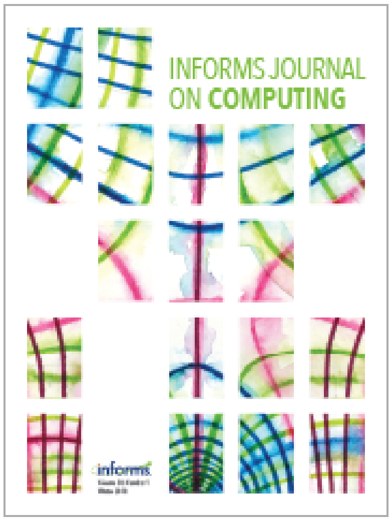


INFORMS Journal on Computing

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

From Interaction to Prediction: A Multi-Interactive Attention-Based Approach to Product Rating Prediction

Li Yu; , Wei Gong; , Dongsong Zhang; , Yu Ding; , Zhe Fu

informs
https://pubsonline.informs.org/journal/ijoc

INFORMS JOURNAL ON COMPUTING
Articles in Advance, pp. 1–15
ISSN 1091-9856 (print), ISSN 1526-5528 (online)

From Interaction to Prediction: A Multi-Interactive Attention-Based Approach to Product Rating Prediction

Li Yu,[a] Wei Gong,[a] Dongsong Zhang,[b,*] Yu Ding,[a] Zhe Fu[c]

[a] School of Information, Renmin University of China, Beijing 100872, China; [b] Belk College of Business & The School of Data Science, The University of North Carolina at Charlotte, Charlotte, North Carolina 28223; [c] College of Computing and Informatics, The University of North Carolina at Charlotte, Charlotte, North Carolina 28223

*Corresponding author

Contact: buaayuli@ruc.edu.cn, https://orcid.org/0000-0001-8503-2535 (LY); gongwei808@hotmail.com (WG); dzhang15@uncc.edu, https://orcid.org/0000-0001-5406-4681 (DZ); dinggyu@163.com (YD); zf2@uncc.edu (ZF)

Received: April 24, 2023
Revised: February 10, 2024; August 4, 2024; November 9, 2024; November 24, 2024
Accepted: November 26, 2024
Published Online in Articles in Advance: January 17, 2025

https://doi.org/10.1287/ijoc.2023.0131

Abstract. Despite increasing research on product rating prediction, very few studies have considered user-item interaction relationships at multiple levels. To address this critical limitation, we propose a novel rating prediction method based on multi-interaction attention (RPMIA) by learning user-item interaction relationships at three levels simultaneously from online consumer reviews for predicting product ratings with reasonable interpretability. Specifically, RPMIA first deploys a multihead cross-attention mechanism to capture the interaction between contexts of items and users. Then, it uses a bilayer gate-based mechanism to extract the aspects of items and users and a self-attention mechanism to learn their interaction at the aspect level. Finally, the aspects of users and items are coupled together to form meaningful user-item aspect pairs via a joint attention. A multitask predictor that integrates a factorization machine and a feedforward neural network is designed to generate a rating prediction. We empirically evaluated RPMIA with seven real-world data sets. The results demonstrate that RPMIA outperforms the state-of-the-art methods consistently and significantly. We also conduct a user study to assess the interpretability of the RPMIA method.

History: Accepted by Ram Ramesh, Area Editor for Data Science & Machine Learning.
Funding: The research is supported by Beijing Social Science Foundation [24XCB012], Suzhou Key Laboratory of Artificial Intelligence and Social Governance Technologies [SZS2023007], and Smart Social Governance Technology and Innovative Application Platform [YZCXPT2023101].
Supplemental Material: The software that supports the findings of this study is available within the paper and its Supplemental Information (https://pubsonline.informs.org/doi/suppl/10.1287/ijoc.2023.0131) as well as from the IJOC GitHub software repository (https://github.com/INFORMSJoC/2023.0131). The complete IJOC Software and Data Repository is available at https://informsjoc.github.io/.

Keywords: rating prediction • user-item interaction • recommender systems • online consumer reviews • deep learning

1. Introduction

With the explosive growth of e-commerce, consumers are frequently overwhelmed by the massive amount of product information available online (Markopoulos et al. 2016). Amazon, one of the world's largest and most well-known e-commerce companies that accounts for nearly 40% of the U.S. e-commerce market in 2023, lists more than 600 million products on its website,[1] providing rich product options but also challenges to consumers when they shop online. To cope with the information overload problem, numerous recommender systems have been developed and deployed for recommending useful products or information to boost business and facilitate consumers' purchase decisions (Adomavicius et al. 2022). Product rating prediction is one of the most crucial tasks in recommender systems that is aimed to predict the ratings of products for an individual consumer so that the products with the highest predicted ratings can be recommended to the consumer (Siering et al. 2018, Zhang et al. 2021).

In recent years, online consumer reviews (OCRs), including product ratings, have become a critical channel for consumers to share their experience with purchased products, gather information about certain products of interest, and make informed purchase decisions (Shin et al. 2020, Guo et al. 2021, Liu et al. 2021). Traditional product rating prediction models for recommender systems primarily relied on product ratings (e.g., collaborative filtering-based recommender systems). They have ignored the context of giving those ratings, which may lead to inaccurate understanding of user preferences. Because OCRs include rich contextual information about both users and items, analyzing the OCRs of the same target item and of the same target users by a product rating prediction model should be able to help the model achieve a more comprehensive understanding of user preference and item characteristics. Some previous studies have proposed methods for making rating predictions based on OCRs (Li et al. 2019, Wang et al. 2021, Zhu et al. 2022, Yu et al. 2023). Those methods can be classified into two categories based on the types of modeling techniques used. The first category is the topic-based method, which employs topic modeling techniques (e.g., latent Dirichlet allocation (LDA)) to model the text of OCRs (Hernández-Rubio et al. 2019). However, it only focuses on the content topic of OCRs while neglecting other key information in OCRs, such as consumer emotions and product aspects (i.e., product features like the color and weight). The second category is the aspect-based method, aiming to extract aspect-related sentiments from OCRs and use them to infer user interests (Hernández-Rubio et al. 2019, Yu et al. 2023). The term "aspects" refers to features, attributes, or dimensions of a product that consumers comment in their reviews, such as lens and battery life of a digital camera or features of a consumer. More recently, increasing studies start exploring deep learning techniques for extracting aspect information from OCRs (Li et al. 2019, Wang et al. 2021, Zhu et al. 2022). For example, Zhu et al. (2022) employed a bidirectional recurrent neural network and a self-attention[2] mechanism to extract aspects from OCRs. The extracted aspects are used to model users and represent items for the rating prediction.

Although some existing studies on rating prediction based on OCRs have already considered aspects in OCRs (Li et al. 2019), very few have delved into the interactions between users and items at a more fine-grained level that can provide more detailed cues about preferences of a target user for product aspects, the perceptions of users about individual aspects of a product, and the association between the user preferences and item aspects so as to improve rating predictions. Specifically, existing studies have the following major limitations. First, few of them distinguish the importance of different contexts and model the interactions between users and items from multiple views at a context level. In this study, the term *context* is referred to as the circumstances or local information of a term (e.g., a number of words before and after a target term) in a review. Existing studies usually extract contexts from reviews through a convolutional neural network (Wu et al. 2019, Wang et al. 2021). Those studies do not consider interaction between users and items at a context level from multiple views, which cannot distinguish the importance of contexts. Second, not all terms in a review are equally important or even relevant to an item aspect. In addition, the same term may have considerably different meanings when the term is associated with different item aspects. For example, when the same term "long" is used to describe the aspect "standby time" of a smartphone and the aspect "boot time" of a PC in two reviews, it represents very different sentiments of users toward those two aspects. The term *sentiments* refers to the emotional tones or attitudes toward products and services expressed in reviews. As a result, a rating prediction model should consider the context of the commented aspects in a review when modeling user representation, referred to as user-item interactions at an aspect level in this paper, which has been largely ignored in current research. Third, existing studies do not fully couple aspects of users and items and distinguish the varying importance of coupled user-aspect and item-aspect pairs to rating prediction. Considering ratings should reflect the user's opinions on item aspects expressed in reviews, it is important that a rating prediction model couples a user aspect and an item aspect to better represent the associations between them, which has rarely been considered in literature.

Motivated to address the above-mentioned limitations of existing research, this study aims to answer the following three overarching research questions. (1) How to identify and extract key information from OCRs at multiple levels? (2) How to learn the interaction relationship between consumers and items from OCRs at multiple levels? (3) Can incorporating such interaction relationships into a rating prediction model improve its performance? To answer these research questions, in this paper, we propose a rating prediction model based on multi-interactive attention (RPMIA) for learning the underlying reason why a consumer gives a certain rating to a product from OCRs by learning the interaction relationships between consumer-specific reviews and product-specific reviews at three levels, including the context, aspect, and aspect pair levels. RPMIA consists of five modules, including an embedding module, a context-aware module, an aspect-aware module, a user-item aspect pairs (UIAP)-aware module, and a prediction module. Specifically, the embedding module generates initial embedding representations of consumers and products. Then, the context-aware module employs a multihead cross-attention method to learn interaction relationships between consumers and products at a context level, and the aspect-aware module extracts and learns the interaction relationships at an aspect level through a bilayer gating-based mechanism and a self-attention mechanism. Furthermore, we propose the UIAP-aware module to further capture the interaction relationships through UIAP identified by an attention mechanism. Finally, a multitask predictor that integrates a factorization machine and a feedforward neural network is used to predict ratings in the prediction module. The empirical evaluation shows superior performance of RPMIA in rating prediction in comparison with state-of-the-art baseline models.

2. Literature Review

2.1. Online Consumer Reviews

There have been extensive studies on user[3] behavior analysis (Sanchez-Loor and Chang 2023, Zhang et al. 2023), review helpfulness (Hong et al. 2017), and rating predictions based on OCRs (Khan and Niu 2021, Yu et al. 2023).

Some deep neural networks, such as convolutional neural networks (CNN) and recurrent neural networks (RNN), have been used to automatically learn the latent features of an item and the preference of a user from review text for generating rating predictions (Wu et al. 2021). Khan and Niu (2021) proposed a novel CNN based model with depthwise separable convolutions for rating prediction based on product reviews. However, existing studies usually generate rating predictions based on the similarity between user representations and item representations, and none of them has considered interaction relationships between users and items at multiple levels when generating rating predictions.

2.2. Review-Aware Recommender Systems

OCRs, particularly the textual content of OCRs, have been exploited by many researchers for product recommendation. Some previous research has employed topic modeling techniques, such as LDA and nonnegative matrix factorization (NMF), to learn topics from OCRs (Hernández-Rubio et al. 2019, Yu et al. 2023). The term *topics* refers to as the central subjects or themes of reviews. Despite encouraging performance, those topic modeling methods ignore important context of terms in reviews. Some other studies make rating predictions based on product aspects commented in reviews. Chen et al. (2015) surveyed how consumer reviews have been exploited to improve product recommendations. Hernández-Rubio et al. (2019) investigated three classes of aspect extracting methods from reviews, including the vocabulary-based method, the double propagation method, and LDA method. The vocabulary-based method leveraged an aspect vocabulary, whereas a social network propagating method was used for the double propagation.

Recently, there has been emerging use of deep learning techniques to extract key information (e.g., context, aspect) of users and items from reviews in review-aware recommender systems. Da'U et al. (2021) employed a Bidirectional Long Short-Term Memory and a coattention network to learn user/item representations from review texts, which were integrated with the transaction data for product recommendation. Musto et al. (2017) extracted aspects from OCR based on sentiments. Kim et al. (2016) proposed a convolutional matrix factorization (ConvMF) method that employed a CNN to learn the representation of contextual information from review text. Wang et al. (2021) used two parallel CNNs to learn the context vectors of users and items from review text, respectively. Wu et al. (2019) proposed a context-aware model based on review text by implementing an attention layer after convolutional layers to represent users and items. Li et al. (2019) extracted user viewpoints and/or item aspects from OCRs to model users and represent items for providing recommendations. Liu et al. (2021) proposed a multilingual aspect-based sentiment analysis method to extract aspects and their associated sentiments simultaneously. Nevertheless, most existing research on review-based recommender systems consider product contexts and aspects discussed in reviews equally important to product ratings, which may not often be the case. In addition, existing studies do not consider interaction relationships between users and items at the context and aspect levels. As a result, existing studies are unable to model users' interest at a fine and sufficient granularity in order to produce accurate rating predictions.

2.3. Attention-Based Recommendation

The attention mechanism has been widely used in deep learning to allow various parts of input to contribute to model output at different levels (Zeng et al. 2021, Rashed et al. 2022). Attention-based recommendation can be categorized into two types: item-level attention and feature-level attention. The item-level attention has been used in both graph convolution network-based recommender systems (Lu et al. 2018) and session-based recommender systems (Yuan and Wu 2019) to differentiate the contributions of neighboring nodes to the user/item embedding learning process. For example, Lu et al. (2018) proposed a coevolutionary recommendation model that learned user and item representations jointly from ratings and review text by a gated recurrent unit (GRU) network using the attention mechanism. The feature-level attention uses the attention mechanism to learn the contributions of user and item representations. For example, Yuan and Wu (2019) proposed a user/item representation learning model through CNN to highlight semantic information in text relevant to users or items.

In summary, despite increasing research on product rating predictions based on OCRs, existing models have not considered deep interaction relationships between users and items at context, aspect, and aspect-pair levels simultaneously when making rating predictions, nor considered the varying importance of different contexts, aspects, and aspect pairs for modeling user preferences, nor empirically explored the interpretability of deep learning-based rating prediction models. As a result, the model performance may be suboptimal.

3. Research Methodology

To address the limitations of existing studies, we propose an RPMIA for rating prediction.

3.1. Problem Formalization and Concepts

Table 1 presents some major symbols and their descriptions. The research problem in this study can be defined as follows: given a group of users $U = \{u_1, u_2, \dots\}$ and an item (i.e., product) set $V = \{v_1, v_2, \dots\}$, RPMIA aims to predict the rating $\hat{r}_{u,v}$ of a user u toward an item v. The candidate items are then ranked in a descending order of $\hat{r}$, and the top K items with the highest predicted ratings will be recommended to user u.

Let's first define the key concepts and symbols associated with RPMIA as follows:

Definition 1 (User-Specific Review Corpus). A user-specific review corpus (URC) is a collection of all reviews written by a user. We denote all reviews written by user u as URC_u.

Definition 2 (Item-Specific Review Corpus). An item-specific review corpus (IRC) is a collection of all reviews of an item written by all users. We denote the collection of all reviews of an item v written by all users as IRC_v.

Definition 3 (Context). A context is denoted as the circumstances or local information of a term at a certain position in a review. We refer to the contexts extracted from URC as user contexts and contexts extracted from IRC as item contexts.

Definition 4 (Aspect). An aspect is referred to as a feature, attribute, or dimension of a product that consumers comment on in their reviews.

Definition 5 (UIAP). A UIAP is a joint unit that combines a user's aspect and an item's aspect.

3.2. Key Design Rationales of RPMIA

There are several key design rationales of the proposed RPMIA method:

i. To improve the performance of rating prediction, it is necessary to address the lack of considerations of the fine-grained multilevel interactions in existing literature.

ii. The contexts extracted from URC may contribute to individual items differently. Similarly, the counterparts from IRC may also have different levels of relevance and importance for different users. Therefore, it is necessary to consider the varying importance of contexts for predicting the rating of a user for the target item.

iii. Not all aspects commented in reviews contribute to the user/item representation equally. Therefore, it is important for a model to identify and give more attention to the important aspects for rating prediction.

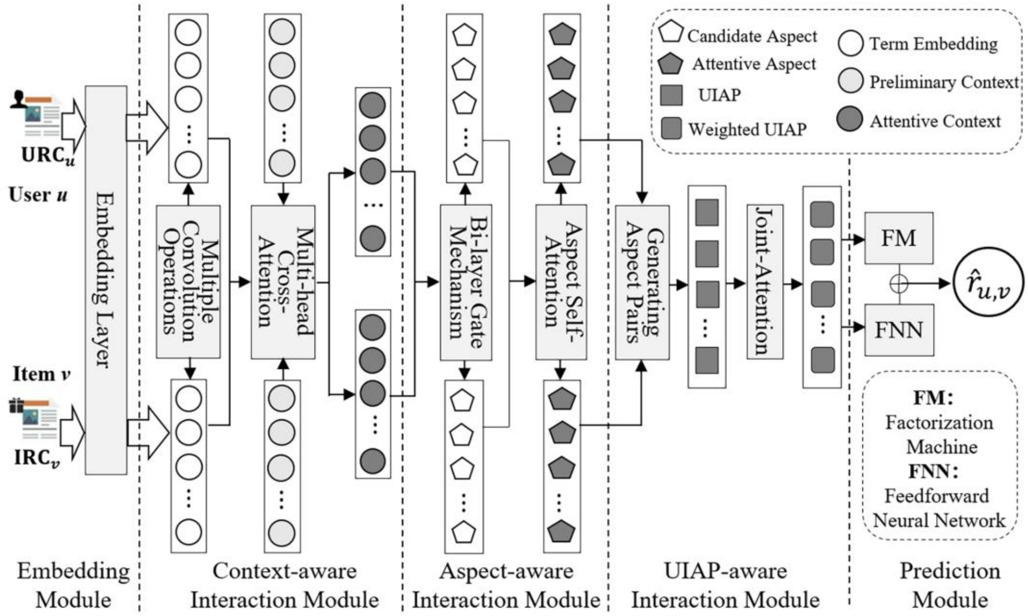
iv. Because users and items usually have a joint effect on rating predictions, it can be helpful to consider UIAPs when coupling interaction relationships.

Based on the above design rationales, we propose the RPMIA model shown in Figure 1. The model consists of an embedding module, a context-aware module, an aspect-aware module, a UIAP-aware module, and a prediction module. Guided by the second design rationale, the context-aware interaction module is designed for learning the interaction relationship between users and items at a context level, which is extracted by multiple convolution operations. By following the third design rationale, the aspect-aware interaction module generates the aspects of a user and an item from their contexts via two steps. This module uses a gating mechanism to get preliminary aspects from the attentive contexts, then deploys a self-attention mechanism to derive attentive aspects from preliminary aspects. Following the fourth rationale, the UIAP-aware module captures the joint effect of the aspects of users and items through a joint-attention mechanism, which learns the influence of different UIAPs on rating prediction. Finally, the prediction module deploys a factorization machine and a feedforward neural network to jointly predict a rating. We gave examples for illustrating the term context, aspects, and UIAPs in Online Appendix 1.

Table 1. Important Symbols and Their Meanings

Symbols	Meanings
u (v)	The user u (the item v)
U (V)	The set of users (items)
URC_u	The merged all reviews of all items written by user u
IRC_v	The merged all reviews of item v written by all users
N (M)	The number of terms in URC_u (IRC_v) after preprocessing
D_u (D_v)	The term vector matrix of user u (item v)
$\tilde{C}^u$ ($\tilde{C}^v$)	The preliminary context representation of user u (item v) extracted from D_u (D_v)
C^u (C^v)	The attentive context representation of user u (item v) extracted from D_u (D_v)
A_q^v (A_p^u)	The qth (pth) attentive aspect of item v (user u)

Figure 1. (Color online) Architecture of RPMIA



3.3. Embedding Module

The embedding module is used to generate initial embedding representations of URC_u of user u and IRC_v of item v as the input of the RPMIA model. The term t_i in a review corpus (i.e., URC_u and IRC_v) is firstly represented by its corresponding embedding e_i. In this research, we employ a pretrained Word2Vec model to obtain a term vector. Then we obtain the term matrices D_u from URC_u and D_v from IRC_v by concatenating the embeddings of terms appeared in URC_u and IRC_v as follows:

$$D_x = [e_1, e_2, \ldots, e_i, ..], \tag{1}$$

where x denotes user u or item v, and e_i denotes the embedding representation of the ith term in D_u or D_v.

3.4. Context-Aware Interaction Module

User reviews and item descriptions provide diverse and complementary contextual information about user preferences and item features. Different parts of a review text may provide varying levels of relevance to different aspects of an item. For example, a sentence praising the battery life of a smartphone contributes specifically to the "battery life" aspect, whereas another sentence about the camera resolution contributes to that aspect. Extracting context helps identify which parts of the text are pertinent to different aspects. By capturing the contextual relevance, RPMIA can understand how different contexts may influence user preferences and item characteristics. Existing methods of extracting review context usually employ a CNN or an attention-based CNN, which cannot capture the contextual relevance from different views. In the context module, we employ a multi-head cross-attention mechanism with CNN-generated contexts, which allows the model to differentiate and appropriately weigh the contributions of contexts to users' viewpoints and items' aspects and ensures that the model leverages the full spectrum of contextual data available, which is crucial for accurate rating predictions.

After that, the context-aware interaction module performs the convolution operation with a sliding window of size T to extract local contexts $\tilde{C}^u_{i,f}$ from the term vector matrix D_u of URC_u, as shown in Equation (2):

$$\tilde{C}^u_{i,f} = \Theta_f(D_{u,i:i+T-1}) = \sigma(W^f * D_{u,i:i+T-1} + b_f), \tag{2}$$

where W^f denotes the convolution weight vector for the convolution operation Θ_f; $D_{u,i:i+L-1}$ is term vector matrix D_u with a sliding window starting at the ith term of a review; b_f is the bias; and σ is a nonlinear activation function. In this study, we employ *ReLU* as the activation function. We use multiple convolution filters to perform multiple convolution operations in order to capture different contexts of D_u from multiple views. Therefore, the module gets the preliminary context $\tilde{C}^u_i$ of the ith term in D_u by concatenating the local contexts generated by all convolution operations.

Using the same method, the module obtains the preliminary context representation $\tilde{C}_j^v$ of the jth term in D_v. Once the module gets the preliminary context representation of all terms in URC_u and IRC_v, it generates the preliminary context matrices of user u and item v, denoted as $\tilde{C}^u$ and $\tilde{C}^v$, respectively. After obtaining the preliminary contexts of user u and item v, we designed a multihead cross-attention mechanism to learn the interaction relationship between the context matrix $\tilde{C}^u$ of the user u and the item context matrix $\tilde{C}^v$ of item v in order to derive the weight importance of contexts. Specifically, we first calculate the average vector of all preliminary contexts of user u and item v as their overall contexts, denoted as $\overline{\tilde{C}^{(h)u}}$ and $\overline{\tilde{C}^{(h)v}}$, respectively. Then, the context-aware interaction module calculates the affinity $\lambda_k^{(h)u}$ (and $\lambda_j^{(h)v}$) of each preliminary context of user u (and item v) with the overall context of item v (user u) for the hth head cross-attention as follows:

$$\lambda_k^{(h)u} = tanh(W_u^{(h)}[\tilde{C}_k^u; \overline{\tilde{C}^v}] + b_u^{(h)}), \quad \lambda_j^{(h)v} = tanh(W_v^{(h)}[\tilde{C}_j^v; \overline{\tilde{C}^u}] + b_v^{(h)}) \tag{3}$$

where $tanh$ is the tangent activation function; $W_u^{(h)}$ and $W_v^{(h)}$ are the learned transformation matrices; and $b_u^{(h)}$ and $b_v^{(h)}$ are bias vectors. Based on the affinity relationships, the context-aware interaction module calculates the attentive weight vectors $\alpha_k^{(h)u}$ for $\tilde{C}_k^{(h)u}$ and $\alpha_j^{(h)v}$ for $\tilde{C}_j^{(h)v}$, respectively, as follows:

$$\alpha_k^{(h)u} = \exp(\lambda_k^{(h)u}) / \sum_{i=1}^{M} \exp(\lambda_i^{(h)u}), \quad \alpha_j^{(h)v} = \exp(\lambda_j^{(h)v}) / \sum_{j=1}^{N} \exp(\lambda_j^{(h)v}). \tag{4}$$

Furthermore, we multiply $\tilde{C}_k^{(h)u}$ and $\tilde{C}_j^{(h)v}$ by the weight vectors $\alpha_k^{(h)u}$ and $\alpha_j^{(h)v}$ to obtain the attentive context $C_k^{(h)u}$ and the attentive context $C_j^{(h)v}$ for the cross-attention h as follows:

$$C_k^{(h)u} = \alpha_k^{(h)u} \tilde{C}_k^{(h)u}, \quad C_j^{(h)v} = \alpha_j^{(h)v} \tilde{C}_j^{(h)v}. \tag{5}$$

We concatenate the attentive contexts generated by all head cross-attentions to obtain the final attentive contexts as follows:

$$C_k^u = C_k^{(1)u} \| \dots \| C_k^{(h)u} \| \dots \| C_k^{(H)u}, \ C_j^v = C_j^{(1)v} \| \dots \| C_j^{(h)v} \| \dots \| C_j^{(H)v}, \tag{6}$$

where C_k^u is the kth attentive context vector of user u; C_j^v is the jth attentive context vector of item v, and H is the number of heads of cross-attention. Finally, the representations C^u of user u and C^v of item v at a context level are obtained as follows:

$$C^u = [C_1^u, C_2^u, \dots, C_k^u, \dots, C_N^u], \ C^v = [C_1^v, C_2^v, \dots, C_j^v, \dots, C_M^v]. \tag{7}$$

3.5. Aspect-Aware Interaction Module

User reviews reflect personal experiences and preferences, whereas item descriptions focus on product features and qualities. In general, it can be well expected that users would be more likely to comment on individual features of a product in OCRs that are more important to them. Therefore, being aware of which aspects of products that a user comments in her reviews would be very valuable to the understanding of her preferences in product aspects and making more accurate ration predictions. Existing methods of extracting aspects from reviews usually rely on topic modeling, which cannot identify valuable aspects on the rating prediction by learning deep association relationships between aspects and ratings. In the aspect-aware interaction module, we design a bilayer gate-based mechanism and a self-attention mechanism to distill the most relevant aspects from contexts, highlighting the relative importance of different aspects.

The aspect-aware interaction module first employs an aspect-specific gating mechanism to extract the influence R_p of the pth aspect A_p of all users, as shown in Equation (8):

$$R_p = \delta(H_p \cdot A_p + b_p), \tag{8}$$

where A_p is the learned embedding of the pth aspect shared by all users; H_p is the transformation matrice of A_p^u; $b_p \in \mathbb{R}^d$ is a bias vector; δ is the sigmoid activation function; and $\cdot$ denotes the inner product operation. By using a general gating mechanism, the module obtains the new embedding representation of C_p^u, which is considered as the candidate aspect, denoted as $\tilde{A}_{i,p}^u$, as shown in Equation (9):

$$\tilde{A}_{i,p}^u = C_i^u \cdot \delta(R_p W_i \cdot C_i^u + b_i'), \tag{9}$$

where W_i is the transformation matrice of C_i^u; and b_i' is the bias vector.

Once the aspect-aware interaction module extracts preliminary aspects from contexts through the bilayer gating mechanism, it generates final aspect representation through a self-attention mechanism. To distinguish those generated aspects from those initially extracted from contexts through self-attention, we refer to the former as attentive aspects. Specifically, this module derives a rudimentary representation $\overline{A}_p^u$ of the pth aspect of user u by taking the average of the aspect-specific embedding $\tilde{A}_{i,p}^u$ of each context C_i^u across all contexts of user u. Then the weighted sum of $\tilde{A}_{i,p}^u$ and the corresponding self-attention weight vector $\beta_{i,p}^u$ are used to get the pth aspect A_p^u of user u as follows:

$$\overline{A}_p^u = \frac{1}{N}\sum_{i=1}^{N}\tilde{A}_{i,p}^u, \quad \beta_{i,p}^u = softmax(\tilde{A}_{i,p}^u \overline{A}_p^u), \quad A_p^u = \sum_{i=1}^{N}\beta_{i,p}^u \tilde{A}_{i,p}^u. \tag{10}$$

Similar to extracting the aspects of a user, the aspect-aware interaction module also extracts the aspects of an item from the corresponding attentive contexts and represents the qth aspect of item v as A_q^v. For model simplicity, we followed the study of Adomavicius et al. (2019) by setting the maximum number of both extracted user aspects and extracted item aspects to five.

3.6. UIAP-Aware Interaction Module

In the real world, a user may have different opinions on different aspects of the same item, and different aspects may be of different levels of importance to the same users. For example, a user's emphasis on item durability and an item's build quality form a meaningful aspect pair. To reflect the joint effect of a user's viewpoints and an item's aspects on the rating prediction, in this module, we propose a concept called UIAP and employ a joint attention to couple the extracted aspects of users and items, forming meaningful aspect pairs. This approach allows RPMIA to focus on specific user preferences and corresponding item characteristics that jointly influence a rating. To the best of our knowledge, no existing studies have ever incorporated such UIAPs in their models.

Specifically, given the aspects A_p^u and A_q^v, we get the representation of the corresponding aspect pair based on Equation (11), denoted as $\tilde{Z}_{p,q}^{u,v}$:

$$\tilde{Z}_{p,q}^{u,v} = (A_p^u \oplus A_q^v)\|(A_p^u \odot A_q^v)\|(A_p^u \otimes A_q^v), \tag{11}$$

where $\otimes$ denotes the cross-product operation; $\odot$ denotes the dot product operation; $\oplus$ denotes the vector addition operation; and $\|$ represents the vector concatenation operation. If we consider A_p^u and A_q^v as the first-order aspects that influence ratings, then $\tilde{Z}_{p,q}^{u,v}$ can be treated as a second-order aspect.

Considering that different UIAPs may have different influences on a rating, the UIAP-aware interaction module employs a self-attention mechanism to learn the weights of individual UIAPs. Specifically, we design a joint-attention mechanism to obtain weighted UIAP, which can be obtained as follows:

$$\overline{\tilde{Z}}^{u,v} = \frac{1}{K^2}\sum_{(p,q)\in\tilde{Z}^{u,v}}\tilde{Z}_{p,q}^{u,v}, \quad \gamma_{p,q}^{u,v} = softmax(\tilde{Z}_{p,q}^{u,v^T}\overline{\tilde{Z}}^{u,v}), \tag{12}$$

$$Z_{p,q}^{u,v} = \gamma_{p,q}^{u,v}\tilde{Z}_{p,q}^{u,v}, \quad Z^{u,v} = \|_{p=1,q=1}^{K,K} Z_{p,q}^{u,v}, \tag{13}$$

where K is the number of aspects of each user and item; $Z_{p,q}^{u,v}$ denotes the attentive UIAP derived by the aspect A_p^u of user u and the aspect A_q^v of item v; $\gamma_{p,q}^{u,v}$ is the obtained weight of $Z_{p,q}^{u,v}$; and $Z^{u,v}$ is the obtained joint UIAP vector representation for user u and items v, which contains aggregated information of all UIAPs associated with user u and item v for inferring a final rating.

3.7. Prediction Module

The prediction module employs the widely used wide and deep learning method (Cheng et al. 2016) to predict the rating $\hat{r}_{u,v}$ of user u for item v as follows:

$$\hat{r}_{u,v} = \delta(W_{wide}\Phi(Z^{u,v}) + W_{deep}\Psi(Z^{u,v}) + b_r), \tag{14}$$

where δ is the sigmoid function, Φ denotes the prediction component of the factorization machine (FM), and Ψ denotes the prediction component of the feedforward neural network (FNN); W_{wide} is the weight vector of the FM model, and W_{deep} is the weight vector of the FNN model. b_r is bias. As shown in Equation (14), FM is used to capture the direct relationship between UIAPs and the rating value, whereas FNN is used to model their deep association relationships:

$$\Phi(X) = m_0 + m^T X + \frac{1}{2}X^T M X, \tag{15}$$

where Φ denotes FNN, and X is UIAP representation ($Z^{u,v}$) of user u and item v in this study; m_0 is the global bias, m is the coefficient vector for its latent feature vector, and M is the learned weight matrix for second-order interaction of the latent feature vector:

$$\Psi(X) = X^{(L)}, \quad X^{(l+1)} = f(W^{(l)} X^l + b^{(l)}) \ (l = 1, 2, \ldots, L), \tag{16}$$

where Ψ denotes FNN, L is the total number of layers of FNN, l is the lth layer, and f is the activation function (rectified linear units (ReLUs)); $W^{(l)}$, $X^{(l)}$, and $b^{(l)}$ are the learned weights, UIAP vectors, and bias at the lth layer, respectively.

3.8. Model Training

We deploy the square loss as the objective function (J_{loss}) of the RPMIA model for parameter optimization, as shown in Equation (17):

$$J_{loss} = \sum_{r_{u,v} \in R} (r_{u,v} - \hat{r}_{u,v})^2 + \lambda_\Theta ||\Theta||^2, \tag{17}$$

where R is the user-item rating matrix; $r_{u,v}$ and $\hat{r}_{u,v}$ are the actual and estimated ratings of user u on item v, respectively; and $\lambda_\Theta ||\Theta||^2$ represents regularization. By using stochastic gradient descent (SGD) and back-propagation, the parameters of RPMIA are optimized based on Equation (17). For parameter updates, RPMIA utilizes RMSprop over minibatches. In addition, RPMIA deploys the dropout strategy to prevent overfitting. After obtaining parameters that minimize the loss function, RPMIA generates the predicted rating scores.

We used the grid search method to obtain optimal hyperparameter values. Specifically, the batch size was set to 256; the learning rate was 10^{-3}; the L_2 coefficient was 10^{-6}; the dropout rate was 0.2; the dimensions of both word vectors and latent feature vectors were 50; the number of convolution kernels was 50; and the size of sliding windows in CNN was three.

4. Empirical Evaluation

We conduct an empirical evaluation of the RPMIA model to assess the overall performance of RPMIA in rating prediction, compared it with the performance of state-of-the-art baseline models, explored the effects of individual design artifacts on RPMIA performance, and investigated the interpretability of RPMIA. The code and data are available at a Github repository (Yu et al. 2024).

4.1. Data Sets

We used seven publicly available OCR data sets collected from Amazon.com on different products, including music instruments, office products, digital music, grocery & gourmet food, video games, tools & home improvement, and sports & outdoors products, which have been widely used for recommendation evaluation in previous studies (Adomavicius et al. 2019). These data sets consist of consumers' product ratings ranging from one to five and corresponding textual reviews. Each consumer or item has at least five reviews. Table 2 presents the descriptive statistics of those data sets.

Similar to the study of Kim et al. (2016), we processed the reviews in those data sets as follows: (1) set the maximum length of reviews to 300; (2) removed stop words; (3) selected the top 20,000 words based on their tf-idf (i.e., term frequency-inverse document frequency) values; and (4) removed all nonvocabulary words from reviews. We randomly selected 70% of data samples from each data set as the training set, 10% for validation, and the remaining 20% as the testing set. In the training set, each product was rated and reviewed by at least one consumer, and each consumer rated and reviewed at least one product.

Table 2. Descriptive Statistics of the Data Sets

Data set	No. of users	No. of items	No. of ratings	No. of words per review	No. of words per user	No. of words per item	Data density
Musical instruments	1,429	900	10,261	32.45	141.32	200.12	0.798%
Office products	4,905	2,420	53,228	48.15	197.93	229.52	0.448%
Digital music	5.540	3,568	64,666	69.57	216.21	266.51	0.327%
Grocery & gourmet food	14,681	8,713	151,254	57.34	177.24	155	0.118%
Video games	24,303	10,672	231,577	72.13	188.79	260.60	0.089%
Tools & home improvement	16,638	10,217	134,345	38.75	162.53	212.48	0.079%
Sports & outdoors	35,598	18,357	296,337	37.59	159.39	154	0.045%

4.2. Baseline Methods

To assess the effectiveness of the proposed RPMIA model, we compared its prediction performance with those of the following state-of-the-art OCR-based recommendation models:

- *DeepCoNN* (Wang et al. 2021) is a deep collaborative neural network that uses two parallel CNNs to learn the latent feature vectors of users and items from user and item reviews, respectively. The learned user and item representations are concatenated and then fed into the FM for rating prediction.
- *D-attn* (Seo et al. 2017) is a convolutional neural network with local and global attentions for rating predictions, enabling an interpretable representation of users and items.
- *CARL* (Wu et al. 2019) is a context-aware user-item representation learning model based on review text that implements an attention layer after convolutional layers.
- *ANR* (Chin et al. 2018) is an aspect-based neural recommender system that deploys a coattention mechanism to learn the representation of aspects of users and items.
- *CARP* (Li et al. 2019) employs a capsule network involving to extract user viewpoints and item aspects and then uses a sentiment capsule architecture with a routing by a bi-agreement mechanism to identify informative logic units for rating prediction.
- *HUARN* (Li et al. 2018) is a method for aspect-aware sentiment analysis that aims to predict a user's sentiment polarities for different aspects of a product in a review, utilizing user attention and aspect attention to learn sentence- and document-level representations. We adapted the last layer of the original model for rating predictions.
- *TFRMF* (Zhu et al. 2022) is a deep learning model based on a bi-directional RNN and a self-attention mechanism to extract user preferences and item attributes from reviews. It then feeds the latent representation vectors of users and items into the matrix factorization model.
- *SentiAttn* (Wang et al. 2022) includes both a sentiment attention mechanism that helps identify user reviews that are most likely to enhance the accuracy of rating prediction and a global attention mechanism that captures the importance of different parts of reviews.

4.3. Evaluation Metrics

Similar to the vast majority of current studies (Liu et al. 2021), we evaluated the performance of the RPMIA model and the baselines with three common evaluation metrics, including mean squared error (MSE), intralist similarity (ILS), and novelty. MSE measures the difference between a model's predicted ratings and real ratings (Wu et al. 2019). The smaller the MSE value, the better the prediction performance of a model. In this study, we calculate MSE as the average MSE of predicted ratings for all items in the testing data set as follows:

$$MSE = \frac{1}{|O_{test}|} \sum_{(u,v) \in O_{test}} (r_{u,v} - \hat{r}_{u,v})^2, \tag{18}$$

where $\hat{r}_{u,v}$ denotes the predicted rating; $r_{u,v}$ is the actual rating; and O_{test} denotes the testing samples. ILS measures the diversity of the recommended items, which denotes how different the recommended items are. It is calculated as follows:

$$ILS = \sum_{u \in U} \frac{2}{|R_u|^2 - 1} \sum_{v_i, v_j \in R_u, i \neq j} d(v_i, v_j), \tag{19}$$

where R_u denotes a ranked recommendation list of user u; $|R_u|$ denotes the length of R_u; and $d(v_i, v_j)$ denotes the distance between recommended items v_i and v_j in R_u, which is calculated by the cosine similarity between the representations of items v_i and v_j.

Novelty measures the extent to which an item is new compared with those items that have been rated by user u, which can be calculated as follows:

$$Novelty = \sum_{u \in U} \sum_{v_i \in R_u} \min_{v_j \in V, i \neq j} d(v_i, v_j), \tag{20}$$

where v_i denotes an item in a recommendation list R_u, and v_j denotes an item in an item set V that user u has rated.

4.4. Overall Performance

We take the average performance of a model with all testing samples over 10 runs as the performance of that model. The overall performances of all models in MSE, ILS, and novelty are shown in Tables 3, 4, and 5, respectively, where the best baseline model performance with each data set is highlighted in boldface. The length of the

Table 3. Comparison of MSE

Data sets	DeepCoNN	D-attn	CARL	ANR	CARP	HUARN	TFRMF	SentiAttn	RPMIA	Improvement (%)
Musical instruments	0.815	0.798	0.776	0.796	0.773	**0.655**	1.168	0.781	0.636	17.51**
Office products	0.759	0.758	0.742	0.724	0.719	**0.699**	0.732	0.721	0.688	4.19**
Digital music	0.942	0.910	0.831	0.867	**0.820**	0.823	0.900	0.935	0.811	0.87*
Grocery & gourmet food	0.979	0.952	0.794	0.822	**0.752**	0.755	1.036	0.834	0.747	0.59**
Video games	1.148	1.190	1.094	1.182	1.084	**0.976**	1.063	1.034	0.788	27.10**
Tools & home improvement	0.983	1.002	**0.949**	0.979	0.962	0.957	1.001	1.023	0.939	2.09*
Sports & outdoors	1.130	0.978	0.989	1.023	0.876	**0.873**	0.923	0.882	0.638	26.86**

$^{**}p < 0.01$; $^{*}p < 0.05$.

recommendation list is set to five. The last column in the tables shows the result of a two-tailed pairwise t-test between RPMIA and the best performed baseline model for each data set. As shown in Tables 3–5, RPMIA significantly and consistently outperforms all baseline models in all three performance measures and with all seven data sets ($p < 0.05$).

4.5. Effects of Parameters on Rating Prediction

In order to verify the hyperparameter settings, we investigated the effects of two key hyperparameters of RPMIA, including the dimension size of term representation vectors d_t and the number of extracted aspects K, on MSE of RPMIA using the testing data sets of music instruments and office products as examples by varying the values of those two variables while keeping other parameter values unchanged.

Figure 2 shows the MSE values of RPMIA with varied dimension sizes for the music instrument and office product data sets. It reveals that RPMIA achieves the optimal performance when the dimension size d_t is 50 for both data sets. Therefore, we choose 50 as the dimension size of the word vector in all our evaluation tasks. Because MSE is the most important performance metric among the three, we mainly focus on MSE in this analysis. In addition, there has not been a well-accepted guideline in literature for determining the optimal number of aspects that should be considered. Previous studies typically chose this number in an ad hoc way (often in the range of three to five) (Chin et al. 2018, Li et al. 2018, Yu et al. 2023). In this study, we empirically evaluated the performance of the proposed model with different numbers of aspects (i.e., three, five, seven, and nine). The results, as presented in Table 6, showed that the proposed model achieved the best prediction performance when it considered five aspects consistently across different products, highlighted in boldface. Thus, we selected five for our model.

4.6. Impacts of Core Interaction Modules on Prediction Performance

We further investigated the impact of each designed interaction module of RPMIA on the performance of RPMIA in rating prediction with the testing data sets. When evaluating a module, we compared the MSE of the ablated RPMIA model without the evaluated module versus the performance of the complete model. Table 7 shows that the MSE of the RPMIA model is consistently lower than the MSEs of any of those ablated models ($p < 0.01$) across all seven data sets, highlighted in boldface, suggesting that removing any of the proposed three interaction modules (i.e., context-, aspect-, and UIAP-aware interaction modules) increases the error in rating prediction. The results validate the positive impact of each individual module on model performance.

4.7. Interpretability Analysis

4.7.1. Example of Model Interpretability. This section is aimed to demonstrate how the ratings predicted by RPMIA can be interpreted. After feeding two rating examples of a certain user shown in Table 8, which are

Table 4. Comparison of ILS

Data sets	DeepCoNN	D-attn	CARL	ANR	CARP	HUARN	TFRMF	SentiAttn	RPMIA	Improvement
Musical instruments	0.454	0.572	0.534	0.608	0.620	0.631	0.623	**0.651**	0.691	10.53**
Office products	0.532	0.491	0.603	0.533	0.692	**0.694**	0.682	0.692	0.711	2.97**
Digital music	0.421	0.687	0.601	**0.691**	0.683	0.685	0.685	0.687	0.691	0.21*
Grocery & gourmet food	0.519	0.501	0.605	0.642	0.681	**0.693**	0.601	0.684	0.702	2.14**
Video games	0.391	0.423	0.538	0.601	0.623	0.662	0.594	**0.672**	0.683	1.28**
Tools & home improvement	0.492	0.501	0.499	0.562	0.581	**0.584**	0.532	0.582	0.609	11.11**
Sports & outdoors	0.432	0.410	0.492	0.502	0.531	**0.569**	0.521	0.562	0.574	2.27**

$^{**}p < 0.01$; $^{*}p < 0.05$.

Table 5. Comparison of Novelty

Data sets	DeepCoNN	D-attn	CARL	ANR	CARP	HUARN	TFRMF	SentiAttn	RPMIA	Improvement
Musical instruments	0.501	0.523	0.602	0.595	0.620	0.628	0.600	**0.630**	0.632	0.36*
Office products	0.491	0.531	0.499	0.525	0.553	0.579	0.548	**0.590**	0.594	3.34**
Digital music	0.531	0.515	0.572	0.569	0.591	**0.603**	0.552	0.600	0.625	10.29**
Grocery & gourmet Food	0.410	0.481	0.431	0.492	0.502	0.531	0.504	**0.529**	0.546	3.27**
Video games	0.514	0.532	0.531	0.562	0.564	**0.581**	0.553	0.575	0.609	8.33**
Tools & home improvement	0.491	0.412	0.462	0.473	0.501	**0.510**	0.500	0.498	0.514	0.41*
Sports & outdoors	0.481	0.483	0.495	0.504	0.511	**0.514**	0.499	0.510	0.527	4.53**

$^{**}p < 0.01$; $^{*}p < 0.05$.

randomly selected from URC, RPMIA generates the predicted ratings of the user u for the two items. The importance weight distribution of review terms in the cross-attention layer is shown in Equation (21):

$$w_p = \Omega\left(\frac{1}{H(2T+1)}\sum_{j=p-T}^{p+T}\sum_{h=1}^{H}\alpha_j^{(h)}\right), \tag{21}$$

where w_p denotes the weight of the pth term of reviews; T denotes the window size of the convolution operation; $\alpha_j^{(h)}$ is the weight of the jth term for the hth head cross-attention, which is computed by Equation (4); and Ω is the normalization method. We adopt a max-normalization scheme for Ω in this study. Finally, the words with importance scores $\delta(w)$ between 0.6 and 0.8 are highlighted in light gray, whereas the words with importance scores larger than 0.8 are highlighted in dark gray, as shown in Online Appendix 2.

In order to analyze the aspects of users and items, the important terms of reviews and their corresponding importance scores are presented in Table 9. In the review of item 1, which has a lower rating, many words reflecting negative viewpoints have larger attention weights, such as "hissing," "leak," "stain," and "moist." In contrast, more positive words are found in the user review for item 2, such as "durable," "cheap," and "appealing." Moreover, it can be observed that the two reviews have different concerned aspects. Thus, we may infer the reason for the user's rating on item 1 as follows: Although item 1 has stable sound, low noise, and good service quality, the user is dissatisfied with its thickness and thus gives it a neutral rating. In contrast, the terms brand and workability in the review of item 2 are highlighted.

Based on the findings, the user's rating of item 1 can be explained as follows: Although the durability, playability, and sound quality of item 1 are highly appreciated, the user gives item 1 a rating of 3 because of its high price. On the other hand, the user gives a high rating of 4.84 to item 2 because there are many positive words with high weight values, such as *works good*, *sounds good*, and *durable*. The above examples show that RPMIA can provide some explanations about the predicted ratings.

Figure 2. Effects of the Dimension Size of Term Vectors on MSE

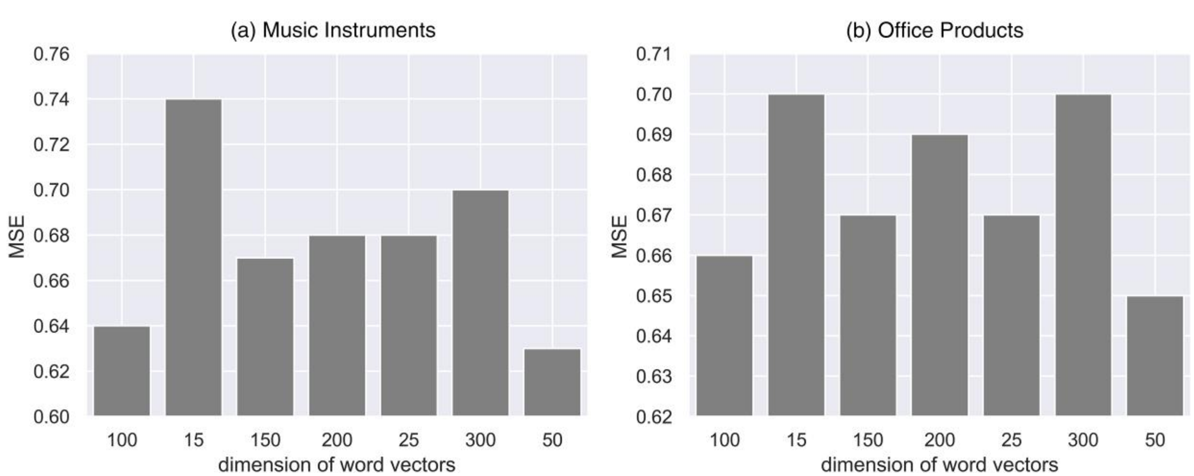


Table 6. Effect of Different Numbers of Aspects on MSE of RPMIA

Data sets	$K = 3$	$K = 5$	$K = 7$	$K = 9$
Musical instruments	0.643	**0.636**	0.638	0.640
Office products	0.703	**0.688**	0.689	0.692
Digital music	0.812	**0.811**	0.817	0.814
Grocery & gourmet food	0.754	**0.747**	0.751	0.755
Video games	0.815	**0.788**	0.804	0.806
Tools & home improvement	0.961	**0.939**	0.941	0.941
Sports & outdoors	0.652	**0.638**	0.643	0.642

4.7.2. User Study on Interpretability of RPMIA. In order to investigate user perceptions of the interpretability of RPMIA, we conducted a laboratory experiment with 35 participants. Those participants were graduate students at the first author's university who were older than 20 years of age and experienced with OCRs. The experiment procedure consisted of the following steps: 60 reviews (i.e., half with highlighted words generated by the D-attn method and the other half with highlighted words generated by RPMIA) were presented to each participant one by one in a random order. After reading each review, each participant provided an assessment score in the range of one (not uninterpretable at all) to five (extremely interpretable) about the interpretability of the highlighted words in the review for the associated rating before moving on to the next review. The participants were not informed about the correspondence between the explainable results and the methods being examined.

After all participants provided the assessment scores of the D-attn and RPMIA methods, all the scores for one method were averaged to derive the final assessment score for that method. A pairwise t-test shows that RPMIA receives significantly higher interpretability assessment scores from the participants (mean = 3.91) than D-attn does (mean = 3.25; $p < 0.01$).

5. Discussion

5.1. Major Findings

In this study, we propose RPMIA, a novel OCR-based and aspect-oriented recommendation model enhanced by multiple user-item interaction relationships discovered from OCRs. There are several major findings of this study. First, RPMIA that incorporates multilevel interactions between users and items based on the attention mechanisms produces significantly better rating predictions than the state-of-the-art baselines. It should be pointed out that different products have different aspects, so reviews on different products may vary significantly in terms of their textual content, leading to significant differences in their embeddings of review text. For example, digital music is a digital product and has unique aspects and context of reviews. The user-digital music interaction and context in reviews may differ from those in reviews of other products, causing the smaller performance improvement of the former. Second, the ablation experiments demonstrate positive impacts of the three proposed design artifacts, including the context-aware interaction module, the aspect-aware interaction module, and the UIAP-aware interaction module, on the rating prediction performance of RPMIA. Third, we evaluate the model interpretability via a laboratory experiment. The result shows that RPMIA provides better interpretability than a baseline model.

5.2. Research Contributions

First and foremost, very few existing rating prediction models have considered interaction relationships between consumers and products at multiple levels. RPMIA is the first OCR-based and aspect-oriented rating prediction model that incorporates three levels of interaction relationships between consumers and products discovered from OCRs simultaneously, including the context, aspect and UIAP levels. Those interaction relationships can

Table 7. MSEs of RPMIA and Its Variant Modules

Methods	Musical instruments	Digital music	Office products	Grocery & gourmet food	Video games	Tools & home improvement	Sports & outdoors
RPMIA-Context	0.643	0.711	0.839	0.774	0.848	0.979	0.671
RPMIA-Aspect	0.644	0.712	0.835	0.778	0.842	0.971	0.668
RPMIA-UIAP	0.640	0.714	0.833	0.795	0.834	0.960	0.664
RPMIA	**0.636**	**0.688**	**0.811**	**0.747**	**0.788**	**0.939**	**0.638**

Note. RPMIA-X represents RPMIA without X-aware module where X is the context, aspect, or UIAP, respectively.

Table 8. Two Rating Examples of a User (ID A1GMWTGXW682GB)

Items	Item IDs	Item name	Actual ratings	Predicted ratings
Item 1	B00IZCSW3M	A string	3.0	2.96
Item 2	B000068NW5	A cable	5.0	4.95

help better identify the preferences and interests of individual consumers for individual products, as well as improve the representations of consumers and items and their associations.

Second, this research designs and evaluates three novel modules for discovering and integrating three levels of interaction relationships in RPMIA. The evaluation clearly shows not only the superiority of RPMIA to existing OCR-based rating prediction models, but also the positive impact of each of those three interaction relationships on rating prediction accuracy. In addition, RPMIA considering all three types of interaction relationships outperforms RPMIA considering any two of those three interaction relationships, suggesting that integrating all three interaction relationships is more effective in improving rating prediction than integrating two relationships only. The designs and evaluations of those artifacts provide both technical insights and theoretical guidance on how to capture and incorporate such interaction relationships in future OCR-based rating prediction models to further improve prediction performance.

Third, existing studies on rating predictions have rarely explored the interpretability of their proposed models, which may have influence on the adoption and deployment of those recommendation models and consumers' trust in recommended products. This study is the first to examine the interpretability of the proposed model through a user experiment. It offers a new dimension and perspective for evaluation of rating prediction models in future research.

5.3. Practical Insights

This study also provides multiple practical insights. First, the findings of this study demonstrate the necessity and benefits of incorporating multiple levels of user-item interaction relationships into rating prediction models to improve their prediction performance. Second, the recommendation model proposed in this study can automatically highlight important key words in the review text that can reflect user preferences and item attributes. Furthermore, it can be used to explain predicted ratings and reveal why an item is recommended to a user. We predict that this interpretability may increase the likelihood of consumers to accept recommendations, help with their purchase decisions, and increase their stickiness to an e-commerce platform. In addition, e-commerce platforms can also profile users and items based on the words highlighted by the proposed model (see examples in Online Appendix 2) to create added value for the platforms.

5.4. Limitations and Future Research

This study has some limitations that may provide future research opportunities. First, because of using a fixed bandwidth for convolutional functions, RPMIA has a limitation in capturing long-term dependencies within text. We call for future research to explore potential guidelines or systematic approaches for determining the optimal length of context when analyzing social media text. Potential areas of exploration include dynamic adjustment of n based on text characteristics and leveraging transformer-based architectures that are capable of capturing long-term dependencies. In addition, although RPMIA utilizes the averaging method to simplify and stabilize the representation of contexts and aspects, future research can explore better methods for effective aggregation of context and aspect representations while keeping the heterogeneity of different representations.

Table 9. Attentive Weights of Highlighted Words in Reviews of a Certain User on Items 1 and 2

Sample reviews	Weights of words	Attentive weights of highlighted words
Review on item 1	>0.9	Right angle(0.93), favorite tone(0.91), comfort(0.91)
	0.8–0.9	Thin(0.89), brand(0.88), build quality(0.86), works good(0.84), leak(0.84), hissing(0.83)
	0.7–0.8	Stain (0.78), card (0.76), high price (0.74)
	0.6–0.7	Packaging (0.69), bunch up (0.69), dialed in (0.67), sounds durable (0.64), good (0.62)
Review on item 2	>0.9	Works good (0.94), pattern (0.93), brand (0.93), smoking deal (0.92), awesome (0.91)
	0.8–0.9	Durable (0.89), maintenance (0.88), sounds good (0.88), appealing (0.88), right angle (0.84)
	0.7–0.8	Build quality (0.79), great (0.77), happy (0.73)
	0.6–0.7	Packaging (0.69), power (0.67), hook up (0.66), favorite (0.63), cumbersome (0.63), stain (0.6)

Endnotes

[1] See https://www.statista.com/statistics.

[2] In this paper, the term attention is referred to as a mechanism in deep learning that attends to a certain part of the input sequence that would arguably be of higher importance.

[3] In this paper, the terms users and consumers are interchangeable.

References

Adomavicius G, Bockstedt J, Curley S, Zhang J (2019) Reducing recommender systems biases: An investigation of rating display designs. *MIS Quart.* 43:19–18.

Adomavicius G, Bockstedt JC, Curley SP, Zhang J (2022) Effects of personalized recommendations versus aggregate ratings on post-consumption preference responses. *MIS Quart.* 46(1):627–644.

Chen L, Chen G, Wang F (2015) Recommender systems based on user reviews: The state of the art. *User Modeling User-Adapted Interaction* 25(2):99–154.

Cheng H-T, Koc L, Harmsen J, Shaked T, Chandra T, Aradhye H, Anderson G, et al. (2016) Wide & deep learning for recommender systems. *Proc. 1st Workshop Deep Learning for Recommender Systems* (Association for Computing Machinery, New York), 7–10.

Chin JY, Zhao K, Joty SR, Cong G (2018) ANR: Aspect-based neural recommender. *Proc. 27th ACM Internat. Conf. Inform. Knowledge Management* (Association for Computing Machinery, New York), 147–156.

Da'U A, Salim N, Idris R (2021) Multi-level attentive deep user-item representation learning for recommendation system. *Neurocomputing* 433:119–130.

Guo X, Chen G, Wang C, Wei Q, Zhang Z (2021) Calibration of voting-based helpfulness measurement for online reviews: An iterative Bayesian probability approach. *INFORMS J. Comput.* 33(1):246–261.

Hernández-Rubio M, Cantador I, Bellogín A (2019) A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews. *User Modeling User-Adapted Interaction* 29(2):381–441.

Hong H, Xu D, Wang GA, Fan W (2017) Understanding the determinants of online review helpfulness: A meta-analytic investigation. *Decision Support Systems* 102:1–11.

Khan ZY, Niu Z (2021) CNN with depthwise separable convolutions and combined kernels for rating prediction. *Expert Systems Appl.* 170:114528.

Kim DH, Park C, Oh J, Lee S, Yu H (2016) Convolutional matrix factorization for document context-aware recommendation. *Proc. 10th ACM Conf. Recommender Systems* (Association for Computing Machinery, New York), 233–240.

Li J, Yang H, Zong C (2018) Document-level multi-aspect sentiment classification by jointly modeling users, aspects, and overall ratings. *Proc. 27th Internat. Conf. Comput. Linguistics* (Association for Computational Linguistics, Stroudsburg, PA), 925–936.

Li C, Quan C, Peng L, Qi Y, Deng Y, Wu L (2019) A capsule network for recommendation and explaining what you like and dislike. *Proc. ACM SIGIR Internat. Conf. Res. Development Inform. Retrieval* (Association for Computing Machinery, New York), 275–284.

Liu P, Zhang L, Gulla JA (2021) Multilingual review-aware deep recommender system via aspect-based sentiment analysis. *ACM Trans. Inform. Systems* 39(2):1–33.

Lu Y, Dong R, Smyth B (2018) Coevolutionary recommendation model: Mutual learning between ratings and reviews. *Proc. World Wide Web Conf.* (Association for Computing Machinery, New York), 773–782.

Markopoulos PM, Aron R, Ungar LH (2016) Product information websites: Are they good for consumers? *J. Management Inform. Systems* 33(3):624–651.

Musto C, de Gemmis M, Semeraro G, Lops P (2017) A multi-criteria recommender system exploiting aspect-based sentiment analysis of users' reviews. *Proc. 11th ACM Conf. Recommender Systems* (Association for Computing Machinery, New York), 321–325.

Rashed A, Elsayed S, Schmidt-Thieme L (2022) Context and attribute-aware sequential recommendation via cross-attention. *Proc. 16th ACM Conf. Recommender Systems* (Association for Computing Machinery, New York), 71–80.

Sanchez-Loor DA, Chang W (2023) Experimental study of the effects of structural assurance, personal experiences, and product reviews on repurchase behavior in e-commerce platforms. *Electronic Commerce Res.* 23(3):1971–2010.

Seo S, Huang J, Yang H, Liu Y (2017) Interpretable convolutional neural networks with dual local and global attention for review rating prediction. *Proc. 11th ACM Conf. Recommender Systems* (Association for Computing Machinery, New York), 297–305.

Shin D, He S, Lee GM, Whinston AB, Cetintas S, Lee KC (2020) Enhancing social media analysis with visual data analytics: A deep learning approach. *MIS Quart.* 44(4):1459–1492.

Siering M, Deokar AV, Janze C (2018) Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews. *Decision Support Systems* 107:52–63.

Wang X, Ounis I, Macdonald C (2022) Effective rating prediction using an attention-based user review sentiment model. Hagen M, Verberne S, Macdonald C, Seifert C, Balog K, Nørvåg K, Setty V, eds. *Proc. 44th Eur. Conf. IR Res.*, Lecture Notes in Computer Science 13185 (Springer, Cham, Switzerland), 487–501.

Wang Z, Xia H, Chen S, Chun G (2021) Joint representation learning with ratings and reviews for recommendation. *Neurocomputing* 425:181–190.

Wu S, Zhang Y, Zhang W, Bian K, Cui B (2021) Enhanced review-based rating prediction by exploiting aside information and user influence. *Knowledge-Based Systems* 222:107015.

Wu L, Quan C, Li C, Wang Q, Zheng B, Luo X (2019) A context-aware user-item representation learning for item recommendation. *ACM Trans. Inform. Systems* 37(2):1–29.

Yu D, Fang A, Xu Z (2023) Topic research in fuzzy domain: Based on LDA topic modelling. *Inform. Sci.* 648:119600.

Yu L, Gong W, Zhang D, Ding Y, Fu Z (2024) From interaction to prediction: A multi-interactive attention-based approach to product rating prediction. https://dx.doi.org/10.1287/ijoc.2023.0131.cd, https://github.com/INFORMSJoC/2023.0131.

Yuan Z, Wu F (2019) Neural review rating prediction with user and product memory. *Proc. 28th ACM Internat. Conf. Inform. Knowledge Management* (Association for Computing Machinery, New York), 2341–2344.

Zeng D, Liu Y, Yan P, Yang Y (2021) Location-aware real-time recommender systems for brick-and-mortar retailers. *INFORMS J. Comput.* 33(4):1608–1623.

Zhang J, Wang C, Chen G (2021) A review selection method for finding an informative subset from online reviews. *INFORMS J. Comput.* 33(1):280–299.

Zhang D, Li W, Niu B, Wu C (2023) A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information. *Decision Support Systems* 166:113911.

Zhu Z, Yan M, Deng X, Gao M (2022) Rating prediction of recommended item based on review deep learning and rating probability matrix factorization. *Electronic Commerce Res. Appl.* 54:101160.