.; Wang, Q.; Yang, Z.; Zhou, L. SPACE: Finding key-speaker in complex multi-person scenes. *IEEE Trans. Emerg. Topics Comput.* **2021**, *10*, 1645–1656. [CrossRef]
6. Wang, Q.; Min, W.; Han, Q.; Liu, Q.; Zha, C.; Zhao, H.; Wei, Z. Inter-domain adaptation label for data augmentation in vehicle re-identification. *IEEE Trans. Multimed.* **2022**, *24*, 1031–1041. [CrossRef]
7. Teicholz, P. *BIM for Facility Managers*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
8. Becerik-Gerber, B.; Jazizadeh, F.; Li, N.; Calis, G. Application areas and data requirements for BIM-enabled facilities management. *J. Constr. Eng. Manag.* **2011**, *138*, 431–442. [CrossRef]
9. Catherine, R.; Cohen, W. TransNets: Learning to Transform for Recommendation. In Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17), Como, Italy, 27–31 August 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 288–296.
10. Seo, S.; Huang, J.; Yang, H.; Liu, Y. Interpretable convolutional neural networks with dual local and global attention for review rating prediction. In Proceedings of the Eleventh ACM Conference on Recommender Systems, Como, Italy, 27–31 August 2017; pp. 297–305.
11. Chen, C.; Zhang, M.; Liu, Y.; Ma, S. Neural attentional rating regression with review explanations. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; pp. 1583–1592.
12. Tay, Y.; Luu, A.T.; Hui, S.C. Multi-pointer co-attention networks for recommendation. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 2309–2318.
13. Shao, J.; Qin, J.; Zeng, W.; Zheng, J. Multipointer Coattention Recommendation with Gated Neural Fusion between ID Embedding and Reviews. *Appl. Sci.* **2022**, *12*, 594. [CrossRef]
14. Richa, B.P. Trust and distrust based cross-domain recommender system. *Appl. Artif. Intell.* **2021**, *35*, 326–351. [CrossRef]
15. Meo, P.D. Trust prediction via matrix factorisation. *ACM Trans. Internet Technol. (TOIT)* **2019**, *19*, 1–20. [CrossRef]
16. Hassan, T. Trust and trustworthiness in social recommender systems Companion. In Proceedings of the 2019 World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 529–532.
17. Huang, J.; Rogers, S.; Joo, E. Improving restaurants by extracting subtopics from yelp reviews. In Proceedings of the iConference 2014 (Social Media Expo), Berlin, Germany, 4–7 April 2014; iSchools: Grandville, MI, USA, 2014; pp. 1–5.
18. Bao, Y.; Fang, H.; Zhang, J. Topicmf: Simultaneously exploiting ratings and reviews for recommendation. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, QC, Canada, 27–31 July 2014.

19. Haruna, K.; Akmar Ismail, M.; Suhendroyono, S.; Damiasih, D.; Pierewan, A.C.; Chiroma, H.; Herawan, T. Context-Aware Recommender System: A Review of Recent Developmental Process and Future Research Direction. *Appl. Sci.* **2017**, *7*, 1211. [CrossRef]
20. Ko, H.; Lee, S.; Park, Y.; Choi, A. A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. *Electronics* **2022**, *11*, 141. [CrossRef]
21. Beheshti, A.; Yakhchi, S.; Mousaeirad, S.; Ghafari, S.M.; Goluguri, S.R.; Edrisi, M.A. Towards Cognitive Recommender Systems. *Algorithms* **2020**, *13*, 176. [CrossRef]
22. Wu, L.; Quan, C.; Li, C.; Wang, Q.; Zheng, B.; Luo, X. A context-aware user-item representation learning for item recommendation. *ACM Trans. Inf. Syst.* **2019**, *37*, 1–29. [CrossRef]
23. Jaiswal, A.; Babu, A.R.; Zadeh, M.Z.; Banerjee, D.; Makedon, F. A Survey on Contrastive Self-teacher Learning. *Technologies* **2021**, *9*, 2. [CrossRef]
24. Albelwi, S. Survey on Self-teacher Learning: Auxiliary Pretext Tasks and Contrastive Learning Methods in Imaging. *Entropy* **2022**, *24*, 551. [CrossRef]
25. Zhou, K.; Wang, H.; Zhao, W.X.; Zhu, Y.; Wang, S.; Zhang, F. S3-Rec: Self-teacher learning for sequential recommendation with mutual information maximization. In Proceedings of the 29th ACM International Conference on Information and Knowledge Management, Virtual, 19–23 October 2020; ACM: Dublin, Ireland, 2020; pp. 1893–1902.
26. Xie, R.; Liu, Q.; Wang, L.; Liu, S.; Zhang, B.; Lin, L. Contrastive Cross-domain Recommendation in Matching. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22), Washington, DC, USA, 14–18 August 2022; Association for Computing Machinery: New York, NY, USA, 2022; pp. 4226–4236.
27. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
28. Chen, Y.; Dai, H.; Yu, X.; Hu, W.; Xie, Z.; Tan, C. Improving Ponzi Scheme Contract Detection Using Multi-Channel TextCNN and Transformer. *Sensors* **2021**, *21*, 6417. [CrossRef] [PubMed]
29. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pretraining of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
30. Dumitrescu, D.; Boiangiu, C.-A. A Study of Image Upsampling and Downsampling Filters. *Computers* **2019**, *8*, 30. [CrossRef]
31. Makarkin, M.; Bratashov, D. State-of-the-Art Approaches for Image Deconvolution Problems, including Modern Deep Learning Architectures. *Micromachines* **2021**, *12*, 1558. [CrossRef]
32. Orukwo, J.O.; Kabari, L.G. Diagnosing Diabetes Using Artificial Neural Networks. *Eur. J. Eng. Res. Sci.* **2020**, *5*, 221–224. [CrossRef]
33. Takato, S.; Shin, K.; Hajime, N. Recommendation System Based on Generative Adversarial Network with Graph Convolutional Layers. *J. Adv. Comput. Intell. Intell. Inform.* **2021**, *25*, 389–396.
34. Chae, D.; Kang, J.; Kim, S.; Lee, J. CFGAN: A generic collaborative filtering framework based on generative adversarial networks. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 137–146.
35. He, R.; McAuley, J. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In Proceedings of the 25th International Conference on World Wide Web, Montreal, QC, Canada, 11–15 April 2016; pp. 507–517.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.