Improving Recommender System via Personalized Reconstruction of Reviews

Zunfu Huang[1,2], Bo Wang[1,2](✉), Hongtao Liu[1], Qinxue Jiang[3], Naixue Xiong[4], and Yuexian Hou[1]

[1] College of Intelligence and Computing, Tianjin University, Tianjin, China
{huangwenjie,bo_wang,htliu,yxhou}@tju.edu.cn
[2] State Key Laboratory of Communication Content Cognition, People's Daily Online, Beijing, China
[3] School of Engineering, Newcastle University, Newcastle upon Tyne, UK
b9064217@Newcastle.ac.uk
[4] Department of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK, USA

Abstract. Textual reviews of items are a popular resource of online recommendation. The semantic of reviews helps to achieve improved representation of users and items for recommendation. Current review-based recommender systems understand the semantic of reviews from a static view, i.e., independent of the specific user-item pair. However, the semantic of the reviews are personalized and context-aware, i.e., same reviews can have different semantics when they are written by different users or towards different items. Therefore, we propose an improved recommendation model by reconstructing multiple reviews into a personalized document. Given a user-item pair, we design a cross-attention model to build personalized documents by selecting important words in the reviews of the given user towards the given item and vice versa. A semantic encoder of personalized document is then designed using a cross-transformer mechanism to learn document-level representation of users and items. Extensive experiments on three real-world datasets demonstrate the effectiveness of the proposed model.

Keywords: Recommender system · Personalized reconstruction · Cross attention · Cross transformer

1 Introduction

Recommender systems have played an increasingly important role in online collaboration by helping people establish connections with interested items, e.g., information pieces, products or other users. Towards more effective items recommendation, an essential problem of recommender systems is to understand the users and items on semantic level. Therefore, auxiliary texts, e.g., reviews are

H. Gao and X. Wang (Eds.): CollaborateCom 2021, LNICST 406, pp. 277–295, 2021.
https://doi.org/10.1007/978-3-030-92635-9_17

widely used to improve the representation of users and items. Auxiliary reviews are an excellent source of understanding users' interests and items' characteristics, which can improve the performance of predicting users' rating on items and alleviate the issue of data sparsity and cold start. In this direction, one main challenge is to obtain valuable semantic information from large-scale auxiliary reviews and another consequent challenge is to make the information acquisition personalized, i.e., depending on each special user and item.

To improve the semantic recommendation, in this work, we focus on rating prediction, which is a main task of recommender systems. On predicting users' rating on items, many current works have been proposed. As one of the most widely used recommendation techniques, collaborative filtering has achieved successful results based on the use of users' ratings of items [1–3]. For example, Probabilistic Matrix Factorization (PMF [3]) uses probability matrix factorization technology to learn the potential factors among the user's rating matrices for items. However, this type of method has a serious problem of sparsity, and the predicted score can only reflect the user's overall satisfaction with a certain product, and lacks interpretability. With the development of e-commerce, users are more and more willing to post their reviews on purchased products on e-commerce platforms. Textual reviews contain rich information that can describe the characteristics of users and products, and the use of review information is proved to be able to alleviate data sparseness and cold start problems. Therefore, a lot of work using review information to enhance the recommender system has been proposed in the task of rating score prediction [4–12]. The initial approaches of using reviews are mainly to obtain potential features in reviews through topic modeling [4,11]. For example, the RMR [13] proposes an interpretable LDA [14] model to extract potential features in item review documents. The shortage of existing topic-based methods is dealing with textual reviews as bag-of-words, which loses the information of word order, resulting in the inability to fully capture the semantic information of the reviews.

Recently, in order to extract potential semantic features from reviews more comprehensively, many neural network-based rating score prediction models have been proposed, such as DeepCoNN [15], NARRE [5], DAML [16] and MRCP [17]. Since the convolutional neural network(CNN) can captures the local features and contextual semantic information of the review text, these methods usually use CNN to extract features of the reviews. For example, the DeepCoNN model learns the feature representation of users and items through the parallel CNN on the semantic features of user and item review documents, and performs rating score prediction.

Current methods of using review information can be mainly divided into two types: document-based methods and review-based methods. Although review-based methods have achieved significant improvement in rating score prediction, they still lack effective interpretability for complex scoring behaviors. Document refers to connecting all reviews of users/items into a long text, and document-based methods aim to combine the semantic of single reviews to a more advanced global semantic. Liu et al. [18] has proved that document-based and review-based methods are complementary in the task of rating score prediction. That is, the document-based method and the review-based method can capture the coarse-grained and fine-grained features of reviews, respectively.

For current document-based methods, we indicate two problems: (1) In order to avoid the long tail effect of review documents, all current document-based methods obtain static partial review documents through data preprocessing, e.g., DeepCoNN [15]. As input, partial review documents lose a lot of review information, and may not be able to fully obtain the coarse-grained characteristics of users and items. (2) Current document-based methods only model the word-level relationship between the documents of users and items. However, the document-level matching is believed to be more effective to utilize global semantic, which is the essential advantage of document-based strategy [16,19].

Therefore, to solve these two problems, we propose to reconstruct reviews into compressed personalized documents which can be dealt with as a unit instead of being cut into pieces. The compressed personalized documents are personalized built by identifying the most informative words from reviews of the certain user towards a certain item or vice versa. The identified words are organized into a personalized document in proper semantic order. A cross attention strategy is designed to effectively identify informative words. Then we encode the global semantics of compressed personalized documents of users and items with improved CNN layers. Furthermore, to learn more comprehensive features, we also incorporate document-level and review-level modeling into a unified framework.

In summary, our major contributions are as follows:

(1) We propose a neural recommendation model that reconstructs reviews into personalized documents and extracts coarse-grained features of users and items by modeling the semantic of personalized documents. To the best of our knowledge, we are the first to build personalized review documents for the recommender system.
(2) We propose a cross-attention mechanism to identify important words for personalized document construction. Personalization is achieved by obtaining in-depth word-level interaction features and context-aware interaction features. We design a novel transformer mechanism to learn the interaction feature representation between user-item pairs.
(3) Experimental results on three real-world datasets show that our model is more accurate in predicting users' rating scores than best-performing baselines. At the same time, it also reveals that personalized reconstruction of reviews helps to better capture coarse-grained features.

2 Related Work

In this section, we will review the recent studies which are most related to the works of review-based recommender system.

2.1 Review-based Recommendation with Topic Modeling

Due to the "data sparsity" and "cold start" problems of the collaborative filtering methods based on user-item interaction, researchers have introduced review information to the recommender system. Initial works use topic modeling techniques to learn potential topic features from review texts [4,6,11,13]. For example, HFT [11] uses a LDA-like topic model to extract potential topic features of reviews for scoring prediction. TopicMF [4] uses MF technology to jointly model the user-item rating data and topic features of reviews. RBLT [20] utilizes ratings to promote reviews, and then combines review text and ratings to model user and item features in the shared space theme. These methods outperform models that rely solely on the user-item rating matrix. However, these bag-of-words-based models ignore the word order in the reviews, and cannot learn the local context information.

2.2 Document-Level Recommendation

Recently, document-level methods have been proposed to improve review-based recommendation. These coarse-grained recommender system methods directly combine reviews into a long document for learning the representation of users and items. For example, DeepCoNN [15] uses parallel CNN to learn the characteristics of users or items from review documents. D-Attn [21] uses local attention to learn the document-level feature representation of users and items. CARL [7] learns the interaction between user-item pairs of potential features based on context awareness, and uses it to make rating score predictions. DAML [16] uses the local and mutual attention of CNN to jointly learn the characteristics of reviews. A common problem of current document-level methods is to ignore the diverse and complex interactions between users and items.

2.3 Review-Level Recommendation

With the introduction of attention mechanism, in order to further improve the performance and interpretability of recommender systems, many fine-grained review-level methods have been proposed. For example, NARRE [5] introduces an attention mechanism to obtain the importance of each review on users or items, and provides review-level explanations for rating score prediction. In order to further filter reviews and words, MPCN [19] uses a common attention network based on Gumbel-softmax to dynamically select reviews and words that are important to target users or items. Dong et al. [22]. believes that user content is heterogeneous while item content is homogeneous. Therefore, they proposed the AHN model, which uses co-attention at the sentence-level and review-level to guide the representation learning of the user/item related to the target item/User. In order to increase the interpretability of fine-grained semantic features, MRCP [17] dynamically learns the feature representation of users and items through a three-layer attention framework of word-level, review-level and aspect-level. Although these review-level-based methods can capture more fine-grained features, they cannot effectively understand the coarse-grained features of users and items.

Compared with current methods, our proposed model can not only reconstruct ordered personalized review documents but also uses a transformer mechanism to capture the in-depth interaction characteristics between users and items.

3 Methodology

In this section, we present our proposed method Recommendation with Personalized Reconstruction of Reviews (PPRR) in detail. The overview of PPRR is shown in Fig. 1. The model has three stages: review document reconstruction network(Re-Doc-Net), document-level compiler(Doc-Net), and review-level compiler(Review-Net). The interactive attention characteristics of users and items are captured from word-level and review-level respectively. We will introduce the details of our model (PPRR) in detail below.

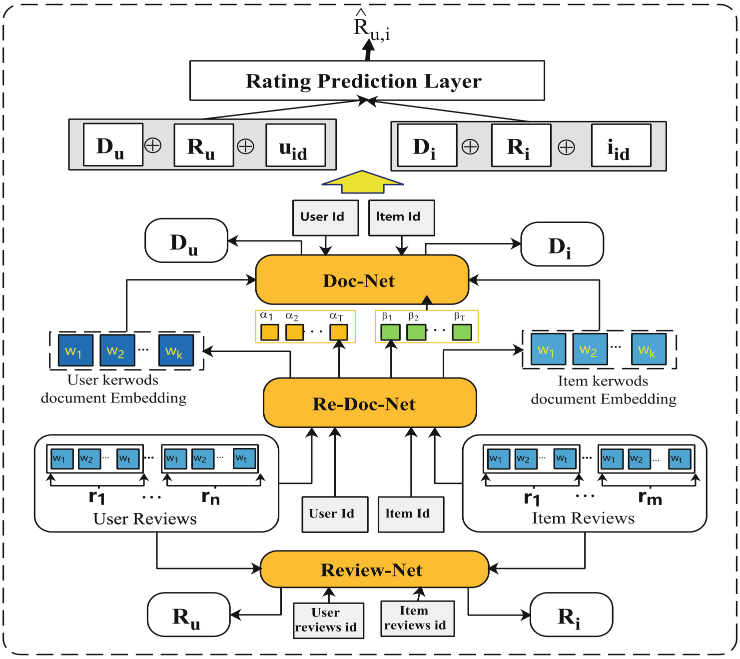

Fig. 1. In the framework of our PPRR method, the three main components are: review document reconstruction network (Re-Doc-Net), document-level encode network (Doc-Net), and review-level encode network (Review-Net).

3.1 Probem Definition

Suppose there is user set U, item set I and the rating matrix $\boldsymbol{R} \in \mathrm{R}^{|U|\times|I|}$, where the entry R_{ui} indicates the rating of user $u \in U$ towards item $i \in I$. For a user u, the reviews written by u can be noted as $r_u = \{r_{u,1}, \cdots, r_{u,n}\}$ where n is the number of reviews. For an item i, all reviews towards i can be represented as $r_i = \{r_{i,1}, \cdots, r_{i,m}\}$ where m is the number of reviews. For a single review r, we represent it as $r = \{w_{r,1}, \cdots, w_{r,t}\}$ where t is the length of r. Additionally, we reconstruct each r_u or r_i into document. We denote the document of r_u as $d_u = \{w_1^u, \cdots, w_T^u\}$, and denote document of r_i as $d_i = \left\{w_1^i, \cdots, w_T^i\right\}$ where w and T are the words in the document and the number of words, respectively.

3.2 Overall Framework of PPRR

The proposed PPRR model has three main stages: dynamically selecting informative words by cross attention to build compressed personalized documents of users and items, encoding the compressed personalized documents with CNN layers, and involving the review-level features into document-level representation.

In detail, we firstly design a personalized document builder to dynamically extract the most informative words in reviews. The informative words are personalized identified according to the semantic relation between certain users and items. That is, given a pair of users and items, we identify the words in the item's texts that are more relevant to the user and vice versa. A cross-attention model is designed to make the personalized identification of informative words. Afterwards, the extracted informative words are organized into compressed personalized documents keeping their order in original reviews, which avoids the disadvantage of bag-of-words.

Sequentially, for obtaining compressed personalized documents, we design a document encoder to learn document-level semantic representations of users and items. We use transformers to involve the context characteristics of documents and the characteristics of each word. Finally, we learn the document-level semantic representation of documents through CNN layers.

As supplement, we also involve a parallel review encoder to learn the review-level features of users and items. At the review-level, we learn the feature representation of each review through the weight of each word obtained by the previous cross attention model. Then, the review feature is used as query vector, and the weight of each review is obtained through the cross-attention mechanism.

Finally, we fuse the document-level and review-level semantics to obtain the final comprehensive representations of users and items for rating score prediction in recommendation.

3.3 Review Document Reconstruction Network (Re-Doc-Net)

Word Embedding Layer. Given a review $r = \{w_1, w_2, \cdots, w_t\}$ which has t words, we fill all words into an embedding matrix $\boldsymbol{W}^{d\times|V|}$ where V is the

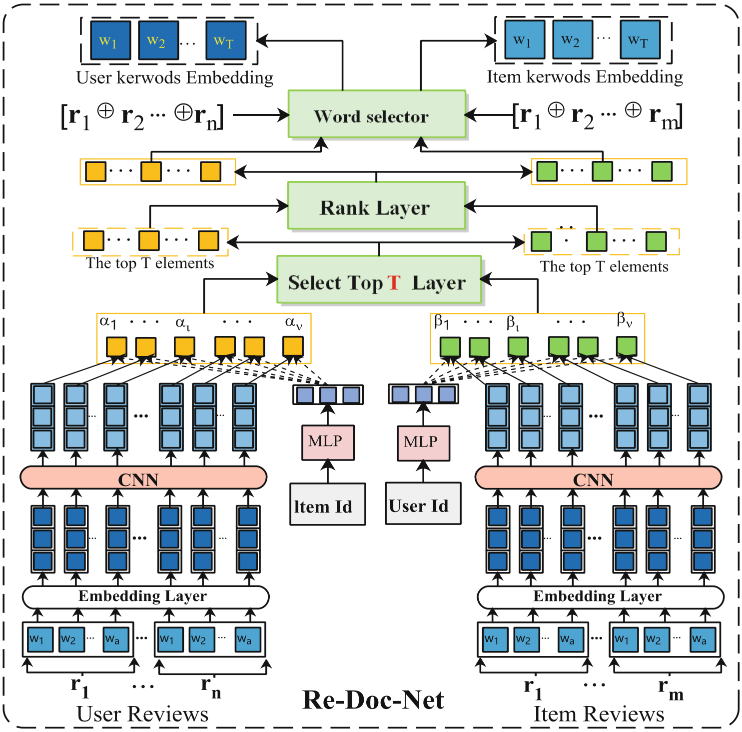

Fig. 2. The document reconstruction network of our PPRR approach, which uses the target ID embedding as the query vector, and obtains the weight of each word through the cross-attention mechanism. In addition, the three operations of **Select Top T Layer**, **Rank layer**, and **Word selector** all sort and select words on the vocabulary without any parameters.

vocabulary of words, d is the dimension of each word vector. Then the embedding vector of review r is $\boldsymbol{r} \in R^{t \times dw}$.

ID Embedding Layer. IDs are usually regarded as identity information of the corresponding user and item in recommender system. Therefore, by encoding the IDs of users and items into low-dimensional vectors $\mathbf{u_{id}}$ and $\mathbf{i_{id}}$, the ID embeddings are denoted as

$$\mathbf{ID^u_r} = \{\mathrm{ID^u_{r,1}} \cdot\cdot, \mathrm{ID^u_{r,n}}\}, \mathbf{ID^i_r} = \{\mathrm{ID^i_{r,1}} \cdot\cdot, \mathrm{ID^i_{r,m}}\} \tag{1}$$

where $\mathbf{ID^u_r} \in \boldsymbol{R}^{n \times z}$ and $\mathbf{ID^i_r} \in \boldsymbol{R}^{m \times z}$ are the matrix of user's reviews ID embedding and item's reviews ID embedding, respectively. z is the dimension of ID embedding.

All ID embeddings are initialized randomly, where the ID embeddings of users and items can not only index the identity of each user, but also learn the attention query vector of the words of each user or item. In addition, user and item review ID embedding can be used to characterize the usefulness of user and item reviews.

Convolution Layer. We utilize CNN to extract the semantic feature of $\boldsymbol{r}$. It consists of K different convolution filters, and each a filter $f \in \mathbb{R}^{l \times dw}$ where l is the filter window size which produces features by applying convolution operator on word vectors matrix. Then, j^{th} filter produces its features as:

$$c_j = \mathrm{ReLU}(\boldsymbol{r} * f_j + b_j) \tag{2}$$

where b_j is the bias, $*$ is the convolution operation and ReLU is a nonlinear activation function. Then, the final features $\boldsymbol{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_K\}$ produced by the K filters. Thus the m-th row of $\boldsymbol{C} \in R^{t \times K}$ is the feature of the m-th word in the review $r_{u,i}$, denoted as $\mathbf{c_m} \in R^K$.

Cross Attention over Word Level. After the above two layers, we have obtained the feature vectors of all IDs ($\mathbf{u_{id}}$, $\mathbf{i_{id}}$) and the semantic features of all words ($\boldsymbol{C}$). For the same user, the importance of words in different item reviews is different. Therefore, in order to capture the importance of words in different contexts (contextual semantic features), we introduce the ID embedding of the target item, generate word-level cross-attention query vectors through a Multilayer Perceptron (MLP), and then capture the importance of words through a cross-attention mechanism:

$$\mathbf{q_w^u} = \mathrm{ReLU}\left(\mathbf{W_w^u i_{id}} + \mathbf{b}\right) \tag{3}$$

$$g_{\mathrm{x}} = \mathbf{q_w^u A c_x}, \quad \alpha_{\mathrm{x}} = \frac{\exp\left(\mathrm{g_x}\right)}{\sum\limits_{\mathrm{j=1}}^{\mathrm{t}} \exp\left(\mathrm{g_j}\right)}, \alpha_{\mathrm{x}} \in (0, 1\,) \tag{4}$$

where $\mathbf{q_w^u}$ is the user word-level query attention vector derived from the corresponding item, $\mathbf{W_w^u}$ is parameter matrix, $\mathbf{A}$ is the harmony matrix in attention, α_x is the attention weight of the m-th word of a review. ReLU is a nonlinear activation function. Then, we aggregate the reviews according to the weight of the words in each review to obtain the feature vector of the review: $\mathbf{h}_{\mathrm{u,x}} = \sum\limits_{j=1}^{t} \alpha_{x,j} \boldsymbol{c}_{x,j}$, then the review feature vector matrices of users and items are $\mathbf{h_u} \in \boldsymbol{R}^{n \times z}$ and $\mathbf{h_i} \in \boldsymbol{R}^{m \times z}$ respectively.

In order to obtain the user's coarse-grained characteristics more comprehensively and to capture the full contextual awareness information (contextual semantic information) to the greatest extent, we dynamically selected the X most important words in the user's reviews under the current target item and reconstructed ordered personalized review documents as shown in Fig. 2. The specific implementation method is as follows:

$$\mathrm{top\_}\alpha_{\mathrm{u}}, \mathrm{top\_idx_u} = \mathrm{TopX}\left(\mathrm{Rank}\left(\boldsymbol{\alpha}\,\right)\right) \tag{5}$$

$$\mathrm{idx_u}, \mathrm{idx\_idx_u} = \mathrm{Rank}\left(\mathrm{top\_idx_u}\right) \tag{6}$$

$$\mathbf{d_u} = \mathrm{Select}\left([\mathbf{r}_1 \oplus \mathbf{r}_2 \cdots \oplus \mathbf{r}_\mathrm{n}], \mathrm{idx_u}\right), \quad \alpha^{\mathrm{u}}_{\mathrm{dw}} = \mathrm{Select}\left(\mathrm{top\_}\alpha_{\mathrm{u}}, \mathrm{idx\_idx_u}\right) \tag{7}$$

where Rank(·) is the sorting method, which sorts the internal parameters in descending order, $\boldsymbol{\alpha}$ is the weight of all words in user reviews, TopX(·) is the method of selecting the top X from the internal parameters, $\mathrm{top\_}\alpha_{\mathrm{u}}, \mathrm{top\_idx_u}$ is the weight and position subscript of the first x words with the highest previous importance, respectively, $\mathrm{idx_u}, \mathrm{idx\_idx_u}$ a represents the weight subscripts and the subscripts of the subscripts after reordering the first X weight subscripts. Select(·) is to select important words and weights in user reviews based on subscripts, where $\mathbf{d_u}$ is a newly constructed vector of ordered review documents, $\alpha^{\mathrm{u}}_{\mathrm{dw}}$ is the weight of each word vector in $\mathbf{d_u}$. The above process only sorts the data, and does not add any parameters to the model. Similarly, we can obtain new ordered item review-documents $\mathbf{d_i}$ and word weights $\alpha^{\mathrm{i}}_{\mathrm{dw}}$ through the above methods.

3.4 Document-Level Encode Network(Doc-Net)

From the review document reconstruction network, we have constructed a new ordered user review document $\mathbf{d_u}$ and item review document $\mathbf{d_i}$. The document-based compilation network is shown in Fig. 3, in order to further obtain user-item complex interaction features and contextual information, We integrate ID embedding information and word position embedding information into each word embedding, and then use the transformer to capture the interaction features at the word-level. The specific calculation method is as follows:

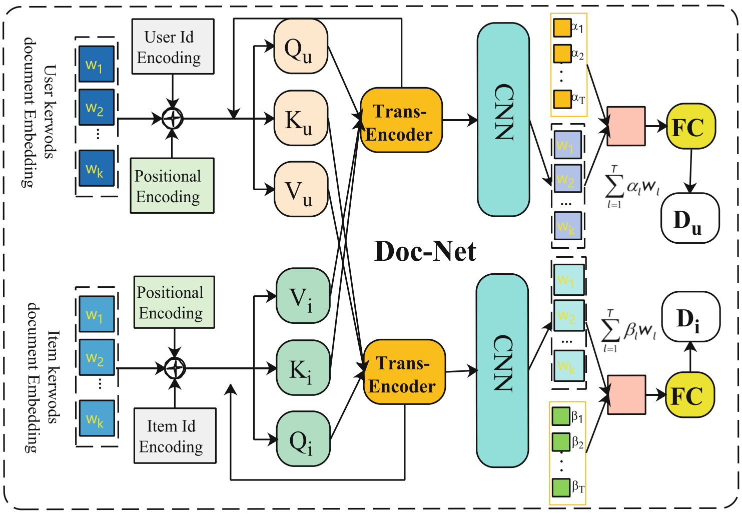

Fig. 3. The Doc-Net of our PPRR approach, which is mainly for feature extraction of personalized review documents, where the cross-transformer mechanism is mainly to dynamically enable word-level information to include target word-level and contextual features.

$$\mathbf{d_w^u} = \mathrm{ReLU}\left(\mathbf{W_w^u u_{id}} + \mathbf{b}\right), \quad \mathbf{d_w^i} = \mathrm{ReLU}\left(\mathbf{W_w^i i_{id}} + \mathbf{b}\right) \tag{8}$$

$$\mathbf{p_u} = \mathbf{d_u} + \mathbf{d_w^u} + \mathbf{pos_u}, \quad \mathbf{p_i} = \mathbf{d_i} + \mathbf{d_w^i} + \mathbf{pos_i} \tag{9}$$

where $\mathbf{W_w^u}$ and $\mathbf{W_w^u}$ are parameter matrices, $\mathbf{d_w^u}$ and $\mathbf{d_w^u}$ are the word-level query attention vector of the review-document, $\mathbf{pos_u}$ and $\mathbf{pos_i}$ are user and item position encodings with the same dimensions as the word embedding, and the encoding method is the same as that of $Transformer$.

$$\begin{aligned} \mathbf{Q_u} = \mathbf{W_u^Q p_u}, \quad \mathbf{K_u} = \mathbf{W_u^K p_u}, \quad \mathbf{V_u} = \mathbf{W_u^V p_u} \\ \mathbf{Q_i} = \mathbf{W_i^Q p_i}, \quad \mathbf{K_i} = \mathbf{W_i^K p_i}, \quad \mathbf{V_i} = \mathbf{W_i^V p_i} \end{aligned} \tag{10}$$

$$\mathbf{d_u^T} = \boldsymbol{Trans}(\mathbf{Q_u}, \mathbf{K_i}, \mathbf{V_i}), \quad \mathbf{d_i^T} = \boldsymbol{Trans}(\mathbf{Q_i}, \mathbf{K_u}, \mathbf{V_u}) \tag{11}$$

where $\mathbf{W_u^Q}$, $\mathbf{W_u^K}$, $\mathbf{W_u^V}$ are parameter matrices related to the user. $\mathbf{W_i^Q}$, $\mathbf{W_i^K}$, $\mathbf{W_i^V}$ are parameter matrices related to the item. $\boldsymbol{Trans}(\cdot)$ is the encoder of the transformer mechanism. $\mathbf{d_u^T}$ and $\mathbf{d_i^T}$ are the encoded review-document vectors.

Then, we capture the feature matrix $\boldsymbol{w_u}$ of all words of $\mathbf{d_u^T}$ through the convolutional neural network. Through the above process, we have obtained the weight $\alpha_{\mathrm{dw}}^{\mathrm{u}}$ of each word embedding in the user review document. Finally, we obtain the representation vector of the review document $\boldsymbol{D_u}$ of user u based on the item i via weighted summation of all words:

$$\boldsymbol{D_u} = \sum_{j=1}^{T} \alpha_u^j \boldsymbol{w_{u,j}} \tag{12}$$

Likewise, We can get the feature representation $\boldsymbol{D_i}$ of the item review document.

3.5 Review-Level Encode Network(Review-Net)

As shown in Fig. 4, in order to capture the fine-grained features of users and items, we introduce user and item review ID embeddings ($\mathbf{ID_r^u} \in \boldsymbol{R}^{n\times z}$, $\mathbf{ID_r^i} \in \boldsymbol{R}^{m\times z}$) to model the quality of reviews, and capture the interactive features between user and item reviews through cross attention method. From the above process, we have obtained the review feature vector matrices $\mathbf{h}_{\mathrm{u,r}}$ and $\mathbf{h}_{\mathrm{i,r}}$ of users and items. The calculation method is as follows:

$$\boldsymbol{f_u} = \mathrm{ReLU}\left(\boldsymbol{W_r^u}\mathbf{h_{u,r}} + \boldsymbol{W_{rid}^u}\mathbf{ID_r^u} + \boldsymbol{b_1}\right) \tag{13}$$

where $\boldsymbol{W_r^u} \in R^{k\times z}$,$\boldsymbol{W_{rid}^u} \in R^{z\times z}$, $\boldsymbol{b}_1$ are model parameters. ReLU is a nonlinear activation function. Then, we obtain the feature $\boldsymbol{f_i}$ of the item in the same way. Then the user-item correlation matrix is calculated as follows:

$$e_j = \boldsymbol{W}\left(\boldsymbol{f_u f_i^T}\right) + \boldsymbol{b}, \quad \alpha_{\mathrm{j}} = \frac{\exp(\mathrm{e_j})}{\sum_{\mathrm{k=1}}^{\mathrm{n}} \exp(\mathrm{e_k})}, \alpha_{\mathrm{j}} \in (0, 1), \tag{14}$$

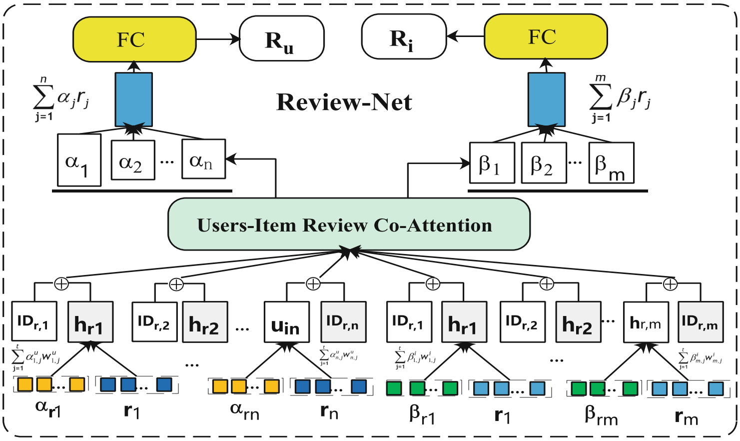

Fig. 4. The Review-Net of our PPRR approach, which mainly extracts the features of each review, and the cross-attention mechanism is mainly to dynamically obtain the weight of each review.

where $\boldsymbol{W} \in R^m$ is the weight matrix of the fully connected layer. $\boldsymbol{\alpha} \in R^n$ is the weight matrix of user reviews, which represents the importance of each review, then the fine-grained feature of users based on reviews is $\boldsymbol{R_u}$:

$$\boldsymbol{R}_u = \boldsymbol{W}_1 \left(\sum_{j=1}^{n} \alpha_j \mathbf{h}_{\mathrm{u,j}} \right) + b. \tag{15}$$

where $\boldsymbol{W}_1 \in R^{k \times z}$ and b are the weight matrix of the fully connected layer. Then, We use the same method to obtain the final fine-grained feature vector $\boldsymbol{R_i}$ of the item based on reviews.

3.6 Rating Score Prediction Layer

After the above process, we have obtained the characteristics of users and items at the document-level and review-level. Since document-level features and review-level features can describe the coarse-grained features and fine-grained features of users/items, respectively, and the ID embedding of users/items can identify their identity information, we can better perform rating score predictions by combining these three features. Then the final latent features of users and items are denoted as $\mathbf{u}$ and $\mathbf{i}$ respectively. The calculation method is as follows:

$$\begin{aligned} \mathbf{u} &= \boldsymbol{D_u} \oplus \boldsymbol{R_u} \oplus \mathbf{u_{id}} \\ \mathbf{i} &= \boldsymbol{D_i} \oplus \boldsymbol{R_i} \oplus \mathbf{i_{id}} \end{aligned} \tag{16}$$

where $\oplus$ is the concatenation operator.

Next, We use Latent Factor Model(PFM), which is widely used for scoring prediction, to make the final score prediction. The predicted rating $\hat{R}_{u,i}$ is computed as follows:

$$\hat{R}_{u,i} = \boldsymbol{W_T^2}\,(\mathbf{u} \oplus \mathbf{i}\,) + \boldsymbol{b_u} + \boldsymbol{b_i} + \boldsymbol{\mu} \tag{17}$$

where $\boldsymbol{W_T^2}$ is the linear transform matrix of the LFM, $\boldsymbol{b}_u$, $\boldsymbol{b}_i$ and $\boldsymbol{\mu}$ are user bias, project bias and global bias, respectively. Since our task is to make rating score predictions, we utilize the Mean Square Error (MSE) function to train our model:

$$L_{sqr} = \sum_{u,i \in \boldsymbol{\Omega}} \left(\hat{R}_{u,i} - R_{u,i}\,\right)^2 \tag{18}$$

where $\boldsymbol{\Omega}$ denotes the set of instances for training, and $R_{u,i}$ is the ground truth rating assigned by the user u to the item i.

4 Experiments and Analysis

4.1 DataSets and Experiments Settings

DataSets. In order to prove the effectiveness of our method, we selected real datasets of three fields of *Digital Music*, *Office Products*, and *Tools Improvement* from Amazon[1] core datasets as the data for our experiment. These datasets contain the actual ratings (1–5) and review texts made by users on the items. Following the preprocessing steps used in [5,17], we solve the long-tail effect of reviews by maintaining the length and number of reviews that can cover $p(p = 0.85)$ percentage of users and items. The detailed information of the datasets is shown in Table 1.

Table 1. Comparison of effects between various models.

Datasets	# Users	# Items	# Rating	# Avg. review length	# Words per user	# Words per item	Density(%)
Digital music	5,540	3,568	64,666	69.57	216.21	266.51	0.327
Office products	4,905	2,420	53,228	48.15	197.93	229.52	0.448
Tools improvement	16,638	10,217	134,345	38.75	162.53	212.48	0.079

For evaluation, for each data set, we randomly split 80% of the data as the training set and 10% as the validation set and test set where the validation set for hyper-parameter validation. We ensure that the training set contains at least one user-item interaction, and does not contain any reviews in the validation set and test set.

[1] http://jmcauley.ucsd.edu/data/amazon/.

Baselines. We compare the proposed PPRR with the following state-of-the-art rating score prediction methods:

- **PMF** [3]**:** Probabilistic matrix factorization is a standard matrix factorization that only utilizes rating data to model the characteristics of users and items.
- **RBLT** [20]**:** This method utilizes a rating-boosted method to combine reviews and ratings to learn the item recommended and user preference distribution in a shared topic space.
- **DeepCoNN** [15]**:** This method utilizes two parallel CNN networks to learn the feature representations of users and items from user review documents and item review documents, and then performs rating score predictions through FM.
- **D-Attn** [21]**:** In order to increase the interpretability of recommendations, the model uses local and global dual attention to capture the interpretable features of users and items.
- **NARRE** [5]**:** This method utilizes a neural attention mechanism to capture the importance and potential features of different reviews, and then combines the reviews and items to make rating predictions.
- **MPCN** [19]**:** This method introduces the Gumbel-softmax pointer mechanism into the neural network to jointly learn and pay attention to the user-item in-depth interaction.
- **DAML** [16]**:** The model uses the dual mechanism of local and mutual attention to jointly learn the features of reviews, and combined the rating features to complete the final rating score prediction.
- **CARP** [7]**:** CARP uses the capsule network to learn the feature representations of users and items from review documents and infer the corresponding emotions to increase the interpretability of rating score predictions.
- **MRCP** [17]**:** MRCP comprehensively learns the feature representation of users/items from word-level, review-level, and aspect-level by using a three-layer attention framework.
- **NRCA** [18]**:** This model proposes a neural network recommendation method under the cross-attention framework, which combines document-level and review-level features of users and items, which can more comprehensively capture the representation of users and items.

In addition to the methods mentioned above, there are many well-known methods, such as HFT [11], ConvMF [23], TARMF [10], ANR [24]. These approaches were not involved because they did not perform as well as other baseline methods in our experiments.

Hyper-parameter Settings. In experiments, we used pre-trained word embeddings with 300-dimensions which are trained on more than 100 billion words from Google News to initialize our word embedding. The dimension of the user or item ID embedding (i.e., d) is 32, the number of convolution filters neurons k is 100 (tuning in 50, 100, 150, 200), and the window size of CNN (i.e., l) is 3. The number of words T we choose to construct the review document is set to 200. For the *Transformer* encoding part, we set the overall default number of layers to 1, and the number of multi-head attention layers to 10. We show the performance of the model under different length of review document in Sect. 4.4. In addition, we utilize dropout technology to alleviate the overfitting problem of the model, and the internal dropout ratio of the transformer mechanism of the model is set to 0.5, and the rest of the dropout ratio is set to 0.75. All weight matrices in the model are initialized with a normal distribution with a mean value of 0.0 and a standard deviation of 0.1, and all biases are initialized to 0.1. The non-linear activation function is $ReLU$.

For the optimization of the model, we utilize the Adam optimization strategy, set the learning rate to 0.004, and the weight decay to 0.001. At the same time, using the validation set to tune the overall hyper-parameters of the model.

Evaluation Metric. For the task of rating score prediction in this work, we utilize the well-known Mean Square Error (MSE) as the evaluation metric. Lower MSE indicates the predicted rating $\hat{R}_{u,i}$ is closer to the ground truth rating $R_{u,i}$:

$$MSE = \frac{1}{|\Omega_{\mathrm{t}}|} \sum_{(u,i)\in\Omega_{\mathrm{t}}} \left(\hat{R}_{u,i} - R_{u,i}\right)^2 \tag{19}$$

where Ω_{t} is the set of the user-item pairs in the testing set.

4.2 Performance Evaluation

The experimental results of all methods over the three datasets are shown in Table 2, in which we have the following observations and analysis:

(1) The review-based method outperforms the rating-based method (i.e., PFM [3]), which demonstrates the advantage of reviews by involving semantic information.
(2) The neural network method outperforms the topic-based method (i.e., RBLT [20]). The main reason is that the neural network can well capture the semantic information in the review and better learn the characteristics of the user/item.
(3) The performance of the method using attention mechanism (i.e., D-Attn, DML, NARRE, MRCP) is generally better than the method without attention (i.e., DeepCoNN), which is mainly because the utility of each word or review may be different, and the attention mechanism can pay attention to these differences.
(4) There is no obvious difference between the document-based method and the review-based method, but our PPRR method achieves the best performance in all data sets by reconstructing the review document at the review-level to capture the fine-grained and coarse-grained features. This experimental result proves the effectiveness of the proposed PPRR model.

Table 2. Comparison of our model **PPRR** and each baseline method on MSE.

Type	Method	Datasets		
		Digital music	Office products	Tools improvement
Rating-based	PFM [3]	1.206	1.092	1.566
Topic-based	RBLT [20]	0.870	0.759	0.983
Document-based	DeepCoNN [15]	1.056	0.860	1.061
	D-Attn [21]	0.911	0.825	1.043
	DAML [16]	0.813	0.705	0.945
	CARP [7]	0.820	0.719	0.960
Review-based	NARRE [5]	0.812	0.732	0.957
	MPCN [19]	0.903	0.769	1.017
	MRCP [17]	0.801	0.702	0.928
Review-and-documnet	NRCA [18]	0.795	0.691	0.929
Proposed method	**PPRR**	**0.779**	**0.678**	**0.923**

4.3 Discussion

In our model, we use word and review level attention to indicate the personalized importance of different words and reviews, and then build a new ordered review document by selecting the most important words, and finally combine the fine-grained features of the review level and coarse-grained features of the document level to learn a comprehensive representation of users and items. In this section, we use ablation experiments to study the effectiveness of the important components of the model, which mainly include three components: review document reconstruction network (Re-Doc-Net), review compiler (Rev-net), and cross-transform (Transformer). Therefore, we designed multiple variants for ablation experiments as follows:

- **PPRR-IT:** the model removes the review-level module, and the input of document-level module is an externally preprocessed static review document(IT).
- **PPRR-ST:** the model removes the review-level module, and the input of the document-level module is reconstructed by the document reconstruction (ST).
- **PPRR-R:** the document-level modules are removed from the model.
- **PPRR-RS:** the model removes the document-level transform mechanism.
- **PPRR-RIS:** the model removes the review-document reconstruction, and uses an external static review document as input in the document-level module.

The experimental results are shown in Table 3. We set all the ablation experimental model parameters to the default values of PPRR. First of all, from the experimental results of the review-level based module (PPRR-R) and document-level module (PPRR-ST), which verified the review-level module and document-level module of the model are effective in capturing the coarse-grained and fine-grained features of users/items. Secondly, the experimental effect of the

Table 3. MSE comparison of different components of model **PPRR**.

Variant	Datasets		
	Digital music	Tools inprovement	Office products
PPRR-IT	0.7806	0.9251	0.6835
PPRR-ST	0.7796	0.9246	0.6808
PPRR-R	0.7823	0.9246	0.6819
PPRR-RS	0.7809	0.9234	0.6796
PPRR-RIT	0.7832	0.9236	0.6813
PPRR	**0.7791**	**0.9230**	**0.6776**

static review-document model (PPRR-IT) is worse than that of the personalized review-document level model (PPRR-ST), which can be seen that the personalized reconstruction of the review-document can more comprehensively capture the user/item Coarse-grained features. In addition, we suppose that each word in the review-document can better learn the user/item features after capturing each word of the target document and the contextual features, so we set up the PPRR-RS model, and the experimental results prove our conclusion.

4.4 Hyper-Parameters Analyses

In this section, we analyze three key hyper parameters in the model: the length of the reconstructed review document, the number of layers of the transformer mechanism, and the number of convolution filters to explore the effectiveness of the hyper parameters for our model. Through experiments, we found that when the length of the reconstructed review document is T = 200, the number of transform mechanism layers is 1, and the number of convolution kernels k = 100, the model has the best effect. Here we only show the length of the reconstructed review-document.

Effect of Review-document Length. Considering that the length of the review document plays a very important role in the overall performance of our model PPRR, we analyze the influence of the length of the review document on {100,200,300,400,500,600} through experiments on all datasets. The experimental results are shown in Table 4. From the experimental results, we find when the length of the review document T = 200, the model works best, which shows that the document length is not as long as possible.

Table 4. The result of different review-document length in PPRR.

Review-document length	Datasets		
	Digital music	Tools inprovement	Office products
T = 100	0.7820	0.9248	0.6811
T = 200	**0.7791**	0.9231	**0.6776**
T = 300	0.7820	0.9304	0.6802
T = 400	0.7838	0.9253	0.6812
T = 500	0.7806	**0.9212**	0.6810
T = 600	0.7795	0.9263	0.6800

5 Conclusion

In this article, we propose a neural recommendation model that can reconstruct reviews into personalized documents and integrate review-level features. In the reconstructed document, we represent the personalized semantic of the reviews of a certain user towards a certain item and vice versa by modeling the personalized importance of each word in the reviews. We propose a cross-transform mechanism to achieve the personalized importance calculating by integrating contextual semantic features of each word. Experimental results on three Amazon public data sets show that the proposed model can effectively improve recommendation performance compared with state-of-the-art baselines. And the reconstruction of personalized documents of reviews is verified to be the essential advantage of the proposed model.

Acknowledgement. This work was supported by a grant from the National Key Research and Development Program of China (2018YFC0809804), State Key Laboratory of Communication Content Cognition (Grant No. A32003), the Artificial Intelligence for Sustainable Development Goals (AI4SDGs) Research Program, National Natural Science Foundation of China (U1736103, 61976154, 61402323, 61876128), the National Key Research and Development Program (2017YFE0111900).

References

1. Koren, Y.: Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2008, pp. 426–434. New York, NY, USA. Association for Computing Machinery (2008)
2. Linden, G., Smith, B., York, J.: Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput. **7**(1), 76–80 (2003)
3. Salakhutdinov, R., Mnih, A.: Probabilistic matrix factorization. In: Platt, J.C., Koller, D., Singer, Y., Roweis, S.T. (eds.), Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, 3–6 December 2007, pp. 1257–1264. Curran Associates Inc (2007)

4. Bao, Y., Fang, H., Zhang, J.: Topicmf: simultaneously exploiting ratings and reviews for recommendation. In: Brodley, C.E., Stone, P. (eds.), Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 27–31 July 2014, Québec City, Québec, Canada, pp. 2–8. AAAI Press (2014)
5. Chen, C., Zhang, M., Liu, Y., Ma, S.: Neural attentional rating regression with review-level explanations. In: Proceedings of the 2018 World Wide Web Conference, WWW 2018, p. 1583C1592, Republic and Canton of Geneva, CHE 2018. International World Wide Web Conferences Steering Committee (2018)
6. Diao, Q., Qiu, M., Wu, C-Y., Smola, A.J., Jiang, J., Wang, C.: Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). In: Macskassy, S.A., Perlich, C., Leskovec, J., Wang, W., Ghani, R. (eds.), The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2014, New York, NY, USA - 24–27 August 2014, pp. 193–202. ACM (2014)
7. Li, C., Quan, C., Peng, L., Qi, Y., Deng, Y., Wu, L.: A capsule network for recommendation and explaining what you like and dislike. In: Piwowarski, B., Chevalier, M., Gaussier, É., Maarek, Y., Nie, J.-Y., Scholer, F. (eds.), Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, 21–25 July 2019, pp. 275–284. ACM (2019)
8. Liu, H., et al.: Hybrid neural recommendation with joint deep representation learning of ratings and reviews. Neurocomputing **374**, 77–85 (2020)
9. Liu, H., et al.: NRPA: neural recommendation with personalized attention. In: Piwowarski, B., Chevalier, M., Gaussier, É., Maarek, Y., Nie, J.-Y., Scholer, F. (eds.), Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, 21–25 July 2019, pp. 1233–1236. ACM (2019)
10. Lu, Y., Dong, R., Smyth, B.: Coevolutionary recommendation model: mutual learning between ratings and reviews. In: Champin, P-A., Gandon, F., Lalmas, M., Ipeirotis, P.G. (eds.), Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, 23–27 April 2018, pp. 773–782. ACM (2018)
11. Julian J. McAuley and Jure Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. In Qiang Yang, Irwin King, Qing Li, Pearl Pu, and George Karypis, editors, Seventh ACM Conference on Recommender Systems, RecSys '13, Hong Kong, China, October 12–16, 2013, pages 165–172. ACM, 2013
12. Wang, X., et al.: Neural review rating prediction with hierarchical attentions and latent factors. In: Li, G., Yang, J., Gama, J., Natwichai, J., Tong, Y. (eds.) DASFAA 2019. LNCS, vol. 11448, pp. 363–367. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-18590-9_46
13. Ling, G., Lyu, M.R., King, I.: Ratings meet reviews, a combined approach to recommend. In: Kobsa, A., Zhou, M.X., Ester, M., Koren, Y. (eds.), Eighth ACM Conference on Recommender Systems, RecSys 2014, Foster City, Silicon Valley, CA, USA - 06–10 October 2014, pp. 105–112. ACM (2014)
14. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993C1022 (2003)
15. Zheng, L., Noroozi, V., Yu, P.S.: Joint deep modeling of users and items using reviews for recommendation. CoRR, abs/1701.04783 (2017)

16. Liu, D., Li, J., Du, B., Chang, J., Gao, R.: DAML: dual attention mutual learning between ratings and reviews for item recommendation. In: Teredesai, A., Kumar, V., Li, Y., Rosales, R., Terzi, E., Karypis, G. (eds.), Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, 4–8 August 2019, pp. 344–352. ACM (2019)
17. Liu, H., Wang, W., Peng, Q., Wu, N., Wu, F., Jiao, P.: Toward comprehensive user and item representations via three-tier attention network. ACM Trans. Inf. Syst. **39**(3), 1–22 (2021)
18. Liu, H., Wang, W., Xu, H., Peng, Q., Jiao, P.: Neural unified review recommendation with cross attention. In: Huang, J., et al. (eds.), Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, 25–30 July 2020, pp. 1789–1792. ACM (2020)
19. Tay, Y., Luu, A.T., Hui, S.C.: Multi-pointer co-attention networks for recommendation. In: Guo, Y., Farooq, F., (eds.), Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, 19–23 August 2018, pp. 2309–2318. ACM (2018)
20. Tan, Y., Zhang, M., Liu, Y., Ma, S.: Rating-boosted latent topics: understanding users and items with ratings and reviews. In: Kambhampati, S. (ed), Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9–15 July 2016, pp. 2640–2646. IJCAI/AAAI Press (2016)
21. Seo, S., Huang, J., Yang, H., Liu, Y.: Interpretable convolutional neural networks with dual local and global attention for review rating prediction. In: Cremonesi, P., Ricci, F., Berkovsky, S., Tuzhilin, A. (eds.), Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys 2017, Como, Italy, 27–31 August 2017, pp. 297–305. ACM (2017)
22. Dong, X., et al.: Asymmetrical hierarchical networks with attentive interactions for interpretable review-based recommendation. In :The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, 7–12 February 2020, pp. 7667–7674. AAAI Press (2020)
23. Kim, D.H., Park, C., Oh, J., Lee, S., Yu, H.: Convolutional matrix factorization for document context-aware recommendation. In: Sen, S., Geyer, W., Freyne, J., Castells, P. (eds.), Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, 15–19 September 2016, pp. 233–240. ACM (2016)
24. Chin, J.Y., Zhao, K., Joty, S.R., Cong, G.: ANR: aspect-based neural recommender. In: Cuzzocrea, A., et al. (eds.) Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, 22–26 October 2018, pp. 147–156. ACM (2018)