# PMAR: Multi-aspect Recommendation Based on Psychological Gap

Liye Shi[1], Wen Wu[1,2(✉)], Yu Ji[1], Luping Feng[3], and Liang He[1]

[1] School of Computer Science and Technology, East China Normal University, Shanghai, China
lyshi@ica.stc.sh.cn, 52205901009@stu.ecnu.edu.cn, lhe@cs.ecnu.edu.cn

[2] Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, School of Psychology and Cognitive Science, East China Normal University, Shanghai, China
wwu@cc.ecnu.edu.cn

[3] School of Data Science and Engineering, East China Normal University, Shanghai, China

**Abstract.** Review-based recommendations mainly explore reviews that provide actual attributes of items for recommendation. In fact, besides user reviews, merchants have their descriptions of the items. The inconsistency between the descriptions and the actual attributes of items will bring users psychological gap caused by the Expectation Effect. Compared with the recommendation without merchant's description, users may feel more unsatisfied with the items (below expectation) or be more impulsive to produce unreasonable consuming (above expectation), both of which may lead to inaccurate recommendation results. In addition, as users attach distinct degrees of importance to different aspects of the item, the personalized psychological gap also needs to be considered. In this work, we are motivated to propose a novel Multi-Aspect recommendation based on Psychological Gap (PMAR) by modelling both user's overall and personalized psychological gaps. Specifically, we first design a gap logit unit for learning the user's overall psychological gap towards items derived from textual review and merchant's description. We then integrate a user-item co-attention mechanism to calculate the user's personalized psychological gap. Finally, we adopt Latent Factor Model to accomplish the recommendation task. The experimental results demonstrate that our model significantly outperforms the related approaches w.r.t. rating prediction accuracy on Amazon datasets.

**Keywords:** Review-based recommendation · Collaborative filtering · Psychological gap · Deep learning

## 1 Introduction

A variety of review-based recommendations have been proposed, which incorporate the valuable information in user-generated textual reviews into the recommendation process [2,14,15]. Recently, deep learning methods have achieved


A. Bhattacharya et al. (Eds.): DASFAA 2022, LNCS 13246, pp. 118–133, 2022.
https://doi.org/10.1007/978-3-031-00126-0_8

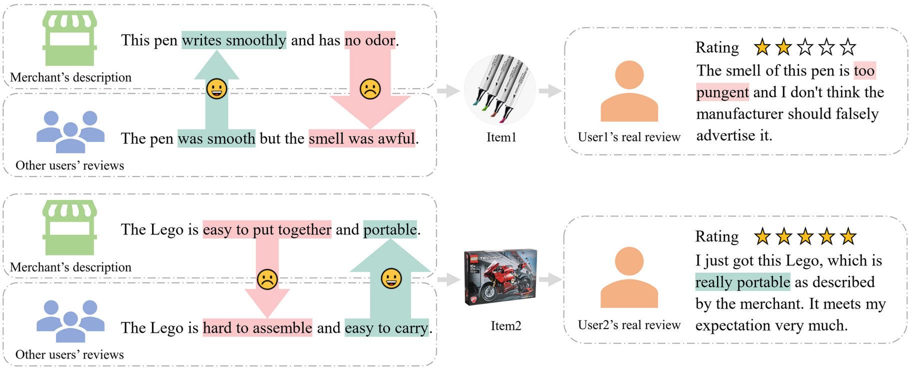


**Fig. 1.** Illsutrates the pshycholgical gap produced by the inconsistency between merchant's description and textual review (Note: red arrows indicate the user feels unsatisfied because the item's actual attributes are not as good as described by the merchant; in contrast, green arrows denote that the user has an impulse to consume the item. The size or thickness of an arrow represents the degree to which users value the corresponding aspect of the item.) (Color figure online)

good performance for review-based recommendations. Although these methods perform well, there are still some limitations that may influence the performance.

Firstly, existing studies mainly model a user's preference for items based on textual reviews written by users to provide actual attributes of items [7,8]. In fact, in addition to reviews, the merchants have their descriptions of the items. The inconsistency of descriptions and actual attributes of items will bring users psychological gap, which is a phenomenon caused by the Expectation Effect [1]. Compared with the recommendations without merchant's description, if an item's actual attributes are lower than expected, the user may be more unsatisfied with the item [17]. Conversely, if the actual attributes of an item are higher than expected, users are more likely to produce unreasonable consumption [18]. Both situations may lead to inaccurate recommendation results. In our work, we consider both user's overall psychological gap and personalized psychological gap in the recommendation process. Generally speaking, we model the overall psychological gap based on the merchant's description and actual attributes reflected by other users' reviews of the item. As for the personalized psychological gap, because users pay attention to different aspects of the item, we assign distinct importance to each aspect's psychological gap.

Figure 1 illustrates one user's overall psychological gap and personalized psychological gap towards different aspects of items in our review-based recommendation process. For item1, we observe that actual attributes (such as smell) reflected by other users' reviews are not as good as the merchant's description, which will cause user1 to be more unsatisfied with item1, and even feel cheated by the merchant. Under this circumstance, the recommendation should reduce the probability of recommending item1 to user1. While for item2, the actual attributes are more than expected relative to the merchant's description (i.e., more than expected), which will make the user more impulsive to buy. In this case, our model would increase the probability of recommending item2 to user2. Regarding personalized psychological gap, take item2 which contains two aspects (i.e., "easy to assemble" and "portable") as an example. Our model will predict that user2 pays the most attention to the "portable" aspect of item2. In this case, user2's psychological gap in the "portable" aspect is considered to be more essential to user2's preference for item2. As for the "easy to assemble" aspect, even if the description of the merchant is inconsistent with the actual attributes of item2 reflected by other user reviews, the psychological gap of user2 based on this aspect has little impact on user2's preference towards item2. In this way, we can infer that item2 still meets user2's preference.

In this paper, we are motivated to propose a Multi-Aspect recommendation based on Psychological Gap (shorten as PMAR). Based on textual review and merchant's description, PMAR models a user's overall psychological gap and personalized psychological gap for better recommendation. We first design an Overall Psychological Gap Module which adopts a gap logit unit to model the user's overall psychological gap based on merchants' descriptions and item reviews. We then propose a Personalized Psychological Gap Module which uses an attention mechanism to select relevant user history reviews based on the semantic information of the review and corresponding item intrinsic information. Besides, a user-item co-attention mechanism is used to obtain the user's personalized psychological gap towards the item. Lastly, we get the user's and item's final representations and apply the Latent Factor Model (LFM) [3] model to complete the rating prediction task. We conduct experiments on several real-world datasets to verify the effectiveness of our model. The experimental results show that our model achieves significantly higher rating prediction accuracy than state-of-the-art models.

In the following sections, we first introduce related work in Sect. 2. We then describe our proposed PMAR model in Sect. 3. We further present the details of our experimental design and results analysis in Sect. 4. Finally, we conclude the paper and indicate some future directions in Sect. 5.

## 2 Related Work

In recent years, how to integrate the rich information embedded in reviews into the recommendation generation has attracted increasing attention [12]. The review-based recommendations can help derive user preferences and item attributes by considering the semantic information reflected in reviews. With the breakthrough of deep learning technology, many works have been proposed to model contextual information from reviews for better recommendation. For example, DeepCoNN [19] adopted two parallel CNNs [4] in the last layer in order to generate potential representations of users and items at the same time. It placed a shared layer on the top to couple these two CNNs together. Neural Attentional Regression model with Review-level Explanations (NARRE) [2] utilized an attention mechanism to select essential reviews when modelling users

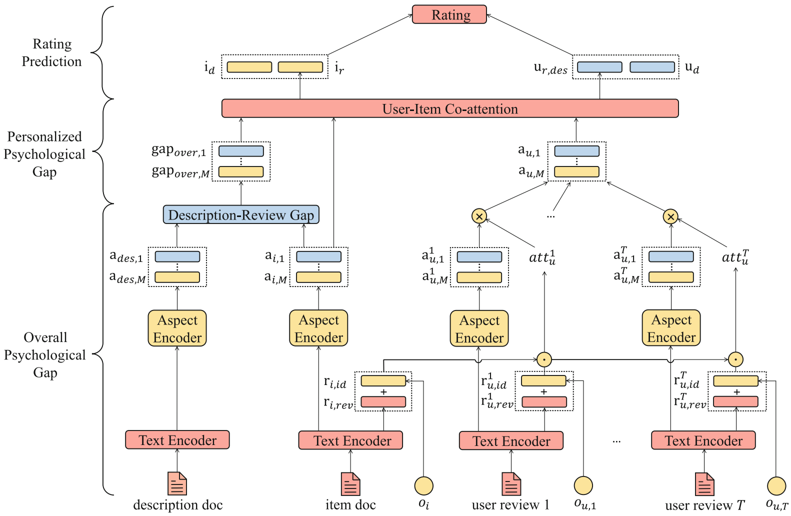


**Fig. 2.** The architecture of PMAR.

and items. Dual Attention Mutual Learning between Ratings and Reviews for Item Recommendation (DAML) [7] utilized local and mutual attention of the CNN to learn the features of reviews jointly. Very recently, MRCP [9] designed a three-tier attention network to select the informative words and reviews for users and items in a dynamic manner. In fact, in addition to reviews, merchants' descriptions of items are also important. The inconsistency of the merchant's descriptions and the item's actual attributes will arise psychological gaps of users [17,18]. For instance, as stated in [17], users would feel a tremendous psychological gap and even worse cheated when the quality of items did not match what the merchant advertised. Besides, [18] argued that when a user bought an item, the user's experiences exceeding expectations towards the item would increase his/her satisfaction.

## 3 Methodology

In this section, we will introduce the details of our PMAR model. The architecture is shown in Fig. 2 including three major modules. The first module is called *Overall Psychological Gap Module*. A Text Encoder layer is integrated to capture the contextual information of words in reviews and descriptions. Then an Aspect Encoder layer is designed to obtain the aspect representations of different aspects. In addition, a Description-Review Gap layer is used to model the user's overall psychological gap on the item based on merchant's description and item reviews. The second module is the *Personalized Psychological Gap Module*, where a User Review Selection layer is adopted to select relevant user history reviews based on the semantic information of reviews and the item information corresponding to reviews. Then a User-Item Co-attention layer is designed to calculate the user's personalized psychological gap. The last module is the *Rating Prediction Module*, where a Latent Factor Model (LFM) is utilized to predict users' ratings for the items.

### 3.1 Overall Psychological Gap Module

As an item's reviews are homogeneous which means each review is written for the same item, we merge all reviews written by distinct users for the item as the item review document. Item $i$'s review document is represented as a word sequence $D_i = \{w_1, w_2, ..., w_{l_i}\}$ where $l_i$ is the length of the review document. Similarly, we treat the description of item $i$ as a description document which is represented as $D_{des}$, and the length of the description document is $l_{des}$. On the other hand, reviews written by a user are for different items (i.e. heterogeneous), so we take the user's reviews as input separately. For each user $u$, the set of his/her reviews is denoted by $S_u = \left\{S_u^1, S_u^2, \cdots, S_u^T\right\}$, where $S_u^t$ denotes the $t$-th review of user $u$, $T$ denotes the number of reviews written by user $u$. Besides, user $u$'s $t$-th review can be represented as a word sequence $S_u^t = \left\{w_1^t, w_2^t, ..., w_{l_u}^t\right\}$, where $l_u$ is the length of each review. Furthermore, the design of this module includes three layers: text encoder layer, aspect encoder layer, and description-review gap layer (see Fig. 2).

- **Text Encoder**

Given item $i$'s review document $D_i = \{w_1, w_2, ..., w_{l_i}\}$, we first project each word to its embedding representation $\mathbf{D}_i = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_{l_i}]$ where $\mathbf{w}_j \in R^{d_w}$ is the embedding vector for the $j$-th word, $d_w$ is the dimension of word embedding, and $l_i$ is the length of item review document. Then CNNs [4] are adopted to capture the context information around each word. Specifically, $f_c$ convolution filters with the sliding window of size $s$ are applied over matrix $\mathbf{D}_i$ to obtain contextual features of each word. The feature matrix $\mathbf{C}_i = [\mathbf{c}_{i,1}, \mathbf{c}_{i,2}, \ldots, \mathbf{c}_{i,l_i}]$ is the resultant feature, where $\mathbf{c}_{i,j}$ is the latent contextual feature vector for word $w_j$ in item $i$'s review document.

- **Aspect Encoder**

As mentioned before, an item has many aspects, such as quality, price and service. Each word in the review plays a distinct role on different aspects, so our goal is to derive a set of aspect-level item representations. Specifically, we first design $M$ different aspect projection matrices over words which are denoted as $\mathbf{AM} = [\mathbf{W}_1, \ldots, \mathbf{W}_M]$, where $M$ is the number of aspect, $\mathbf{W}_m \in R^{d_w \times d_w}$ represents the $m$-th aspect projection matrix. To ensure our model is end to end, the aspect in our model is defined as an implicit feature, and there is no need to pre-process to extract explicit aspects with other tools. The aspect-specific document embedding matrix $\mathbf{G}_{i,m}$ for item $i$'s review document and the $m$-th aspect can be calculated as:

$$\mathbf{G}_{i,m} = [\mathbf{g}_{i,1,m}, \mathbf{g}_{i,2,m}, \ldots, \mathbf{g}_{i,l_i,m}] \tag{1}$$

$$\mathbf{g}_{i,j,m} = \mathbf{W}_m \mathbf{c}_{i,j} + b_m \tag{2}$$

where $\mathbf{g}_{i,j,m}$ is the word $w_j$'s embedding towards the $m$-th aspect, $b_m$ is the bias. Since we need to select important words under different aspects, we generate $M$ aspect query vectors which are denoted as $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_M]$, where $\mathbf{v}_m$ represents the $m$-th aspect query vector. Then the $m$-th aspect representation of item $i$'s review document $\mathbf{a}_{i,m}$ can be represented as:

$$\mathbf{a}_{i,m} = \sum_{j=1}^{l_i} \beta_{i,j,m} \mathbf{g}_{i,j,m} \tag{3}$$

$$\beta_{i,j,m} = \frac{\exp\left(\mathbf{v}_m^\top \mathbf{g}_{i,j,m} / \tau\right)}{\sum_{k=1}^{l_i} \exp\left(\mathbf{v}_m^\top \mathbf{g}_{i,k,m} / \tau\right)} \tag{4}$$

Here, $\tau$ is the temperature parameter that is used to sharpen weights of important words, and $\beta_{i,j,m}$ is the importance of word $w_j$ towards the $m$-th aspect. Similarly, we can get the $m$-th aspect representation of item $i$'s description document $\mathbf{a}_{des,m}$. As mentioned above, the input of user reviews is a series of reviews. So we regard each review of user $u$'s review set as a piece of document and repeat the steps similar to item document. Then $\mathbf{C}_u^t = \left[\mathbf{c}_{u,1}^t, \mathbf{c}_{u,2}^t, \ldots, \mathbf{c}_{u,l_u}^t\right]$ is the feature matrix of the $t$-th review of user $u$ after convolution process. The aspect-specific review embedding matrix $\mathbf{G}_{u,m}^t$ for user $u$'s $t$-th review and the $m$-th aspect can be calculated as:

$$\mathbf{G}_{u,m}^t = \left[\mathbf{g}_{u,1,m}^t, \mathbf{g}_{u,2,m}^t, \cdots, \mathbf{g}_{u,l_u,m}^t\right] \tag{5}$$

$$\mathbf{g}_{u,j,m}^t = \mathbf{W}_m \mathbf{c}_{u,j}^t + b_m \tag{6}$$

where $\mathbf{g}_{u,j,m}^t$ is the embedding of word $w_j^t$ towards the $m$-th aspect, and $b_m$ is the bias. Then we get the $m$-th aspect representation of user $u$'s $t$-th review $\mathbf{a}_{u,m}^t$ as:

$$\mathbf{a}_{u,m}^t = \sum_{j=1}^{l_u} \beta_{u,j,m}^t \mathbf{g}_{u,j,m} \tag{7}$$

$$\beta_{u,j,m}^t = \frac{\exp\left(\mathbf{v}_m^\top \mathbf{g}_{u,j,m}^t / \tau\right)}{\sum_{k=1}^{l_u} \exp\left(\mathbf{v}_m^\top \mathbf{g}_{u,k,m}^t / \tau\right)} \tag{8}$$

where $\tau$ is the temperature parameter, and $\beta_{u,j,m}^t$ is the importance of word $w_j^t$ towards the $m$-th aspect.

- **Description-Review Gap**

In reality, in addition to the reviews, the merchants have their own descriptions of the items. In this layer, we propose a Description-Review Gap layer to model the overall psychological gap between users and items based on item descriptions and item reviews. According to the Aspect Encoder layer, we have obtained the representation of each aspect of item's description document and review document. In order to calculate the difference between the description and the review in each aspect, inspired by CARP [6], we design a gap logit unit to represent the overall psychological gap of the $m$-th aspect between item description document and reviews document as follows:

$$\mathbf{gap}_{over,m} = [(\mathbf{a}_{i,m} - \mathbf{a}_{des,m}) \oplus (\mathbf{a}_{i,m} \odot \mathbf{a}_{des,m})] \tag{9}$$

### 3.2 Personalized Psychological Gap Module

As users focus on different aspects of items when purchasing items, they will have a personalized psychological gap towards different aspects of items. In this module, we first use a User Review Selection layer to select important user reviews based on textual message and corresponding item intrinsic message. Then we design a User-Item Co-attention layer to model the user's personalized psychological gap.

- **User Review Selection**

In this layer, we use an attention mechanism to extract important user reviews according to the current item. We average each word in the item review to get the semantic representation of the document review. Similarly, we average the words in each user review to get the semantic representation of the current review. Unlike the previous method of extracting essential user reviews based on semantic similarity alone, we consider both the semantic information of the reviews and the item ID information corresponding to each user's review. We add two information as the representation of each review. Similarly, for item document reviews, as the document reviews of items are all written to the current item, we add the ID information of the current item. Specifically, user $u$'s fusion representation $\mathbf{r}_u^t$ towards the $t$-th review can be represented as follows:

$$\mathbf{r}_u^t = \mathbf{r}_{u,rev}^t + \mathbf{r}_{u,id}^t \tag{10}$$

$$\mathbf{r}_{u,rev}^t = avg(\mathbf{c}_{u,1}^t, \mathbf{c}_{u,2}^t, ..., \mathbf{c}_{u,l_u}^t), \quad \mathbf{r}_{u,id}^t = \mathbf{o}_{u,t} \tag{11}$$

where $\mathbf{o}_{u,t}$ is the corresponding item's ID embedding about $t$-th review of user $u$, and $avg$ is the average pooling method. Similarly, we represent the item document review as follows:

$$\mathbf{r}_i = \mathbf{r}_{i,rev} + \mathbf{r}_{i,id} \tag{12}$$

$$\mathbf{r}_{i,rev} = avg(\mathbf{c}_{i,1}, \mathbf{c}_{i,2}, ..., \mathbf{c}_{i,l_i}), \quad \mathbf{r}_{i,id} = \mathbf{o}_i \tag{13}$$

where $\mathbf{r}_i$ is the item $i$'s fusion representation based on its review and ID embedding, and $\mathbf{o}_i$ is the item $i$'s ID embedding. Then we use an attention mechanism to select important user reviews according to item $i$ as follows:

$$att_u^t = \frac{\exp\left(\mathbf{r}_i^\top \mathbf{r}_u^t\right)}{\sum_{k=1}^{T} \exp\left(\mathbf{r}_i^\top \mathbf{r}_u^k\right)} \tag{14}$$

where $att_u^t$ represents the importance of user review $S_u^t$ towards item $i$. We then get different aspects' weighted sum representation of user reviews as:

$$\mathbf{a}_{u,m} = \sum_{t=1}^{T} att_u^t \mathbf{a}_{u,m}^t \tag{15}$$

- **User-Item Co-attention**

As stated above, users value different aspects when facing different items. The psychological gaps caused by different aspects of descriptions and comments will also be different. We design a user-item co-attention mechanism to model the importance of different aspects. Formally,

$$weight_m = \mathbf{W}_{uv,m}[\mathbf{a}_{u,m}, \mathbf{a}_{i,m}, \mathbf{a}_{u,m} \odot \mathbf{a}_{i,m}] + b_{uv,m} \tag{16}$$

$$att_m = \frac{\exp(weight_m)}{\sum_{k=1}^{M} \exp(weight_k)} \tag{17}$$

where $att_m$ represents the importance of the $m$-th aspect when user $u$ values the item $i$, $\mathbf{W}_{uv,m} \in R^{3d_w \times 1}$ is the weight parameter, and $b_{uv,m}$ is the bias. Then we get the user's personalized psychology gap:

$$\mathbf{gap}_{per} = \sum_{m=1}^{M} att_m \mathbf{gap}_{over,m} \tag{18}$$

Similarly, user $u$'s and item $i$'s representation from reviews can be denoted as:

$$\mathbf{u}_r = \sum_{m=1}^{M} att_m \mathbf{a}_{u,m}, \quad \mathbf{i}_r = \sum_{m=1}^{M} att_m \mathbf{a}_{i,m} \tag{19}$$

We then concat user $u$'s review representation and personalized psychology gap representation, and pass the obtained vector to the Multilayer Perception to get the user $u$'s representation $\mathbf{u}_{r,des}$:

$$\mathbf{u}_{r,des} = \mathbf{W}_p[\mathbf{u}_r, \mathbf{gap}_{per}] + b_p \tag{20}$$

where $\mathbf{W}_p \in R^{3d_w \times d}$ is the weight parameter, $d$ is the dimension of latent factors, and $b_p$ is the bias.

### 3.3 Rating Prediction Module

Although the user representation $\mathbf{u}_{r,des}$ learned from reviews contains rich aspect information of users, there are some latent characteristics of users which can not be captured in reviews but can be inferred from the rating patterns. Thus, we also represent users according to their ids to capture the latent factors of users. The final representation $\mathbf{u}_f$ of user $u$ is the concatenation of the user representation $\mathbf{u}_{r,des}$ learned from reviews and the user embedding $\mathbf{u}_d \in R^d$ from user ID.

Similarly, we can get the final representation $\mathbf{i}_f$ of item $i$. Formally, it can be represented as follows:

$$\mathbf{u}_f = [\mathbf{u}_d, \mathbf{u}_{r,des}], \quad \mathbf{i}_f = [\mathbf{i}_d, \mathbf{i}_r] \tag{21}$$

We concatenate two representations as $\mathbf{x}_{u,i}$=[$\mathbf{u}_f$, $\mathbf{i}_f$], and then pass it into a Latent Factor Model (LFM) [3]. The LFM function is defined as follows:

$$\hat{y}_{u,i} = \mathbf{W}_x(\mathbf{x}_{u,i}) + b_u + b_i + \mu \tag{22}$$

where $\hat{y}_{u,i}$ denotes the predicted rating,$\mathbf{W}_x$ denotes the parameter matrix of the LFM model, $b_u$ denotes the user bias, $b_i$ denotes the item bias, and $\mu$ denotes the global bias.

### 3.4 Objective Function

We optimize the model parameters to minimize recommendation task loss $\mathcal{L}$. Concretely, the mean squared error is used as the loss function of recommendation task loss, which measures the divergences between the rating score that the model predicts and the gold rating score that a user gives to an item. The recommendation task loss can be calculated as:

$$\mathcal{L} = \frac{1}{|\Omega|} \sum_{u,i \in \Omega} (\hat{y}_{u,i} - y_{u,i})^2 \tag{23}$$

where $\Omega$ denotes the set of instances for training, $y_{u,i}$ is the gold rating score, and $\hat{y}_{u,i}$ is the predicted rating score of the user $u$ to the item $i$ separately.

## 4 Experiments

### 4.1 Experiment Setup

*Datasets and Evaluation Metric.* We used four publicly accessible datasets from Amazon 5-core[1] (i.e., *Automotive* (shorten as "Auto"), *Office Products* (Office), *Grocery and Gourmet Food* (Food) and *Toys and Games* (Toys)), which included the review information that users wrote for items they had purchased. We filtered items that did not have descriptions. Following NARRE [2], we randomly split the dataset into training (80%), validation (10%), and testing (10%) sets. At least one interaction per user/item was included in the training set.

The statistics of the datasets are summarized in Table 1. As for the evaluation, we used Mean Square Error (MSE) [16] to measure the rating prediction. Concretely, we computed the square error between the predicted rating $\hat{y}_{u,i}$ and the ground truth $y_{u,i}$, where $\Omega$ indicated the set of the user-item pairs in the testing set (see Eq. 24).

$$MSE = \frac{1}{|\Omega|} \sum_{u,i \in \Omega} (\hat{y}_{u,i} - y_{u,i})^2 \tag{24}$$

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

**Table 1.** Statistics of datasets used for rating prediction task

| Dataset | Auto | Office | Food | Toys |
|---|---|---|---|---|
| Number of users | 996 | 2,370 | 10,446 | 15,096 |
| Number of items | 683 | 973 | 6,305 | 9,814 |
| Number of ratings | 6,553 | 24,120 | 100,980 | 125,944 |
| Words per user | 279.27 | 834.24 | 454.57 | 408.57 |
| Words per item | 368.49 | 453.93 | 395.37 | 406.09 |
| Density of ratings | 0.96% | 1.05% | 0.15% | 0.09% |

In addition, to make a more intuitive comparison, we used Accuracy Improvement Percentage (AIP) to measure the accuracy improvement percentage of our proposed model against other compared methods (see Eq. 25).

$$AIP = \frac{MSE_{comparedmethod} - MSE_{ourmethod}}{MSE_{comparedmethod}} \tag{25}$$

For each dataset, we performed fivefold cross-validation to avoid any biases. As the data were not normally distributed, we adopted permutation test [11] for significance tests. In our experiments, pre-trained word embeddings were adopted from Google News [10], where the word embedding size was set to 100. We kept the number and the length of reviews covering $pe$ percent users, where $pe$ was set to 0.85 for all four datasets. For item doc and description doc, we kept the length of doc 500 and 100 respectively. Adam was used to updating parameters when training. The learning rate was determined by grid search amongst $\{0.0001, 0.0002, 0.001, 0.002\}$, the dropout ratio was explored amonst $\{0.0, 0.1, ..., 0.9\}$, and the batch size was set amonst $\{32, 64, 128, 256\}$. The window size $s$ was empirically set as 3 and $\tau$ was set as 0.5. All hyper-parameters were tuned according to the validation set.

*Compared Methods.* We compared our approach with eight related methods. These algorithms can be classified into three categories: *rating-based* (NMF and SVD), *review-based* (e.g. DeepCoNN, D-ATTn, NARRE, DAML, NRMA and MRCP) and *variations of our method* (e.g. PMAR-O, PMAR-P and PMAR-ID). The detailed description of each method is given below.

- **NMF** [5]: Non-negative Matrix Factorization is a traditional model which casts the MF model within a neural framework and combines the output with multi-layered perceptrons.
- **SVD** [3]: Singular Value Decomposition is a matrix factorization model which reduces the dimension of a rating matrix and eliminates its sparsity.
- **DeepCoNN** [19]: Deep Cooperative Neural Networks is a neural network which learns the representations of users and items from reviews by using convolutional neural networks.
- **D-ATTn** [13]: Dual Attention CNN Model is an interpretable, dual attention-based CNN model which combines reviews and ratings for product rating prediction.

- **NARRE** [2]: Neural Attentional Rating Regression with Review-level Explanations is a model which uses attention mechanism to explore the usefulness of reviews.
- **DAML** [7]: Dual Attention Mutual Learning is a model which utilizes local and mutual attention of CNN to learn the features of reviews.
- **NRMA** [8]: Neural Recommendation Model with Hierarchical Multi-view Attention is a model which designs a review encoder with multiview attention to learn representations of reviews from words.
- **MRCP** [9]: Multi-aspect Neural Recommendation Model with Context-aware Personalized Attention is a most recently model which designs three encoders to extract hierarchical features of reviews, aspects and users/items.
- **PMAR-O**: This method removes the Description-Review Gap Layer from our PMAR model. That is, the model does not consider the psychological gap.
- **PMAR-P**: As the second variation of our method, this method removes the User-Item Co-attention Layer from our PMAR model so that the weight of the psychological gap of each aspect is the same. That is, it only considers the overall psychological gap rather than the personalized psychological gap.
- **PMAR-ID**: As another variation of our method, this method uses textual message alone to model user's and item's fusion representations in Eq. 10 and Eq. 12.

### 4.2 Overall Performance Comparison

The overall comparison results are shown in Table 2. We observe that our proposed PMAR method is the best in terms of MSE, and the improvements against the best baselines are significant ($p$-value $< 0.05$ via permutation test). Concretely, the advantages against MRCP in Office and Toys are 2.64% and 2.06% respectively (refer to the value of $\Delta$MRCP). The possible reason is that our method not only considers the multi-aspect information in reviews which can construct diverse user preferences and item characteristics, but also the personalized psychological gap brought by the inconsistency of merchant's description and the item's actual attributes to the user. Another interesting observation from Table 2 is that our PMAR model achieves the largest improvement against other baselines in Office datasets (e.g., $\Delta$ MRCP = 2.64% in Office vs. 1.89% in Food). It is reasonable because the density of Office dataset is higher relative to others (see Table 1). In this case, our PMAR model is able to learn user preference and item characteristic more comprehensively, leading to the more accurate personalized psychological gap construction.

Furthermore, when compared with three variations (i.e., PMAR-O, PMAR-P and PMAR-ID), our complete model PMAR also performs better in all four datasets. Among the variations, compared with our PMAR model, PMAR-O which does not consider the psychological gap has the largest drop in the results of all datasets. Specifically, in Office, our PMAR model's advantage against PMAR-O is 2.61% while against PMAR-P is 1.72%. It is reasonable because PMAR-O models neither overall psychological gap nor personalized psychological gap.

**Table 2.** Overall comparison results of rating prediction are measured by MSE (Note: * denotes the statistical significance for $p$-value $< 0.05$ compared to the best baseline, the boldface indicates the best model result of the dataset, and the underline indicates the best baseline result of the dataset.)

| Method | | Auto | Office | Food | Toys |
|---|---|---|---|---|---|
| Rating-based | NMF | 1.0083 | 0.8388 | 1.2707 | 1.0486 |
| | SVD | 0.8276 | 0.7626 | 1.0293 | 0.8931 |
| Review-based | DeepCoNN | 0.7473 | 0.7338 | 0.9881 | 0.8501 |
| | D-ATT | 0.7532 | 0.7397 | 0.9815 | 0.8355 |
| | NARRE | 0.7685 | 0.7124 | 0.9919 | 0.8217 |
| | DAML | <u>0.7320</u> | 0.7033 | 0.9908 | 0.8711 |
| | NRMA | 0.7658 | 0.7118 | 0.9891 | 0.8229 |
| | MRCP | 0.7565 | <u>0.6967</u> | <u>0.9729</u> | <u>0.8191</u> |
| Variations of our method | PMAR-O | 0.7347 | 0.6965 | 0.9657 | 0.8112 |
| | PMAR-P | 0.7201 | 0.6902 | 0.9561 | 0.8062 |
| | PMAR-ID | 0.7302 | 0.6816 | 0.9599 | 0.8066 |
| Our Method | PMAR | **0.7137*** | **0.6783*** | **0.9545*** | **0.8022*** |
| AIP of PMAR | $\Delta$NMF | 29.22% | 19.13% | 24.88% | 23.50% |
| | $\Delta$SVD | 13.76% | 11.05% | 7.27% | 10.18% |
| | $\Delta$DeepCoNN | 4.50% | 7.56% | 3.40% | 5.63% |
| | $\Delta$D-ATT | 5.24% | 8.30% | 2.75% | 3.99% |
| | $\Delta$NARRE | 7.13% | 4.79% | 3.77% | 2.37% |
| | $\Delta$DAML | <u>2.50%</u> | 3.55% | 3.66% | 7.91% |
| | $\Delta$NRMA | 6.80% | 4.71% | 3.50% | 2.52% |
| | $\Delta$MRCP | 5.66% | <u>2.64%</u> | <u>1.89%</u> | <u>2.06%</u> |
| | $\Delta$PMAR-O | 2.86% | 2.61% | 1.16% | 1.11% |
| | $\Delta$PMAR-P | 0.89% | 1.72% | 0.17% | 0.50% |
| | $\Delta$PMAR-ID | 2.26% | 0.48% | 0.56% | 0.55% |

In this way, when there is a great gap between the item's actual attributes and the merchant's description, the psychological gap of the user will have a large impact on the accuracy for recommendation. In addition, the suboptimal results of PMAR-P model (e.g., $\Delta$ PMAR-P = 0.89% in Auto vs. 1.72% in Office) validate the importance of the personalized psychological gap. It is likely because the variation ignores that users pay distinct attentions to different aspects, which will cause users' personalized psychological gaps, resulting in a great impact on accuracy for the recommendation. Besides, the degraded MSEs of PMAR-ID demonstrate that combining both ID and review messages can capture textual message as well as intrinsic attributes of the reviewed item (e.g., $\Delta$ PMAR-ID = 0.56% in Food vs. 0.55% in Toys). Although PMAR-ID uses both the overall psychological gap and personality psychological gap, it uses textual message alone which can not capture intrinsic attributes of the reviewed item.

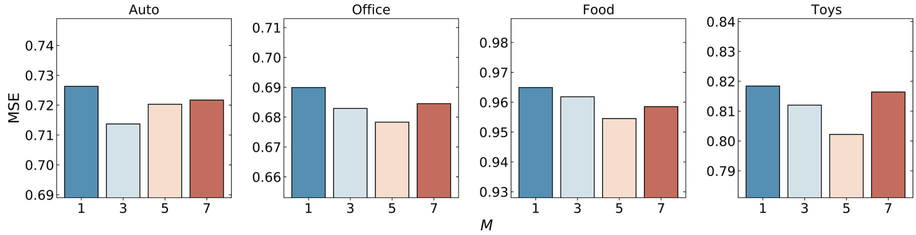


**Fig. 3.** Experimental results with the change of Aspects' Number.

### 4.3 Influence of the Number of Aspects

To investigate whether different number of aspects will influence the performance of our PMAR model, we change $M$ (refer to Sect. 3.1). Following MRCP, the aspects' number $M$ is set amongst $\{1, 3, 5, 7\}$. We can observe from Fig. 3 that the multi-aspect model (i.e., $M > 1$) always performs better than that with only one aspect (i.e., $M = 1$), indicating the effectiveness of multi-aspect diverse representations of users and items. Another finding is that when $M$ equals to 5, our PMAR model performs the best in three datasets (i.e., Office, Food and Toys). In contrast, our PMAR achieves the best result in the Auto dataset when $M$ equals to 3. We argued that the value of the optimal aspects' number might be related to the average number of words per user and each item (e.g., Words Per User: 279.27 in Auto vs 834.24 in Office; Words Per Item: 368.49 in Auto vs 453.93 in Office). That is, more words may contain richer aspects of users and items. In addition, setting the aspects' number to 5 would be an optimal choice for most recommendations.

### 4.4 Case Study

To better understand how our model facilitates the recommendation system based on the user's psychological gap caused by the inconsistency of the merchant's description and the item's actual attributes, we conducted a case study in Fig. 4. We randomly sampled two user-item pairs from Auto and Office datasets. For each user-item pair, we set $M$ as 5 and extracted the most valued aspect by the current user (i.e., the highest attention weight according to $att_m$ in Eq. 17) and another aspect that the user less valued. Similar to CARP [6], to better visualize an aspect, we retrieved the top-$K$ phrases whose weight is the sum of the weights (i.e., $\beta_{m,j,i}$ in Eq. 4 and $\beta^t_{u,j,m}$ in Eq. 8) of the constituent words in the convolutional window. We then selected the most informative sentences containing these phrases to represent the corresponding aspect. Here, we choose $K = 30$ which is the same setting as CARP [6]. In Fig. 4, red and green which are predicted by our PMAR model indicate the phrases in reviews and descriptions of the user's most valued and less valued aspects respectively. As a reference, we manually displayed the parts matched well by the most valued aspects in the target user-item review with yellow color. In addition, $y_{u,i}$ represents the user's

| user1 – item1 (Heavy-Duty Dual Propane Tank Cover): $y_{u,i} = 3.0$ $\hat{y}_{u,i} = 3.3$ $\hat{y}'_{u,i} = 3.9$ | | | aspect importance |
|---|---|---|---|
| aspect 5 | item description: | … is made of heavy-duty polypropylene and protects propane gas bottles. | 0.287 |
| | item history reviews: | It's not as heavy duty as I thought. | |
| aspect 4 | item description: | Easy to assemble in about 15 minutes with supplied hardware. | 0.100 |
| | item history reviews: | The cover was easy to assemble. | |
| user history reviews: | | … hang very secure to under sink door. | |
| target review: | | I guess I expected 'Heavy duty' to be really heavy duty. I felt the product was somewhat flimsy. I think the product should not advertize it to be more than what it is. | |
| user2 – item2 (EnduraGlide Dry-Erase Markers): $y_{u,i} = 5.0$ $\hat{y}_{u,i} = 4.8$ $\hat{y}'_{u,i} = 4.1$ | | | aspect importance |
| aspect 2 | item description: | … write smoother, and erase cleanly. | 0.396 |
| | item history reviews: | … and erase really well. I think this is a really nice set of dry-erase markers. | |
| aspect 1 | item description: | … that delivers bold, continuous color on dry-erase boards. | 0.028 |
| | item history reviews: | The colors (assorted option) aren't as bold as I may have liked them. | |
| user history reviews: | | Of course does not leave any residue either. | |
| target review: | | I was looking to a inexpensive dry-erase marker set when a few of my oldmarkers dried out. These work perfectly. The print erases as expected. | |

**Fig. 4.** Example study of two user-item paris from Auto and Office.

gold rating of the current item, $\hat{y}_{u,i}$ represents the rating that our PMAR model predicts, and $\hat{y}'_{u,i}$ denotes the rating predicted by PMAR-O (i.e., w/o Overall Psychological Gap) model which is a variation of our PMAR model.

**Example 1:** the first user-item pair is a user who gave a tank cover "Heavy-Duty Dual Propane Tank Cover" a low score (i.e. $y_{u,i} = 3.0$). We can observe that our PMAR model predicts that user1 assigns the highest weight on aspect 5 of item1 (e.g., aspect importance: 0.287 of aspect 5 vs. 0.100 of aspect 4) which is regarding the heavy-duty aspect and s/he likes sturdy items according to his/her history reviews. This indicates aspect 5 brings a greater psychological gap to user1 than other aspects. Besides, item1 in aspect 5 is not as heavy-duty as described by the merchant. Although item1 is easy to assemble as advertised by the merchant in aspect 4 (see Fig. 4), the recommendation system should still reduce the probability of recommending item1 to user1. In the rating prediction stage, our model PMAR predicts that user1's rating for item1 is 3.3 points while PMAR-O model predicts 3.9 points. It is reasonable because the variant does not consider the impact of user1's psychological gap which causes user1 to be more disappointed with item1, leading to the inaccuracy for recommendation. Finally, it can be seen that user1's target review denotes that user1 is not satisfied with item1 and considers that the merchant falsely advertises the heavy-duty aspect s/he values most, which is in accordance with our PMAR model's prediction.

**Example 2:** compared with the first pair, the second one illustrates a user-item pair towards a pen "EnduraGlide Dry-Erase Markers" by user2 with a high rating (i.e., $y_{u,i} = 5.0$). It can be seen that our PMAR model predicts that user2 values aspect 2 of item2 most (e.g., aspect importance: 0.396 of aspect 2 vs. 0.028 of aspect 1) which is about the erase-cleanly aspect. Besides, user2 likes items that do not leave any residue according to his/her history reviews. This implies that item2 erases cleanly as the merchant's description in aspect 2 (see Fig. 4), leading to a greater psychological gap to user2 than other aspects. Although regarding the less valued aspect 1 in which item2 is not as bold as advertised by the merchant, the recommendation system should still increase the probability of introducing item2 to user2. As for rating prediction, our model PMAR and PMAR-O models predict that user2's ratings for item2 are 4.8 and 4.1 points respectively. The possible reason is that the variant does not consider the impact of user2's psychological gap which causes user2 to be more satisfied with item2. Finally, we can observe that user2's target review denotes that item2 satisfies user2's expectation of the erase-cleanly aspect, which also meets the prediction of our PMAR model.

## 5 Conclusions and Future Work

In this work, we are motivated by the Expectation Effect to propose a novel PMAR model which learns a user's overall and personalized psychological gap. Concretely, we first construct an Overall Psychological Gap Module which uses a logit unit to calculate the overall psychological gap of merchants description and item reviews. Secondly, we design a Personalized Psychological Gap Module which uses an attention mechanism to select relevant user reviews and item reviews based on review semantic information and the corresponding item intrinsic information. A user-item co-attention mechanism is adopted to calculate the user's personalized psychological gap. We evaluate our model on four real-world datasets (i.e., Auto, Office, Food and Toys). The experimental results show that our model significantly outperforms other state-of-the-art methods in terms of rating prediction accuracy.

**Acknowledgements.** We thank editors and reviewers for their suggestions and comments. We also thank National Natural Science Foundation of China (under project No. 61907016) and Science and Technology Commission of Shanghai Municipality (under projects No. 21511100302, No. 19511120200 and 20dz2260300) for sponsoring the research work.

## References

1. Anderson, R.E.: Consumer dissatisfaction: the effect of disconfirmed expectancy on perceived product performance. J. Mark. Res. **10**(1), 38–44 (1973)
2. Chen, C., Zhang, M., Liu, Y., Ma, S.: Neural attentional rating regression with review-level explanations. In: WWW, pp. 1583–1592 (2018)
3. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009)
4. Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D.: Face recognition: a convolutional neural-network approach. IEEE Trans. Neural Netw. **8**(1), 98–113 (1997)
5. Lee, D., Sebastian Seung, H.: Algorithms for non-negative matrix factorization. In: NIPS, pp. 556–562 (2001)

6. Li, C., et al.: A capsule network for recommendation and explaining what you like and dislike. In: SIGIR, pp. 275–284 (2019)
7. Liu, D., Li, J., Du, B., Chang, J., Gao, R.: DAML: dual attention mutual learning between ratings and reviews for item recommendation. In: KDD, pp. 344–352 (2019)
8. Liu, H., et al.: Hierarchical multi-view attention for neural review-based recommendation. In: NLPCC, pp. 267–278 (2020)
9. Liu, H., et al.: Toward comprehensive user and item representations via three-tier attention network. ACM Trans. Inf. Syst. **39**(3), 1–22 (2021)
10. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: NIPS, pp. 3111–3119 (2013)
11. Nichols, T.E., Holmes, A.P.: Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. **15**(1), 1–25 (2002)
12. Peng, Q., et al.: Mutual self attention recommendation with gated fusion between ratings and reviews. In: DASFAA, pp. 540–556 (2020)
13. Seo, S., Huang, J., Yang, H., Liu, Y.: Interpretable convolutional neural networks with dual local and global attention for review rating prediction. In: RecSys, pp. 297–305 (2017)
14. Wang, X., et al.: Neural review rating prediction with hierarchical attentions and latent factors. In: DASFAA, pp. 363–367 (2019)
15. Wang, X., Xiao, T., Tan, J., Ouyang, D., Shao, J.: MRMRP: multi-source review-based model for rating prediction. In: DASFAA, pp. 20–35 (2020)
16. Willmott, C.J., Matsuura, K.: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. **30**(1), 79–82 (2005)
17. Wu, X.F., Zhou, J., Yuan, X.: The research of factors which effect B2C E-commerce trust-based on the mechanism of process. In: AMR, pp. 2583–2586 (2012)
18. Xiao, B., Benbasat, I.: E-commerce product recommendation agents: use, characteristics, and impact. MIS Q. **31**, 137–209 (2007)
19. Zheng, L., Noroozi, V., Yu, P.S.: Joint deep modeling of users and items using reviews for recommendation. In: WSDM, pp. 425–434 (2017)