Expert Systems With Applications 203 (2022) 117317

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Learning persona-driven personalized sentimental representation for review-based recommendation

Peipei Wang [a], Lin Li [a,*], Ru Wang [b], Xinhao Zheng [a], Jiaxi He [c], Guandong Xu [d]

[a] School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan, 430070, China
[b] School of Information Science and Engineering, Shandong Normal University, Jinan, 250358, China
[c] National University of Singapore, 119077, Singapore
[d] School of Computer Science and Advanced Analytics Institute, University of Technology Sydney, Sydney, NSW 2008, Australia

ARTICLE INFO

MSC:
00-01
99-00

Keywords:
Review mining
Sentimental representation learning
Attention neural networks
Review-based recommendation

ABSTRACT

A large amount of information exists in many e-commerce and review websites as a valuable source for recommender systems. Recent solutions focus on exploring the correlation between sentiment and textual reviews in the review-based recommendation. However, these studies usually pay less attention to the differences of different users in sentimental expression styles or language usage habits when a user writes reviews. In this work, we argue that the individual reviewing behavior is closely related to personality, and sentimental expression is a manifestation of personality. Therefore, we propose a novel Persona-driven Sentimental Attentive Recommendation model (named PSAR) via personalized sentimental interactive representation learning for the review-based recommendation. The proposed model is devised to learn fragment-level and sequence-level personalized sentimental representation simultaneously from reviews. Besides, an attentive persona-driven interaction module is designed to capture word-level usage habits and sentence-level analogous tones. Comprehensive experimental results on four real-world datasets demonstrate that our model outperforms the state-of-the-art methods.

1. Introduction

With the increasing growth of available online information, personalized recommendation (Liu, Zheng, Li, Zhang, et al., 2022; Shen et al., 2021; Tewari & Barman, 2018) is playing an increasingly critical role in alleviating information overload. It is widely used in various fields, such as e-commerce (Bao et al., 2014; Chong et al., 2020; Li et al., 2021), social media (Liu, Zheng, Li, Zhang, et al., 2022; Ning et al., 2019) and industrial (Liu, Zheng, Li, Shen et al., 2022) fields. Not only can it help sellers gain revenue growth through accurate marketing, but users can spend less time discovering the items or services they are interested in. Collaborative Filtering (CF) (Ebesu et al., 2018; Xue et al., 2019) is extensively used in the most successful personalized recommendation, which models user preferences and item features based on historical interaction records such as user ratings and click behaviors. Most CF technologies are based on matrix factorization (MF) to compute the user–item satisfaction score. MF based on bayesian personalized ranking (BPRMF) (Rendle et al., 2009) and generalized MF (GMF) (He et al., 2017) are usually the basic models to improve top-N recommendation. However, these MF methods (Yi et al., 2019) suffer from sparsity problems, and it is also hard to model users' fine-grained preferences and furnish explainable recommendations.

In most online e-commerce platforms (e.g., Amazon), users are allowed to express their attitudes or opinions utilizing reviews apart from ratings. The textual reviews usually contain rich information and fine-grained features. Therefore, review-based methods (Lei et al., 2016; Li et al., 2021; Wu et al., 2018) are proposed and Li et al. (2021) consider historical rating behavior and review latent factor representation learning. Although these studies have achieved significant improvement, we argue that they still have inherent limitations. Most existing methods neglect the differences of different users in sentimental expression styles or language usage habits. In essence, sentimental expression is a manifestation of personality, and different users with different personalities have different sentimental expression habits. For example, when a user writes reviews, on the same sentimental five-star rank, some users usually use gentle words, while others use stimulating words.

Specifically, personality refers to the characteristic pattern in a person's thinking, feeling, and decision making (Siddique et al., 2019).

* Corresponding author.
E-mail addresses: ppwang07@whut.edu.cn (P. Wang), cathylilin@whut.edu.cn (L. Li), ruwang0929@gmail.com (R. Wang), 248681@whut.edu.cn (X. Zheng), jiaxihe1989@gmail.com (J. He), Guandong.Xu@uts.edu.au (G. Xu).

https://doi.org/10.1016/j.eswa.2022.117317
Received 6 April 2021; Received in revised form 5 April 2022; Accepted 23 April 2022
Available online 2 May 2022
0957-4174/© 2022 Elsevier Ltd. All rights reserved.

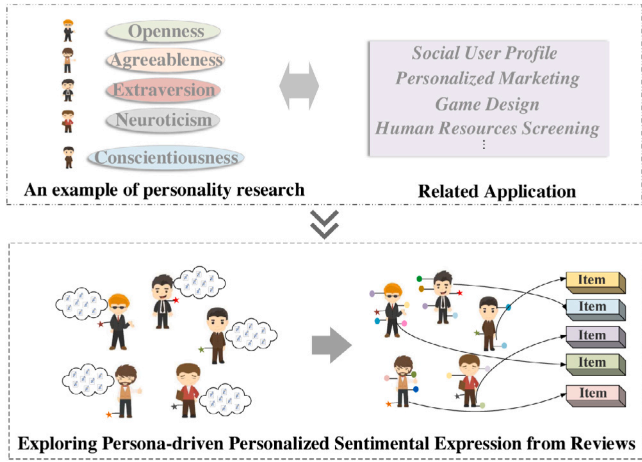


Fig. 1. A schematic of the persona-driven recommendation. The upper box indicates the related applications based on personality research with the Big-Five personality as an example. The bottom box denotes the pipeline of our method for the review-based recommendation. The five-pointed stars with colors represent user personality, and the circles with colors mean personalized sentimental features from reviews.

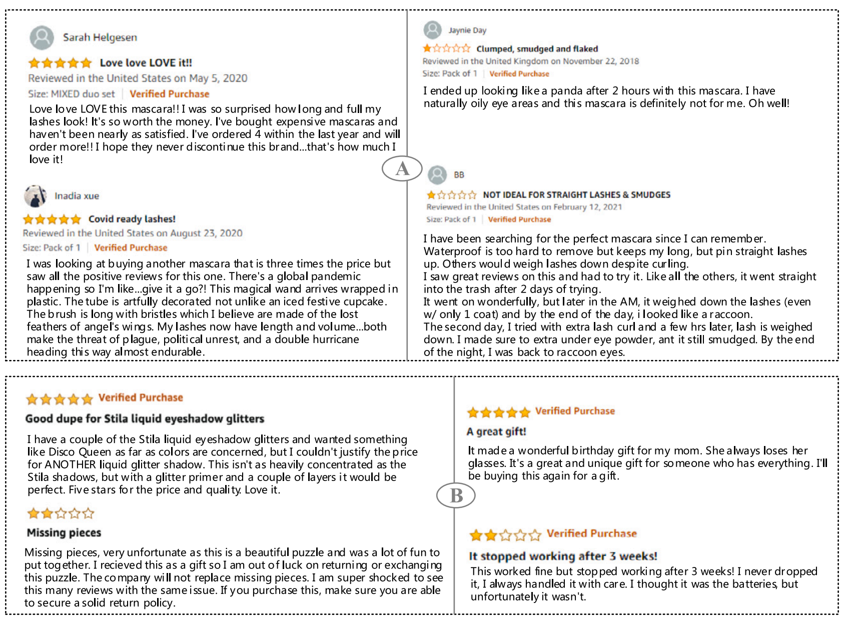


Fig. 2. Examples of reviews for users' sentimental expression on Amazon. Example A is the review documents of different users on the same item. Example B is the review documents of the user in different items.

In other words, it is a quality that a person has acquired during a comparatively long period. While the sentiment is a manifestation of personality and it can be perceived in an instant. Generally speaking, personality is to sentiment what climate is to weather. Personality traits exhibit the difference of preference and habit in people, and it is an integral part of social multiple interactions (Chen, 2011; Wu, Li et al., 2019). Therefore, research on personality has become increasingly significant in many related applications (see Fig. 1) such as social user profiles, personalized marketing, game design, job recruitment, and so on. Above all, in sociological research, words, phrases, and sentences written by people could reflect personality to some extent (Lin et al., 2019; Yarkoni, 2010). As illustrated in Fig. 1, inspired by the advancement of these applications, we hold that the free-text reviews could reflect the users' personalities. In Fig. 2, we show review examples of different sentimental expression on Amazon. Fig. 2(A) is the review documents of different users on the same item, which shows that different users differ in language usage and tone when expressing positive (five-star) or negative (one-star) sentiments on the same item. Fig. 2(B) is the review documents of the user in different items, indicating the differences of sentimental expression styles of users in writing reviews of items. For example, on the left of Fig. 2(A), differently, two users use "love love love and I was so surprised" and "I believe are made of the lost feathers of angel's wings" to express their likes respectively. Intuitively, the difference in user personality leads to the difference in sentimental expression styles or language usage habits. In this case, we aim to utilize personality (same as "persona" below) to drive interactive sentimental representation learning from reviews and help to address the above-mentioned concerns.

To this end, we propose a Persona-driven Sentimental Attentive Recommendation model (PSAR), which can cope with the mentioned problem effectively. As shown in Fig. 3, as for the interactive sentimental representation learning, we first adopt interactive fragmentary and sequential features to learn sentimental expression from reviews respectively, and then concatenate them to obtain the final personalized sentimental representation. In this part, we design an Attentive Persona-based Interaction (API) module to learn the interaction of similar and synonymous words or phrases, as well as the interaction of sentences with the analogous sentimental tone. Besides, we employ two

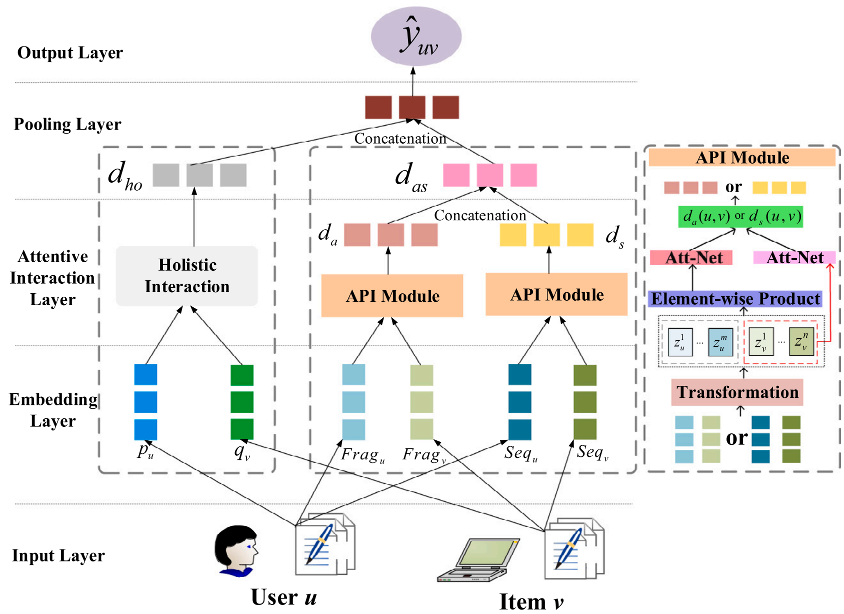


Fig. 3. An overview of persona-based attentive interaction recommendation model.

neural attention networks in the API module to select informative interactions and pick up the users' most concerned item features separately. In addition, the holistic interaction modeling enhances further the connotation of the original sentences. Finally, the results of interactive sentimental representation learning and holistic interaction modeling are concatenated as the input to the output layer, predicating the final satisfaction score. The main contributions of our work can be summarized as follows.

- To the best of our knowledge, we are the first to explore persona-driven sentimental representation modeling from textual reviews, enabling to understand the user's behaviors. Specifically, we explore the complementary benefits of the sequence-based and fragment-based reviews to discover differences in personalized sentimental expression.
- We propose a novel Persona-driven Sentimental Attentive Recommendation model (PSAR) with personalized sentimental representation. PSAR effectively captures sentimental expression habits by using the interaction of similar and synonymous words or phrases, as well as sentences with the analogous sentimental tone.
- Experimental results on four real-world datasets demonstrate that the proposed model is significantly better than existing state-of-the-art methods on tasks of prediction precision, review mining, and recommendation interpretability.

The remainder of this paper is organized as follows. In Section 2, the existing works related to our method are presented. Section 3 describes the proposed PSAR thoroughly. In Section 4, we introduce experimental settings and evaluate the performance of PSAR by comparing some baselines on various real-world datasets. Finally, this paper is concluded and some future research points are shown in Section 5.

2. Related work

As mentioned before, most existing efforts focused on free-text reviews written by users to reach recommendation precision by a large margin and provide the interpretable recommended results (Wang et al., 2018). In the light of our target points, we roughly review the related work from two aspects: review-based recommendation, personality-driven recommendation.

2.1. Review-based recommendation

Conventionally, review-based recommendation methods (Bao et al., 2014; Li et al., 2021; Xu et al., 2019) integrate the review information and ratings to improve recommendation performance, such as EFM (Zhang et al., 2014), TriRank (He et al., 2015), sCVR (Ren et al., 2017), and NAREE (Chen et al., 2018). Deep matrix factorization (Liu, Zheng, Li, Shen et al., 2022) is one of the most promising frameworks in the field of review-based recommendation algorithms. Ling et al. (2014) propose a context-based collaborative filtering unified model by using reviews and ratings. DeepCoNN (Zheng et al., 2017) is a state-of-the-art deep learning-based method and it uses two parallel convolutional structures to model user and item review documents for rating prediction. More specifically, EFM, TriRank and sCVR first extract item features and phrase-level user opinions from reviews, and then yield feature-level explanations in terms of users' preferences on the specific item features. NAREE unitizes an attention network to learn the usefulness of reviews and provide more informative explanations. Wu, Quan et al. (2019) consider textual reviews and rating scores via CNNs and attention mechanisms.

Analogously, aspect-based methods (Hou et al., 2019; Liang et al., 2019; Luo et al., 2014) also extract aspect features from the textual reviews to identify users' interests. For example, Luo et al. (2014) define and address aspect identification problem and develop a topic model to generate ratable aspects. ANR (Chin et al., 2018) exhibits an aspect-based representation model for users and items by an attention-based component. Especially, recent AARM (Guan et al., 2019) outperforms significantly many state-of-the-art methods, which models the aspect interactions to alleviate data sparsity problem in reviews. Obviously, the above methods contribute to the explainable recommendation. However, these methods just extract words or phrases from free-text reviews, which pay less attention to the integrity of reviews and may twist the meaning of original sentence-level reviews. Besides, the above solutions lack research on the correlation between sentimental expression and individual reviewing behavior. Therefore, we aim to discover the persona-driven sentimental representation for the review-based recommendation. The proposed model employs word-level interactions and sentence-level interactions to capture the word usage habit and overall sentimental tone respectively.

2.2. Personality-driven recommendation

Personality, as a domain-independent and stable factor, is widely regarded as a crucial explanation for a person's miscellaneous outward manifestations (Sun et al., 2020). Thus, personality has been researched in various applications (Feng & Qian, 2013; Hu & Pu, 2011; Li et al., 2016; Siddique et al., 2019) including recommender systems. For example, Hu and Pu (2011) incorporate human personality into collaborative filtering framework to tackle cold start problem and verify their model in the movie scene. Personality mining techniques are reported on the precise music and game recommendation (Klec, 2017; Shen et al., 2020; Yang & Huang, 2020). In the content-based recommender systems (Shu et al., 2018; Wu, Quan et al., 2019), Ning et al. (2019) propose a friend recommendation based on Big-Five personality traits and hybrid filtering for accurate and robust recommendation. Yakhchi et al. (2018) understand users' behaviors by detecting the users' personality type implicitly with the aid of the Linguistic Inquiry and Word Count (LIWC) tool. Dhelim et al. (2021) propose a personality-aware product recommendation to predict the topics he/she is interested in by combining the user's personality traits and matching the user's personality facets to relevant items. Although these methods extract user personality and can recommend diver items to target users in a specific application scenario, they fail to consider the connection between personality and sentiment. Therefore, different from these personality-driven recommendation, we target the persona-driven recommendation method in reviews to explore persona-driven personalized sentimental expression. In our PSAR, we employ attentive multiple interactions to enhance the feature connections and capture users' sentimental expression habits from reviews.

3. Persona-based sentimental attentive recommendation

In this section, we introduce the proposed PSAR model in detail. Our PSAR achieves interactive sentimental representation learning. As a core module, API module is described to learn sequence-based (sentence-level) and fragment-based (word-level) attentive interactions severally and enhances the feature connections. In addition, the holistic interaction modeling is designed to enhance further the connotation of the original sentences. Afterward, the rating prediction and network learning are presented. The architecture of our PSAR model is shown in Fig. 3. As illustrated in Fig. 3, the user and item are embedded from original-level, fragment-level, and sequence-level. The original-level embedding is used to achieve holistic interaction. The fragment-level and sequence-level embedding are fed into API module to achieve attentive interaction respectively. After interaction, the obtained representations are concatenated to merge information in the pooling layer. The rating prediction is finally completed in the output layer.

3.1. Persona-driven personalized sentimental interactive representation learning

As stated in Section 1, the study of personality has been successfully applied to various scenarios. Inspired by these studies, in this work, we argue that the individual reviewing behavior is closely related to personality, and sentimental expression is a manifestation of personality. The motivation of this paper is to explore a persona-driven personalized sentimental representation method for an effective review-based recommendation. Intuitively, the personalized sentimental expression habits of users expressing opinions, attitudes and opinions in reviews mainly reflect in word-level and sentence-level two aspects. At the level of words, the expression behavior of different users is reflected in the usage of words. At the sentence level, different users' sentimental expression behavior is reflected in the overall sentimental expression tone.

To capture personalized sentimental expression from textual reviews, we learn sentimental word usage habits and overall tone styles in word-level and sentence-level respectively. We further combine them to achieve persona-driven personalized sentimental representation. Word2vec is used to pre-train word-level embeddings and BERT is adopted to pre-train sentence-level embeddings from reviews. API module is designed to capture similar or synonymous personalized sentimental expressions.

3.1.1. Attentive persona-driven interaction

We take sentence-level feature (Seq_u, Seq_v) as an example to introduce our API module, and the word-level feature is processed in the same way. We define Mat_{train} as a trainable matrix and to customize the pre-trained sentence-level embedding e. We normalize these customized embeddings as follows.

$$z = \frac{Mat_{tra}e}{\|Mat_{tra}e\|}, \tag{1}$$

where the denominator denotes the Euclidean norm. To discover sentimental expression habits, we define interaction between two features as the element-wise product (denoted as "·") of their embedding vectors. According to the interaction calculation of different features, we can identify the synonymous or similar sentimental expression. We then project the vector representation of sentence-level features, the sentence-level interaction between $i \in S_u$ and $j \in S_v$ is defined as the element-wise product of their embedding vector z_i and z_j respectively. Therefore, we represent the results of the sentence-level feature interaction as follows.

$$F(u,v) = \{z_i \cdot z_j(x_i x_j)\}_{i\in S_u, j\in S_v}, \tag{2}$$

where $x_i \in \{0,1\}$ is the masking indicator, if the value of x_i is zero and it is the meaningless feature. To purify the interaction and avoid noisy interactions, we give more attention to some important interactions than others. Thus, we design a neural attention network to focus on important interactions. The attention network is defined as:

$$\begin{aligned} \tilde{\ell}_{ij} &= \mathrm{H}_{att1}{}^{T}\mathrm{ReLU}(u,v), \\ \ell_{ij} &= \mathrm{softmax}(\tilde{\ell}_{ij}) = \frac{exp(\tilde{\ell}_{ij})}{\sum_{j'\in S_v} exp(\tilde{\ell}_{ij'})}, \end{aligned} \tag{3}$$

where H_{att1} is a learnable vector, and ℓ_{ij} is the value of the interaction between i and j by this attention network. ℓ is used as the importance weight as follows:

$$\Gamma_i = \sum_{j\in S_v} \ell_{ij}(z_i \cdot z_j(x_i x_j)). \tag{4}$$

To obtain a user's most concerned item features, we estimate the user's preferences towards different item features by assigning weights. $\Lambda_{v,i}$ denotes the overall similarity. We calculate the importance of each item feature as:

$$\Lambda_{v,i} = z_i \cdot \textstyle\sum_{j\in S_v} z_j. \tag{5}$$

To measure the importance of different item features, and the attention layer is defined as follows.

$$\begin{aligned} \tilde{\eta}_{uvi} &= \mathrm{H}_{att2}{}^{T}\Lambda_{v,i}, \\ \eta_{uvi} &= \mathrm{softmax}(\tilde{\eta}_{uvi}) = \frac{exp(\tilde{\eta}_{uvi})}{\sum_{i'\in S_u} exp(\tilde{\eta}_{uvi'})}, \end{aligned} \tag{6}$$

where H_{att2} is a learnable vector and $\eta_{u,v,i}$ denotes the importance of each item feature in user u's interest on item v. Observingly, H_{att1} and H_{att2} are two different vectors by the different attention network layers. Next, we represent user u's overall satisfaction on item v:

$$d_s(u,v) = \sum_{i\in S_u} \eta_{uvi}\Gamma_i. \tag{7}$$

Now, we have obtained the sentence-level sentimental representation by attentive interaction. Due to the interaction operation, the overall analogous style of sentimental expression is explored. Meanwhile, we consider the word-level usage habits from reviews and extract word-level features to conduct interaction for fine-grained word-level sentimental representation. In the next section, these two representations are combined.

Table 1
Statistical details of our final datasets.

	Luxury beauty	Cell phones and accessories	Industrial and scientific	Clothing, shoes, and jewelry
#Users	3,818	27,879	5,597	39,387
#Items	1,581	10,429	4,076	23,033
#Ratings &Reviews	34,265	194,439	77,063	278,677

3.1.2. Attentive aggregation representation

Similarly (see Section 3.1.1), we also can generate word-level interactive sentimental representation d_a. As for d_a, it is worth mentioning that we extract adjectives from reviews that can reflect sentiment expression obviously. Finally, d_a and d_s are combined to obtain d_{as}. The operation formula is as follows and $Concat$ denotes concatenation.

$$d_{as} = Concat\left(d_a, d_s\right). \tag{8}$$

As illustrated in Fig. 3, it is the output of interactive sentimental representation. By designing two neural attention networks in our API module, we can discover informative feature interactions and pick up the users' most concerned item features respectively. Notably, the sequence-based and fragment-based features are considered jointly to explore sentimental word usage habits and overall tone styles, contributing to the personalized sentimental representation learning.

To further enhance the connotation of the original sentences, we explore a holistic interaction modeling method. The calculation equation is similar to latent factor models. d_{ho} is the output of this holistic interaction modeling part. The formula is as follows. d_{as} and d_{ho} are concatenated as the input of rating prediction.

$$d_{ho}(u, v) = \mathrm{p}_u \cdot \mathrm{q}_v. \tag{9}$$

3.2. Rating prediction

As shown in Fig. 3, we concatenate d_{ho} and d_{as} into one vector to merge information. A regression layer is stacked above it. $\hat{y}(u, v)$ denotes user u's overall satisfaction score on item v.

$$\hat{y}(u, v) = \mathrm{W}_{Rat} \left\langle d_{ho}(u, v), d_{as}(u, v)\right\rangle. \tag{10}$$

Actually, the rating prediction task is a regression problem. In this paper, the standard squared loss is adopted as the loss function and it is a common objective function:

$$Loss_1 = \sum_{u,v\in\tau} (\hat{y}_{u,v} - y_{u,v})^2, \tag{11}$$

where τ represents the set of instances for training, and $y_{u,v}$ is the ground truth by the user u to the item v.

To avoid the overfitting as much as possible, L^2 regularization is adopted on the user and item embedding matrix and kernel matrix of the output layer. The $Loss_2$ is minimized to fit our PSAR from data.

$$Loss_2 = Loss_1 + \varphi * \frac{\sum w^2}{Num_w}, \tag{12}$$

where φ adjusts the L^2 regularization intensity. The numerator denotes the L^2-norm of three matrices and the denominator is the number of elements in the matrix.

4. Experiments

In this section, we first describe our experimental setup including datasets, comparison baselines, evaluation metrics, and experiment details. We then analyze our experimental results and discuss the performance of our model compared to baselines in terms of overall performance, ablation study, interaction and interpretability, and sensitivity study. We have a discussion at the end of this section.

Table 2
Statistics of ratings and reviews on Amazon datasets.

	Dataset 1	Dataset 2	Dataset 3	Dataset 4
#Ratings only	574,628	3,447,249	1,758,333	5,748,920
#5-core	34,278	19,4439	77,071	278,678
#Null	13	0	8	1
#Ratings & Reviews	34,265	194,439	77,063	278,677
Ratio(actual adopted/overall)	5.963%	5.640%	4.382%	4.847%

4.1. Experimental setup

4.1.1. Data analysis

Datasets. In our experiments, we follow review-level state-of-the-art methods and select four publicly accessible datasets from Amazon 5-core[1] to evaluate our model. Here are at least five reviews for each user and item. The record consists of user, item, rating, textual review, and helpfulness votes in the dataset. In our PSAR model, we focus on users, items, textual reviews, supplemented by ratings. We adopt the subset of Amazon datasets, that is, "Luxury Beauty (Dataset 1)", "Cell Phones and Accessories (Dataset 2)", "Industrial and Scientific (Dataset 3)", and "Clothing, Shoes, and Jewelry (Dataset 4)". These datasets with different scales are selected from various domains, which could cover different recommendation scenarios. Among them, "Clothing, Shoes, and Jewelry" is the largest dataset, while "Luxury Beauty" is the smallest one. The statistical details of our final datasets are shown in Table 1.

Statistics of reviews. To further explain the upper bound of the proposed scheme, we study some statistics about the ratio of reviewed items with overall items purchased by users (i.e., rating only), and the length of reviews. As shown in Table 2, statistics of ratings and reviews on Amazon are provided. The number of "ratings only" is the total number of overall items purchased by users. 5-core denotes that all users and items have at least five reviews. We use 5-core with null values removed in our experiment. We can observe that the ratio of our actual adopted reviews (5-core with null value removed) to "rating only" on dataset 1 (Luxury Beauty) is higher than the others. Meanwhile, Fig. 4 shows the length distribution of reviews roughly on four datasets. A number of intervals are divided in terms of the review length (number of words) and we provide a ratio of each interval review with overall reviews as shown in Fig. 5. It is observed that if the length of long text is set to more than 100 words, the ratio of long text to all reviews on the four datasets is 33.3%, 25.1%, 10.7% and 13.9% approximately.

4.1.2. Comparison baselines

To justify the performance of our model, the proposed PSAR compares with seven models, namely BPRMF (Rendle et al., 2009), GMF (He et al., 2017), DeepCoNN (Zheng et al., 2017), NARRE (Chen et al., 2018), AARM (Guan et al., 2019), AARM+, CARL (Wu, Quan et al., 2019), and EDMF (Liu, Zheng, Li, Shen et al., 2022). These baselines are described in detail as follows.

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

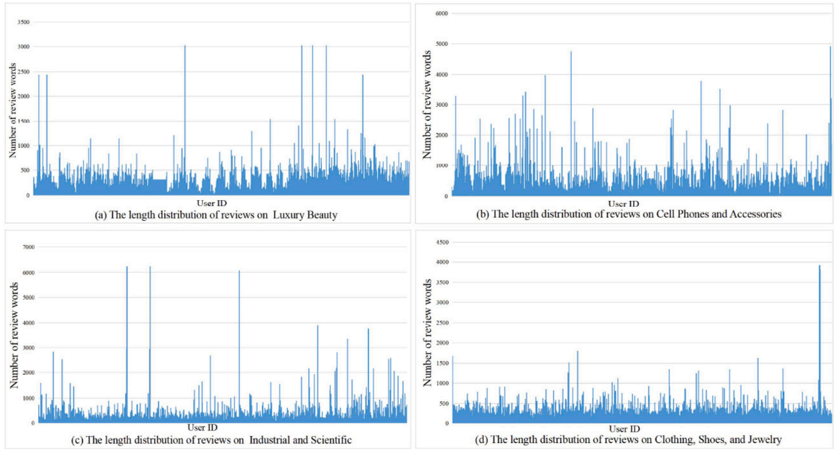


Fig. 4. The length distribution of reviews on four datasets.

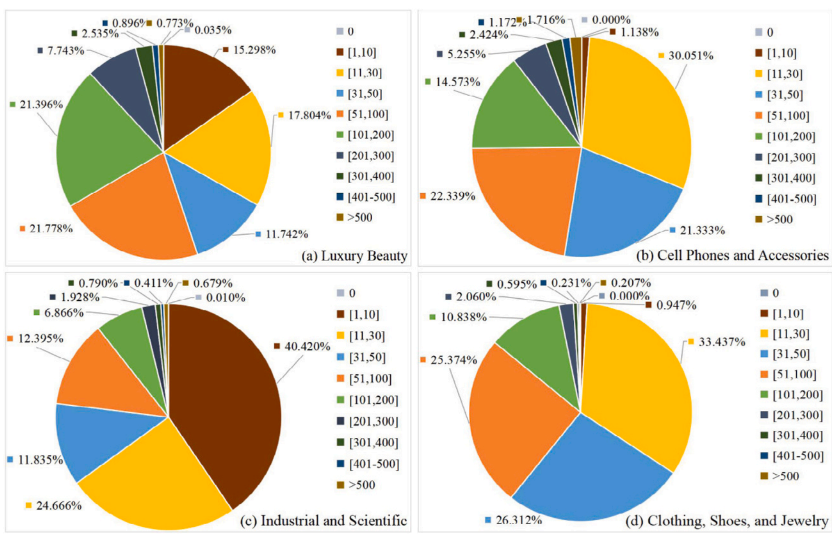


Fig. 5. Statistics of review words on four datasets.

- **BRPMF** (Rendle et al., 2009): This is a matrix factorization method based on Bayesian Personalized Ranking (BPR) loss function. It mainly uses a pair-wise learning way and it is a popular baseline for the top-N recommendation.
- **GMF** (He et al., 2017): GMF is short for generalized matrix factorization and it is one of the state-of-the-art neural networks-based recommendation methods.
- **DeepCoNN** (Zheng et al., 2017): DeepCoNN[2] is the state-of-the-art method, which models user and item jointly from textual reviews. It unitizes review information fully for rating prediction.
- **NARRE** (Chen et al., 2018): NARRE[3] is a rating regression model for explainable recommendations, which adopts attention networks to learn useful reviews for users.
- **CARL** (Wu, Quan et al., 2019): CARL[4] is a context-aware user–item representation learning model for rating prediction. A dynamic linear fusion strategy based on neural attention mechanism is used to explore context-aware representations for each user–item pair in reviews.
- **AARM** (Guan et al., 2019): The attentive aspect-based recommendation model adopts word-level technique to identify the user's interests from reviews. The authors have shown that it can achieve better performance, however, it only uses the review information.
- **AARM**+: To strengthen the performance of AARM, we add rating information to AARM. It is an AARM plus model by considering rating and review. In experiments, we achieve AARM+ model based on the released code for the AARM.[5]
- **EDMF** (Liu, Zheng, Li, Shen et al., 2022): This is an efficient deep matrix factorization with review feature learning recommendation framework. It considers sparsity constraint and mainly uses convolutional neural network and word-level attention to extract interactive features in reviews.

4.1.3. Evaluation metrics

In our experiments, two standard evaluation metrics are adopted to evaluate the performance of all methods, namely Normalized Discounted Cumulative Gain (NDCG) and Hit Ratios (HR). NDCG measures the accuracy of the ranking and HR calculates the actual user preferences hit in the recommended list. The higher the value, the better the

[2] https://github.com/chenchongthu/DeepCoNN
[3] https://github.com/chenchongthu/NARRE
[4] https://github.com/WHUIR/CARL
[5] https://github.com/XinyuGuan01/Attentive-Aspect-based-Recommendation-Model

Table 3
Overall performance comparison results (%) of baselines and our model in terms of NDCG and HR on "Luxury Beauty", "Cell Phones and Accessories", "Industrial and Scientific", and "Clothing, shoes, and jewelry" datasets.

	Luxury beauty		Cell phones and accessories		Industrial and scientific		Clothing, shoes, and jewelry	
	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR
BPRMF (Rendle et al., 2009)	18.486	28.487	1.959	5.351	5.278	15.087	0.601	1.767
GMF (He et al., 2017)	19.280	30.487	2.306	6.327	6.111	17.253	1.144	2.795
DeepCoNN (Zheng et al., 2017)	21.483	32.986	3.636	9.913	11.483	28.986	1.310	3.286
NARRE (Chen et al., 2018)	23.013	33.715	4.117	10.586	12.891	30.029	1.892	4.766
CARL (Wu, Quan et al., 2019)	23.939	34.220	4.277	10.713	13.097	29.330	1.924	4.833
AARM (Guan et al., 2019)	23.869	34.101	4.364	10.940	13.232	30.425	1.956	4.915
AARM+	24.115	34.530	4.476	11.268	13.511	30.686	1.983	4.986
EDMF (Liu, Zheng, Li, Shen et al., 2022)	24.109	34.603	4.525	11.393	14.254	31.658	2.095	5.258
PSAR(ours)	**25.284**	**37.781**	**5.090**	**11.676**	**15.261**	**34.097**	**2.389**	**5.317**

performance. The calculation formulas are as follows.

$$DCG = \sum_{i=1}^{N} \frac{2^{rel_i} - 1}{\log_2(i+1)}, \quad NDCG = \frac{DCG}{IDCG}, \tag{13}$$

where rel_i denotes the graded relevance and IDCG is the maximum DCG under ideal conditions.

$$HR@N = \frac{\#hit@N}{|D_{test}|}, \tag{14}$$

where $\#hit@N$ represents the number of hits and $|D_{test}|$ is the total number of test cases in the test set.

4.1.4. Experiments details

We carry out our method based on Tensorflow.[6] We randomly split the dataset and 80% records are training sets, while the rest of 20% records are put into test sets. We test the batch size of [128, 256, 512] and gain the optimal setting with the batch size of 512. The learning rate is searched in [0.001, 0.003, 0.01] and the optimal learning rate is 0.003. The regularization parameter is searched in [0, 0.1, 0.01,0.001] and it is set to 0.1. The ReLU activation function is employed in our model. For implementation details of baselines, we follow their optimal parameter settings. Besides, we set the dimension to 128 and 768 for the word embedding and sentence embedding respectively and utilize the Adaptive Moment Estimation (Adam) optimizer. For the word-level feature extraction, we focus on extracting adjectives in reviews that can reflect sentiment expression obviously. To avoid overfitting, we apply dropout on the element-wise product module and after the API module, and we set dropout to 0.5. The detailed sensitivity analysis is studied in Section 4.5.

4.2. Overall performance comparison

The overall performance of PSAR and the baselines (see Section 4.1.2) are shown on the top-N recommendation task in terms of NDCG and HR in Table 3. The best results are marked in bold and our PSAR model achieves optimal performance on the selected four datasets. From the results in Table 3, several observations can be reported:

(1) In general, the review-based recommendation methods (DeepCoNN, NARRE, CARL, AARM, AARM+, and PSAR) perform better than collaborative filtering models (BPRMF and GMF) which just use the rating matrix as the input. Obviously, the textual reviews usually contain rich information and fine-grained features. Therefore, the better-quality representation leads to more accurate recommended results.

(2) The attention-based methods (NARRE, CARL, AARM, AARM+, and PSAR) generally outperform non-attention methods (BPRMF, GMF, and DeepCoNN). Thus, it can be seen that different features have different influence weights and all features generally cannot be treated equally. In this case, the attention mechanism can capture the user's various preferences and assign a high weight to the user's most concerning item feature.

(3) As illustrated in Fig. 5, in "Cell Phones and Accessories" and "Clothing, Shoes, and Jewelry" datasets, the proportion of reviews with less than 10 words is much smaller than the other two datasets. Our PSAR achieves obvious improvement in terms of NDCG on "Cell Phones and Accessories" and "Clothing, Shoes, and Jewelry" datasets. The sentence-level technique of our model contributes to the performance improvement.

(4) The AARM+ model gains better results than AARM model on the four datasets. AARM+ combines the rating and textual review features to obtain rich information, and the improvement ratio varies for each dataset. As can be seen in Table 3, the maximum improvements achieved by AARM+ model are 2.57% for NDCG, 2.99% for HR compared to AARM model. The results demonstrate the significance of aggregating the ratings and reviews.

(5) From Table 3, generally, our PSAR gets the best results and EDMF achieves suboptimal results. EDMF performs better than AARM, AARM+, and CARL on large-scale datasets. Compared with EDMF, on average, our PASR improves 6.76% and 7.19% for NDCG and HR, respectively. Besides, PASR improves approximately 16.33% and 11.42% in terms of NDCG and HR over CARL model, gets 15.01% (NDCG) and 9.44% (HR) improvement over AARM on average. We hold that the reasons are as follows. Firstly, our PSAR considers the ratings to obtain informative features for representation. Secondly, we employ the sentence-level technique to exact the original sentence feature in textual reviews for sentimental representation. Meanwhile, we combine word-level and sentence-level reviews to capture personalized sentimental expression for accurate and informative representation. Therefore, the experimental results show the high accuracy of the proposed PSAR on all datasets.

4.3. Ablation study

In this subsection, we conduct an ablation study to explain how different components in our PSAR model contribute to the overall performance and further justify the effectiveness of our model designs. The proposed PSAR merges the ratings and reviews and adopts word-level and sentence-level interactions to complete personalized sentimental representation. We analyze our model and compare it to the following four variants.

- **PSAR-W**: This variant only employs word-level technology and model users' preferences from the textual reviews. It is similar to AARM model and does not utilize rating information.
- **PSAR-WR**: In contrast to PSAR-W, we add rating features to PSAR-WR. It not only considers the ratings and reviews, but also utilizes word-level representation for users and items. It is the same as AARM+ model.

[6] https://www.tensorflow.org/

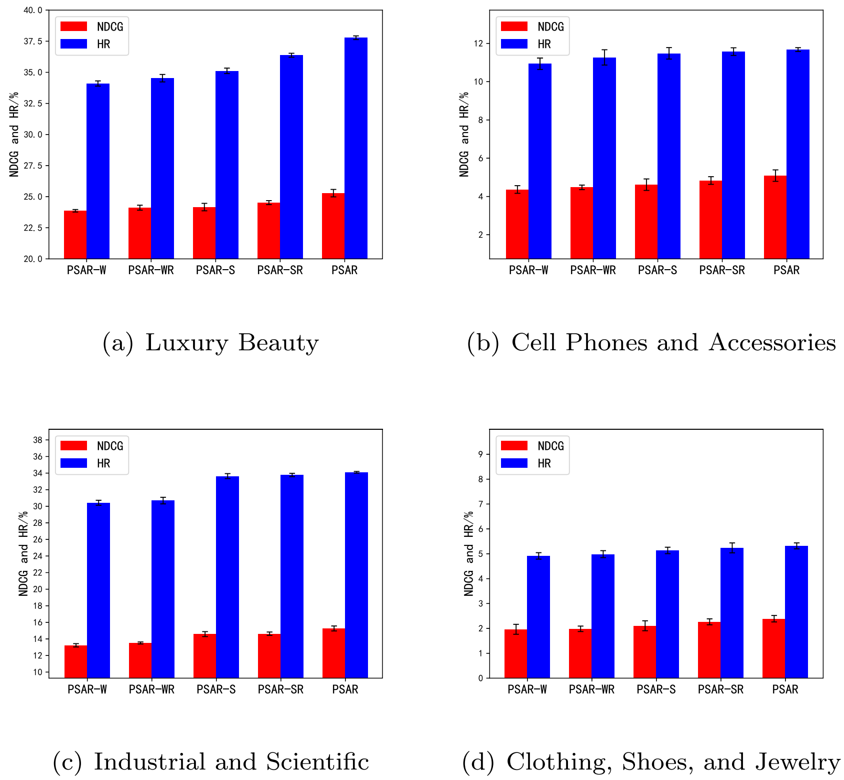


Fig. 6. Performance comparison of our PSAR and its variants in terms of NDCG and HR on the four datasets.

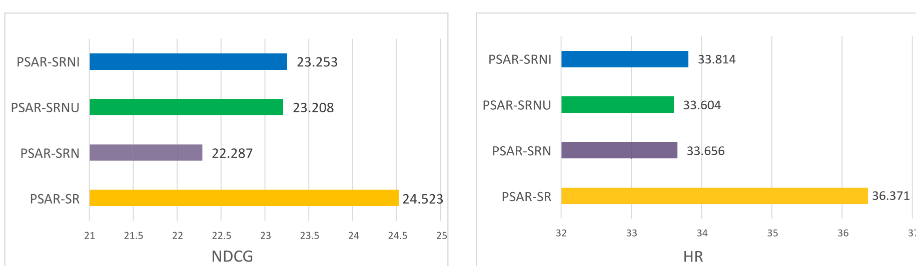


Fig. 7. Interaction evaluation performance on Luxury Beauty dataset.

- **PSAR-S**: From a technical point of view, it uses sentence-level representation to reserve the meaning of the original sentence in reviews. However, this variant does not take advantage of ratings.
- **PSAR-SR**: As the basic variant of PSAR, rather than adopting word-level technique, this method takes into account sentence-level representation and fully exploits the ratings and reviews in records.

The results of the ablation study on all evaluation protocols from four datasets are shown in Fig. 6. We can make the following observations. First, concerning results in Fig. 6, if we do not consider ratings, namely PSAR-W and PSAR-S, we can find that PSAR-S gets better results than PSAR-W, and the average improvements achieved by PSAR-S are 6.12% for NDCG. Besides, PSAR-S gains 5.74% improvement on average for HR. Second, PSAR-SR outperforms PSAR-WR for NDCG and HR, and increases 8.06% and 5.77% improvement on average respectively. Third, PSAR integrates the above variants and achieves the best performance. The average improvement is 4.54% for NDCG and 1.83% for HR compared to the best variant PSAR-SR. Accordingly, as the integrated model, the proposed PSAR improves the recommendation performance further by exploiting reviews and ratings. More importantly, PSAR can learn users' expression habits based on fragmentary (word-level) and sequential (sentence-level) features simultaneously for more effective recommendations.

4.4. Interaction and interpretability

To analyze the interaction performance, we perform non-interaction experiments on "Luxury Beauty" dataset as shown in Fig. 7. In this subsection, because of the ingenuity of our sentence-level PSAR-SR method, we compare it with PSAR-SRN (non-interaction). The results demonstrate that PSAR-SR outperforms PSAR-SRN for NDCG and HR. In the meantime, we model the user and item separately without considering the interaction, named PSAR-SRNU and PSAR-SRNI. As illustrated in Fig. 7, we can see that the results of PSAR-SRNU and PSAR-SRNI are pimping, however, the value of PSAR-SRN is lower than them. It shows that the interaction between different features is indispensable for the review-based recommendation in line with our interactive personalized sentimental representation.

Reviews are usually used to express users' opinions, sentiments, and attitudes, which can provide constructive information and suggestions to help other users make decisions. The proposed PSAR model simultaneously learns the weight of each sentimental feature from reviews. In addition, our interactive sentimental representation modeling is proposed driven by personality to further understand user's behaviors. In our PSAR model, a user's sentimental expression is decomposed into sequence-based and fragment-based features and it usually stays

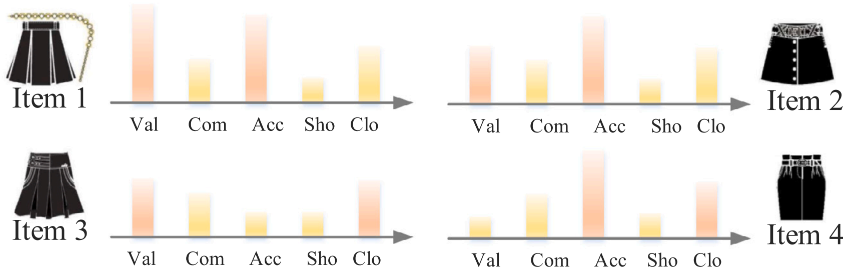


Fig. 8. Interpretation for why “user 972” rated “item 1”,“item 2”,“item 3”, and “item 4” with 5, 3, 2, and 2, respectively, from Clothing.

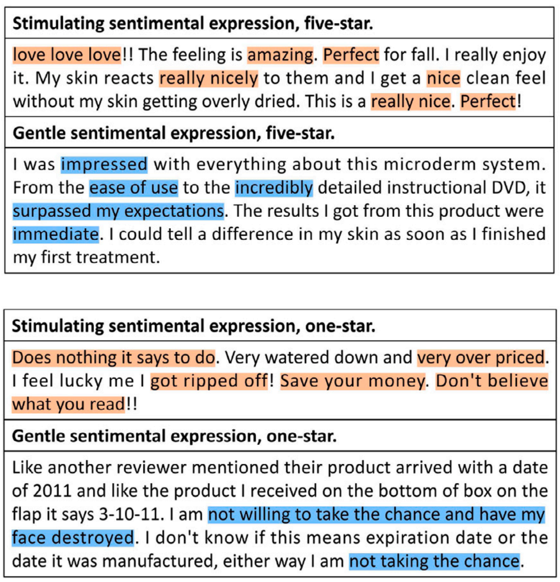


Fig. 9. Sentimental expression cases of different users in reviews on Luxury Beauty dataset. Orange and blue represent stimulating and gentle expressions respectively.

Table 4
Top five words based on word-level feature for a user (index 972) from Clothing. Each column is the related “interpretation” label.

Value	Comfort	Accessories	Shoes	Clothing
Price	Size	Ring	Socks	Shirt
Quality	Fit	Pretty	Foot	Black
Worth	Comfortable	Dress	Boots	Top
Fits	Small	Beautiful	Comfort	Feel
Cute	Bra	Earrings	Sandals	Soft

unchanged over some time. For example, as shown in Table 4, the top five words of user 972 from “Clothing” dataset are discovered. According to our sentimental expression habits analysis, the target user focuses on “Value” and “Accessories”. From the results in Fig. 8, item 1 gets a high score on “Value” and “Accessories” than other items, and it provides the interpretation for why user 972 rated item 1, item 2, item 3, and item 4 with 5, 3, 2, and 2 respectively. We use the abbreviation Val, Com, Acc, Sho, and Clo to denote Value, Comfort, Accessories, Shoes, and Clothing respectively in Fig. 8.

Case analysis. To further understand the interpretability of our persona-driven personalized sentimental representation, in Fig. 9, several cases are selected for sentimental expression of different users in five-star and one-star reviews on “Luxury Beauty” dataset. The word-level sentimental representation modeling of our PSAR can achieve the mining of sentimental language usage habits. The overall sentimental expression habit can be explored by the sentence-level sentimental representation modeling of our PSAR. These highlighted words or sentences in Fig. 9 are in line with our personalized sentimental representation design.

4.5. Sensitivity study

Regularization parameter. As mentioned before, φ is used to adjust L^2 regularization intensity. [0,0.1,0.01,0.001] is tested in our experiment and by observing the results in Fig. 10, when φ is equal to 0, 0.1, 0.01, and 0.001 respectively, the difference in results was relatively small. The robustness of the parameter can be verified on the four datasets. Considering the optimal results on different datasets, the value of φ is set to 0.1 in our model.

Dropout ratio. To avoid overfitting, the dropout strategy is employed on the element-wise product module and after the API module. Fig. 11 shows the performance of our PSAR model for different dropout ratios in NDCG and HR on the four datasets. It can be seen that when the dropout ratio equals to 0.5, our PSAR performs better on the four datasets. Generally speaking, as the size of the dataset increases, the robustness of dropout becomes better.

Activation function. We take “Luxury Beauty” dataset as an example and in the last layer of our model, the results using different activation functions (ReLU and Sigmoid) are compared. The best experimental results with 600 epochs are as follows. When Sigmoid activation function is used, the value of NDCG is 24.912 and the value of HR is 37.185. When ReLU activation function is used, the value of NDCG is 25.284 and the value of HR is 37.781. It is suitable for our PSAR to use ReLU activation function than Sigmoid. Therefore, ReLU is adopted in our model.

Convergence analysis. As shown in Fig. 12, we record the value of training loss along with different epochs using the optimal parameter setting on “Luxury Beauty”, “Cell Phones and Accessories”, “Industrial and Scientific”, and “Clothing, Shoes, and Jewelry” datasets. Several observations can be shown from the convergence results. First, with more epochs, the training loss of our PSAR gradually declines and the proposed model improves the recommendation performance on all datasets by learning effective representation. These results reveal the rationality of our learning scheme. Second, when the epochs reach about 200, the PSAR model has converged to some extent on “Luxury Beauty” and “Cell Phones and Accessories” datasets. For the other two datasets, when the epochs reach about 300, the results converge. Third, compared with (a) and (b) in Fig. 12, Fig. 12(a) converges faster and Fig. 12(b) learns more stably. Lastly, for datasets with different scales, there are some differences in the trend of training loss. The convergence result of “Luxury Beauty” dataset is better than the other three datasets.

4.6. Discussion

Overall, by our experimental results and analysis, the proposed PSAR achieves the best recommendation performance compared with

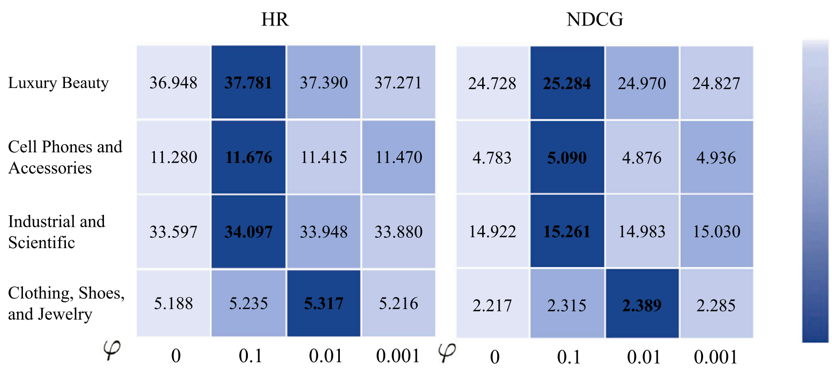


Fig. 10. Regularization parameter discussion of our PSAR. The robustness of the parameter can be confirmed on the four datasets.

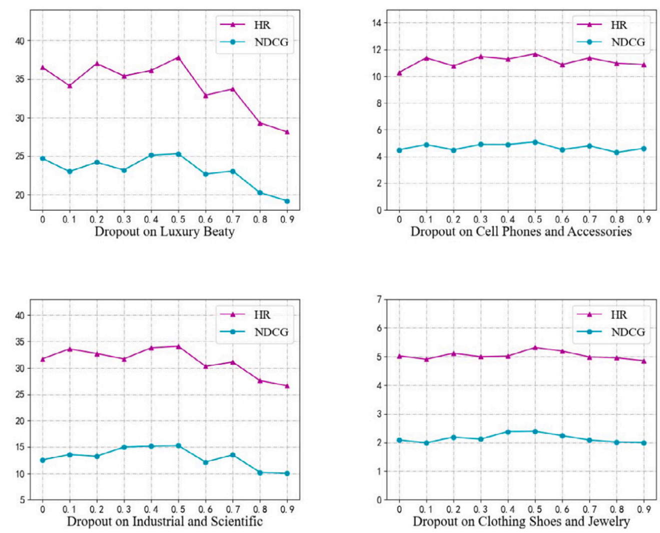
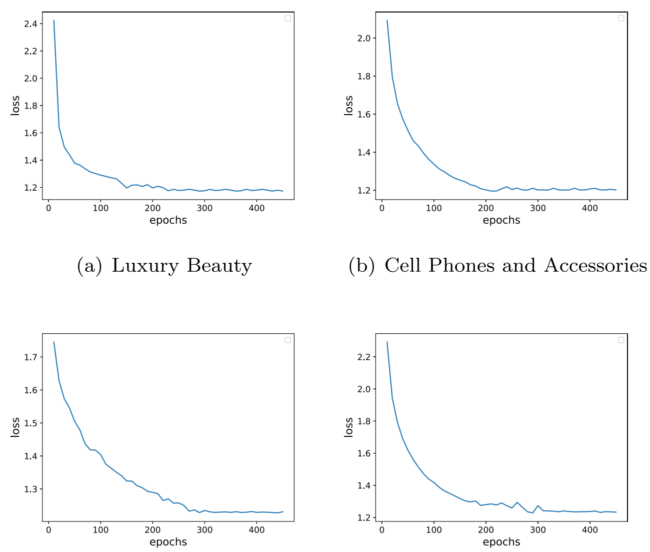


Fig. 11. Performance of our PSAR w.r.t. the dropout ratio in terms of HR and NDCG on the four datasets.



(a) Luxury Beauty (b) Cell Phones and Accessories

(c) Industrial and Scientific (d) Clothing, Shoes, and Jewelry

Fig. 12. Training loss of our PSAR w.r.t. the number of epochs on the four datasets.

all baselines on our selected datasets. The discussion of our model can be outlined as follows:

(1) Our PSAR is driven by personality, that is, different personality traits exhibit the difference of reviewing expression habits in people, providing sufficient design guideline study.

(2) The persona-driven personalized sentimental representation is developed from word-level and sentence-level. Through the modeling of review information at these two levels, users' sentimental expression habits in reviewing can be explored.

(3) Compared with the recent AARM, our PSAR achieves 15.01% (NDCG) and 9.44% (HR) improvement on average. The experimental results and analysis further match our motivation and verify the feasibility and effectiveness of the model design.

5. Conclusion and future work

We have presented a novel Persona-driven Sentimental Attentive Recommendation model (PSAR) via personalized sentimental representation learning. As a core component, the personalized sentimental representation captures sentimental expression habits effectively by employing the interaction of similar and synonymous words or phrases, as well as sentences with the analogous sentimental tone. Experimental results on the public datasets show that our model greatly improves the recommendation accuracy compared with the state-of-the-art methods. With the explored multi-level sentimental expression from reviews and estimated feature weights, the proposed model can provide interpretable recommendation results. In the future, we will try to classify the personality types and integrate the classified personality characteristics into the model to achieve neural personality mining for the review-based recommendation.

CRediT authorship contribution statement

Peipei Wang: Methodology, Software, Investigation, Resources, Writing – review & editing. **Lin Li:** Supervision, Conceptualization, Formal analysis, Resources. **Ru Wang:** Methodology, Investigation, Formal analysis. **Xinhao Zheng:** Resources, Software. **Jiaxi He:** Supervision, Writing – review & editing. **Guandong Xu:** Supervision, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is supported by Key Research and Development Project of Hubei Province (No. 2021BAA030) and Youth Project of National Natural Science Foundation of China (No. 62106070).

References

Bao, Y., Fang, H., & Zhang, J. (2014). TopicMF: Simultaneously exploiting ratings and reviews for recommendation. In *Proceedings of the twenty-eighth conference on artificial intelligence* (pp. 2–8). AAAI Press.

Chen, Y. (2011). Interface and interaction design for group and social recommender systems. In *Proceedings of the conference on recommender systems* (pp. 363–366). ACM.

Chen, C., Zhang, M., Liu, Y., & Ma, S. (2018). Neural attentional rating regression with review-level explanations. In *Proceedings of the 2018 world wide web conference on world wide web* (pp. 1583–1592). ACM.

Chin, J. Y., Zhao, K., Joty, S. R., & Cong, G. (2018). ANR: aspect-based neural recommender. In *Proceedings of the 27th ACM international conference on information and knowledge management* (pp. 147–156). ACM.

Chong, X., Li, Q., Leung, H., Men, Q., & Chao, X. (2020). Hierarchical visual-aware minimax ranking based on co-purchase data for personalized recommendation. In *Proceedings of the world wide web conference on world wide web* (pp. 2563–2569). ACM / IW3C2.

Dhelim, S., Ning, H., Aung, N., Huang, R., & Ma, J. (2021). Personality-aware product recommendation system based on user interests mining and metapath discovery. *IEEE Transactions on Computational Social Systems*, *8*(1), 86–98.

Ebesu, T., Shen, B., & Fang, Y. (2018). Collaborative memory network for recommendation systems. In *Proceedings of the 41st international conference on research & development in information retrieval* (pp. 515–524). ACM.

Feng, H., & Qian, X. (2013). Recommendation via user's personality and social contextual. In *Proceedings of the 22nd international conference on information and knowledge management* (pp. 1521–1524). ACM.

Guan, X., Cheng, Z., He, X., Zhang, Y., Zhu, Z., Peng, Q., & Chua, T. (2019). Attentive aspect modeling for review-aware recommendation. *ACM Transactions on Information Systems*, *37*(3), 28:1–28:27.

He, X., Chen, T., Kan, M., & Chen, X. (2015). TriRank: Review-aware explainable recommendation by modeling aspects. In *Proceedings of the 24th international conference on information and knowledge management* (pp. 1661–1670). ACM.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. (2017). Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web* (pp. 173–182). ACM.

Hou, Y., Yang, N., Wu, Y., & Yu, P. S. (2019). Explainable recommendation with fusion of aspect information. *World Wide Web*, *22*(1), 221–240.

Hu, R., & Pu, P. (2011). Enhancing collaborative filtering systems with personality information. In *Proceedings of the conference on recommender systems* (pp. 197–204). ACM.

Klec, M. (2017). The influence of listener personality on music choices. *Computer Science*, *18*(2), 163–178.

Lei, X., Qian, X., & Zhao, G. (2016). Rating prediction based on social sentiment from textual reviews. *IEEE Transactions Multimedia*, *18*(9), 1910–1921.

Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., & Dolan, W. B. (2016). A persona-based neural conversation model. In *Proceedings of the 54th annual meeting of the association for computational linguistics* (pp. 994–1003). The Association for Computer Linguistics.

Li, D., Liu, H., Zhang, Z., Lin, K., Fang, S., Li, Z., & Xiong, N. N. (2021). CARM: confidence-aware recommender model via review representation learning and historical rating behavior in the online platforms. *Neurocomputing*, *455*, 283–296.

Liang, B., Du, J., Xu, R., Li, B., & Huang, H. (2019). Context-aware embedding for targeted aspect-based sentiment analysis. In *Proceedings of the 57th conference of the association for computational linguistics* (pp. 4678–4683). Association for Computational Linguistics.

Lin, Q., Jiayu, C., Jonathan E., R., & Jiahui, L. (2019). Personality predicts words in favorite songs. *Journal of Research in Personality*, *78*, 25–35.

Ling, G., Lyu, M. R., & King, I. (2014). Ratings meet reviews, a combined approach to recommend. In *Proceedings of the eighth conference on recommender systems* (pp. 105–112). ACM.

Liu, H., Zheng, C., Li, D., Shen, X., Lin, K., Wang, J., Zhang, Z., Zhang, Z., & Xiong, N. N. (2022). EDMF: Efficient deep matrix factorization with review feature learning for industrial recommender system. *IEEE Transactions on Industrial Informatics*, 1–11. http://dx.doi.org/10.1109/TII.2021.3128240.

Liu, H., Zheng, C., Li, D., Zhang, Z., Lin, K., Shen, X., Xiong, N. N., & Wang, J. (2022). Multi-perspective social recommendation method with graph representation learning. *Neurocomputing*, *468*, 469–481.

Luo, W., Zhuang, F., Cheng, X., He, Q., & Shi, Z. (2014). Ratable aspects over sentiments: Predicting ratings for unrated reviews. In *Proceedings of the international conference on data mining* (pp. 380–389). IEEE Computer Society.

Ning, H., Dhelim, S., & Aung, N. (2019). PersoNet: Friend recommendation system based on big-five personality traits and hybrid filtering. *IEEE Transactions on Computational Social Systems*, *6*(3), 394–402.

Ren, Z., Liang, S., Li, P., Wang, S., & de Rijke, M. (2017). Social collaborative viewpoint regression with explainable recommendations. In *Proceedings of the tenth international conference on web search and data mining* (pp. 485–494). ACM.

Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2009). BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence* (pp. 452–461). AUAI Press.

Shen, T., Jia, J., Li, Y., Ma, Y., Bu, Y., Wang, H., Chen, B., Chua, T., & Hall, W. (2020). PEIA: personality and emotion integrated attentive model for music recommendation on social media platforms. In *Proceedings of the thirty-fourth AAAI conference on artificial intelligence* (pp. 206–213). AAAI Press.

Shen, X., Yi, B., Liu, H., Zhang, W., Zhang, Z., Liu, S., & Xiong, N. (2021). Deep variational matrix factorization with knowledge embedding for recommendation system. *IEEE Transactions on Knowledge and Data Engineering*, *33*(5), 1906–1918.

Shu, J., Shen, X., Liu, H., Yi, B., & Zhang, Z. (2018). A content-based recommendation algorithm for learning resources. *Multimedia Systems*, *24*(2), 163–173.

Siddique, F. B., Bertero, D., & Fung, P. (2019). GlobalTrait: Personality alignment of multilingual word embeddings. In *Proceedings of the thirty-third AAAI conference on artificial intelligence* (pp. 7015–7022). AAAI Press.

Sun, X., Liu, B., Meng, Q., Cao, J., Luo, J., & Yin, H. (2020). Group-level personality detection based on text generated networks. *World Wide Web*, *23*(3), 1887–1906.

Tewari, A. S., & Barman, A. G. (2018). Sequencing of items in personalized recommendations using multiple recommendation techniques. *Expert Systems with Applications*, *97*, 70–82.

Wang, X., He, X., Feng, F., Nie, L., & Chua, T. (2018). TEM: tree-enhanced embedding model for explainable recommendation. In *Proceedings of the 2018 world wide web conference on world wide web* (pp. 1543–1552). ACM.

Wu, J., Li, X., Chiclana, F., & Yager, R. R. (2019). An attitudinal trust recommendation mechanism to balance consensus and harmony in group decision making. *IEEE Transactions on Fuzzy Systems*, *27*(11), 2163–2175.

Wu, L., Quan, C., Li, C., & Ji, D. (2018). PARL: let strangers speak out what you like. In *Proceedings of the 27the international conference on information and knowledge management* (pp. 677–686). ACM.

Wu, L., Quan, C., Li, C., Wang, Q., Zheng, B., & Luo, X. (2019). A context-aware user-item representation learning for item recommendation. *ACM Transactions on Information Systems*, *37*(2), 22:1–22:29.

Xu, Y., Yang, Y., Han, J., Wang, E., Zhuang, F., Yang, J., & Xiong, H. (2019). NeuO: Exploiting the sentimental bias between ratings and reviews with neural networks. *Neural Networks*, *111*, 77–88.

Xue, F., He, X., Wang, X., Xu, J., Liu, K., & Hong, R. (2019). Deep item-based collaborative filtering for Top-N recommendation. *ACM Transactions on Information Systems*, *37*(3), 33:1–33:25.

Yakhchi, S., Ghafari, S. M., & Beheshti, A. (2018). CNR: cross-network recommendation embedding user's personality. In *Proceedings of the 19th international conference on web information systems engineering* (pp. 62–77). Springer.

Yang, H., & Huang, Z. (2020). Mining personality traits from social messages for game recommender systems. *Knowledge-Based Systems*, *165*, 157–168.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363–373.

Yi, B., Shen, X., Liu, H., Zhang, Z., Zhang, W., Liu, S., & Xiong, N. (2019). Deep matrix factorization with implicit feedback embedding for recommendation system. *IEEE Transactions on Industrial Informatics*, *15*(8), 4591–4601.

Zhang, Y., Lai, G., Zhang, M., Zhang, Y., Liu, Y., & Ma, S. (2014). Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *Proceedings of the 37th international conference on research and development in information retrieval* (pp. 83–92). ACM.

Zheng, L., Noroozi, V., & Yu, P. S. (2017). Joint deep modeling of users and items using reviews for recommendation. In *Proceedings of the tenth international conference on web search and data mining* (pp. 425–434). ACM.