

Article

Review Rating Prediction Based on User Context and Product Context

Bingkun Wang ¹, Shufeng Xiong ^{2,*}, Yongfeng Huang ³ and Xing Li ³

¹ School of Computer, Pingdingshan University, Pingdingshan 467000, China; bkwang77@163.com

² School of Information Engineering, Zhengzhou University, Zhengzhou 450000, China

³ Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China; yfhuang@tsinghua.edu.cn (Y.H.); xing@cernet.edu.cn (X.L.)

* Correspondence: xsf@whu.edu.cn; Tel.: +86-159-3753-5000

Received: 28 August 2018; Accepted: 1 October 2018; Published: 9 October 2018



Abstract: With the explosion of online user reviews, review rating prediction has become a research focus in natural language processing. Existing review rating prediction methods only use a single model to capture the sentiments of review texts, ignoring users who express the sentiment and products that are evaluated, both of which have great influences on review rating prediction. In order to solve the issue, we propose a review rating prediction method based on user context and product context by incorporating user information and product information into review texts. Our method firstly models the user context information of reviews, and then models the product context information of reviews. Finally, a review rating prediction method that is based on user context and product context is proposed. Our method consists of three main parts. The first part is a global review rating prediction model, which is shared by all users and all products, and it can be learned from training datasets of all users and all products. The second part is a user-specific review rating prediction model, which represents the user's personalized sentiment information, and can be learned from training data of an individual user. The third part is a product-specific review rating prediction model, which uses training datasets of an individual product to learn parameter of the model. Experimental results on four datasets show that our proposed methods can significantly outperform the state-of-the-art baselines in review rating prediction.

Keywords: sentiment classification; review texts; user-specific model; product-specific model

1. Introduction

The rapid development of Web 2.0 and e-commerce has led to a proliferation in the number of online user reviews. Online reviews contain a wealth of sentiment information that is important for many decision-making processes, such as personal consumption decisions, commodity quality monitoring, and social opinion mining. Mining the sentiment and opinions that are contained in online reviews has become an important topic in natural language processing, machine learning, and Web mining.

The sentiment classification of online reviews is the most fundamental and important work in natural language processing [1,2]. At present, the review sentiment classification (positive and negative) has been widely studied, but the review sentiment classification cannot meet the demand for fine-grained review sentiment analysis [3–7]. For example, how do consumers choose the most appropriate product from several products that are being reviewed? Horrigand's research shows that consumers are willing to pay an extra 20% to 99% to buy a five-star rated product instead of a four-star rated one [8]. This shows that the nuances of product ratings can lead to dramatic changes

in product sales. In public opinion monitoring, the government not only wants to understand the positive and negative sentiment categories of the reviews, but also wants to further understand the positive and negative sentiment intensity to distinguish the urgency of public opinion events and take different measures. As a result, researchers are paying more and more attention to review rating predictions (RRP).

Existing RRP methods are mainly based on the review content. From the perspective of natural language processing, the review content is converted into a text feature vector, and then a linear regression model is used to model the RRP [9–13]. For example, Qu et al. regarded the RRP as a feature engineering problem, and improved the performance of RRP by extracting different features, such as words, lexical patterns, syntactic structures, and semantic topics from the review content [10]. Zhang et al. used deep CNN on word embedding to learn the features of reviews, which are then fed into a softmax layer to obtain sentiment predictions [13].

Existing RRP methods that are based on review content have an implicit hypothesis: the sentiment polarity and intensity expressed by the same sentiment words is consistent, and the sentiment polarity and intensity expressed by different sentiment words is different. However, this implicit assumption does not hold in the actual scenario. For example, a demanding user gave a five-point rating to the Huawei mobile phone with the sentiment word “good”, while another tolerant user might give a four-star rating with the same sentiment word “good”. Different users use same sentiment words to express different sentiment intensity, which reflects the user context of sentiment word is very important to RRP. For another example, the sentiment word “simple” in the review “this movie is simple” expresses negative sentiment. While in the review “this phone is simple to use”, it expresses positive sentiment. For different products, the same sentiment words are used to express different sentiment polarities, which reflect the product context of sentiment word has a certain impact on RRP. Through the above analysis, it can be found that RRP is not only related to the review text content, but is also related to the user context information and the product context information.

Some studies have shown that user context information is an important data source for online review rating prediction [14–17]. Wang et al. believe that the rating is not entirely determined by the review content, because a demanding user may comment on all products with harsh words, even if he gives a very high rating to the product [14]. Some researchers believe that product context information is another important data source for online review rating prediction. Tang et al. believe that a product has a collection of product-specific words suited to evaluate it [15]. For example, people prefer using “sleek” and “stable” to evaluate a smartphone, while like to use “wireless” and “mechanical” to evaluate a keyboard. Wu et al. believe that the same words may have different or even opposite sentiment meanings in the reviews of different products [16]. For example, “easy” is a positive sentiment in “this digital camera is easy to use”, however, it may have a negative sentiment meaning in “the ending of this movie is easy to guess”.

The main problem with the existing RRP methods that are based on review content is that the user context dependency and product context dependency of the sentiment word cannot be processed very well based only on the review content information. So, we propose a review rating prediction method based on user context and product context (RRP + UC + PC). Our method firstly models the user context dependency of sentiment word, and it proposes a review rating prediction method that combines the review content and the user context. Then, the product context dependency of sentiment words is modeled, and a review rating prediction method based on review content and product context is proposed. Finally, a review rating prediction method that is based on user context and product context is proposed.

The main contributions of this paper can be summarized as:

- (1) We propose a review rating prediction method based on user context and product context.
- (2) Aiming at the problem of user context dependency of sentiment words, a review rating the prediction method based on review content and user context is proposed. Aiming at the problem

of product context dependency of sentiment words, we proposed a review rating prediction method based on review content and product context.

- (3) We conduct comprehensive experiments on four large-scale datasets for RRP. We find that our methods significantly outperform the state-of-the-art baselines.

The remainder of this paper is organized as follows: we describe the related works in Section 2. Our methods are described in Section 3. In Section 4, we describe experimental setup and discuss the performance evaluation results. Section 5 concludes the paper and it outlines future research directions.

2. Related Work

2.1. RRP Based on Review Content

Existing RRP methods mostly focused on review content to mine the sentiment information [10,18–20]. Pang and Lee proposed and studied the RRP problem (the score is generally divided into 1 to 5 points) [18]. Since the ratings have a certain order, the RRP are generally formatted as regression problems. Pang and Lee used SVM regression model and SVM multi-classifier (one-to-many strategy), respectively, for sentiment prediction. The experimental results show that SVM regression model shows better performance than SVM multi-classifier in RRP. Because the numerical score is not a classification attribute value, the classification model is not as effective as the regression model.

Qu et al. proposed a bag-of-opinions review text representation model that is different from the traditional bag-of-words text representation model, and then obtained the sentiment phrases through the bag-of-opinions review text representation model [10]. Each viewpoint is represented as a triple, which is a sentiment word, a modifier, and a negative word (such as in “not very good”, “good” is a sentiment word, “very” is a modifier, and “not” is negative word). For the sentiment classification that only requires positive and negative tendencies, the modifier does not play a key role, but in RRP, both the modifier and the negative word play an important role. In a labeled domain-independent corpus (from multiple domains), a constrained ridge regression model is firstly used to learn each viewpoint feature (sentiment intensity and score), and then extend the traditional unigrams feature through viewpoint features to achieve RRP.

There are also some researches based on review content that take into account the reviewers and items information [14,19]. Wang et al. believe that the rating is not entirely determined by the review content, because a tolerant user may comment on all items with positive words, even if he give a lower rating to the item [14]. Following the work of [14], a method of incorporating user and item into review content is proposed in [19]. Li et al. takes into account the information of user and item when mining the review content, and uses tensor factorization to learn the parameters of the regression model and achieve RRP.

2.2. Missing Rating Prediction in User-Item Rating Matrix

A study that is relevant to RRP is the missing rating prediction for the user-item rating matrix in the recommendation system. The difference between RRP and the missing rating prediction in the user-item rating matrix is that the RRP is based on the review information published by the user to predict the review rating; and the missing rating prediction in the user-item rating matrix is based on history rating that is written by users in the user-item rating matrix to predict the missing rating. Two different types of rating prediction studies are used to achieve rating prediction from different perspectives. Therefore, the prediction method of missing rating in the user-item rating matrix has important reference significance for RRP.

The missing rating prediction methods for the user-item rating matrix in the recommendation system mainly include: K-nearest neighbor (KNN) and matrix factorization (MF). KNN-based rating prediction mainly includes a method that is based on user similarity and method based on item similarity [21,22]. The key idea of KNN-based method is to calculate the similarity between users based on the user-item rating matrix information, and then predict the target user’s missing rating on

the item by the history rating of the K users with the highest similarity to the target user. The idea of KNN-based method is similar to that while using the user similarity method, except that the user is replaced by an item. The key idea of MF is to project users and items into a shared latent factor space, and then models interactions between users and items with the inner item of latent factor vectors of users and items [23,24].

2.3. Review-Based Recommendation

When the user-item rating matrix is sparse, the recommendation performance will be significantly reduced. Therefore, some research work considers review content information to improve missing rating predictions in user-item rating matrix. The effectiveness of using review content information in the recommendation has been extensively discussed and demonstrated in many existing studies [13,25–27].

While considering user review content information, some research works generate latent factors for users and items by integrating topic models in the collaborative filtering framework [28–33]. One of the early studies of using review content to improve missing rating predictions in a user-item rating matrix was proposed in [34]. This work found that reviews often include information, such as price, service, positive, or negative sentiments that can be used for missing rating predictions in use-item rating matrix. Hidden factors as topics were proposed to discover potential aspects from item or user reviews using topic model in [30]. This approach achieved significant improvements over models that only use user-item ratings matrix or review content.

Existing review-based recommendation methods predict the missing ratings in the user-item rating matrix from the history of review texts written by a user. While in our paper, we mainly study the rating prediction of an existing review. Existing review-based recommendation methods provide a common RRP recommendation model for all the users. In contrast, our approach builds a user-specific and product-specific review rating prediction model for each user and each product.

3. RRP Based on User Context and Product Context

3.1. Problem Description

Suppose there is an online review site that contains N items and M users. The M users have published T reviews and corresponding ratings on N items. Our goal is how to effectively predict the rating of reviews by using review content information. That is, we want to find a function f : input of the function f is review content; the output of the function f is review rating.

3.2. RRP Method Based on Review Content and User Context

The review content is the most important source of information for RRP. The existing RRP methods that are based on review content mainly use a vector space model to represent the review content, and then implement RRP through a linear regression model. Specifically, user u has posted a review \mathbf{r}_{ui} on the item i . The existing RRP function based on the review content is as followed:

$$\hat{\vartheta}_{ui} = \mathbf{w}^T \mathbf{r}_{ui} \quad (1)$$

Here, $\hat{\vartheta}_{ui}$ is the predicted rating of user u for item i ; \mathbf{w} is the parameters of the function; and, \mathbf{r}_{ui} is the vector representation of review content.

Due to the differences in sentiment expression of different users in online social media, the general RRP model built for all users cannot accurately understand the special sentiment expression of each user. It is the most intuitive method to design a personalized RRP method for each user by using the personal review content published by each user in the online social media. However, in online social media, the personal review data published by a single user is usually very limited. Therefore, based on the personal review data alone, it is impossible to accurately train a RRP model for each user.

Psychology and sociology studies have shown that, although online users express their sentiments in a personalized way, different online users share many of the same sentiment expressions [35]. For example, “happy” and “good” are often used to express the positive sentiment among different users. Therefore, making full use of the shared sentiment information between different users can help to solve the problem of insufficient data for a single user.

Based on the above analysis, we propose a RRP model that is based on the review content and user context. In order to model the sentiment commonality of different users and sentiment personality of individual user, the RRP model is decomposed into two components, a common one and a user-specific one. The common part is shared by all users to characterize the sentiment information shared by different users and to use all user data for training. The user-specific part is unique to each user and it is used to characterize each user’s specific sentiment expression and to use a single user’s data for learning.

Specifically, user u has posted a review \mathbf{r}_{ui} on the item i . The RRP model based on the review content and user context is as followed:

$$\hat{v}_{ui} = (\mathbf{w} + \mathbf{w}_u)^T \mathbf{r}_{ui} \tag{2}$$

Here, \hat{v}_{ui} is the predicted rating of user u for item i ; \mathbf{w} and \mathbf{w}_u are the common and user-specific parameters in RRP model; and, \mathbf{r}_{ui} is the vector representation of review content.

To calculate the parameter vector \mathbf{w} and \mathbf{w}_u , given the training data set and, we use the least squares error loss principle to minimize the objective function in training data set.

$$\min_{\mathbf{w}, \mathbf{w}_u} \sum_{\text{trainsets}} (v_{ui} - (\mathbf{w} + \mathbf{w}_u)^T \mathbf{r}_{ui})^2 + \lambda (\|\mathbf{w}\|^2 + \|\mathbf{w}_u\|^2) \tag{3}$$

Here, λ is the regular coefficient, $\|\mathbf{w}\|$ and $\|\mathbf{w}_u\|$ are the regular term of the parameter vector \mathbf{w} and \mathbf{w}_u . To estimate the parameter vector \mathbf{w} and \mathbf{w}_u , we use a stochastic gradient descent algorithm to solve this optimization problem. We use the following update rules to learn the parameters \mathbf{w} and \mathbf{w}_u .

$$\mathbf{w} \leftarrow \mathbf{w} + \eta (\varepsilon_{ui} \mathbf{r}_{ui} - \lambda \mathbf{w}) \tag{4}$$

$$\mathbf{w}_u \leftarrow \mathbf{w}_u + \eta (\varepsilon_{ui} \mathbf{r}_{ui} - \lambda \mathbf{w}_u) \tag{5}$$

Here, $\varepsilon_{ui} = v_{ui} - (\mathbf{w} + \mathbf{w}_u)^T \mathbf{r}_{ui}$, η is learning rate. After getting \mathbf{w} and \mathbf{w}_u , given reviews in test datasets, we can use $\hat{v}_{ui} = (\mathbf{w} + \mathbf{w}_u)^T \mathbf{r}_{ui}$ to predict the review rating.

3.3. RRP Based on Review Content and Product Context

Due to the differences in sentiment expression of different products in reviews, the general RRP model that is built for all products cannot accurately understand the special sentiment expression of each product. It is the most intuitive method to design a personalized RRP method for each product by using the personal review content published for each product. However, the personal review data published for a single product is usually very limited. Therefore, based on the personal review data alone, it is impossible to accurately train a RRP model for each product.

Although each product has its own unique sentiment expression, different products have many identical sentiment expressions. For example, general sentiment words such as “perfect”, “excellent”, and “bad” often express a constant sentiment tendency in different products. Therefore, making full use of the shared sentiment information between different products can help to solve the problem of insufficient data for a single product.

Based on the above analysis, we propose a RRP model based on the review content and product context. In order to model the sentiment commonality of different products and sentiment specific of individual product, the RRP model is decomposed into two components, a common one and a product-specific one. The common part is shared by all products to characterize the sentiment

information shared by different products and to use all product data for training. The product-specific part is unique to each product and it is used to characterize each product-specific sentiment expression and to use a single product’s data for learning.

Specifically, user u has posted a review \mathbf{r}_{ui} on the item i . The RRP model that is based on the review content and product context is as followed:

$$\hat{v}_{ui} = (\mathbf{w} + \mathbf{w}_i)^T \mathbf{r}_{ui} \tag{6}$$

Here, \hat{v}_{ui} is the predicted rating of user u for item i ; \mathbf{w} and \mathbf{w}_i are the common and product-specific parameters in RRP model; and, \mathbf{r}_{ui} is the vector representation of review content.

To calculate the parameter vector \mathbf{w} and \mathbf{w}_i , given the training data set and, we use the least squares error loss principle to minimize the objective function in training data set.

$$\min_{\mathbf{w}, \mathbf{w}_i} \sum_{\text{trainsets}} (v_{ui} - (\mathbf{w} + \mathbf{w}_i)^T \mathbf{r}_{ui})^2 + \lambda (\|\mathbf{w}\|^2 + \|\mathbf{w}_i\|^2) \tag{7}$$

Here, λ is the regular coefficient, $\|\mathbf{w}\|$ and $\|\mathbf{w}_i\|$ are the regular term of the parameter vector \mathbf{w} and \mathbf{w}_i . To estimate the parameter vector \mathbf{w} and \mathbf{w}_i , we use a stochastic gradient descent algorithm to solve this optimization problem. We use the following update rules to learn the parameters \mathbf{w} and \mathbf{w}_i .

$$\mathbf{w} \leftarrow \mathbf{w} + \eta(\varepsilon_{ui} \mathbf{r}_{ui} - \lambda \mathbf{w}) \tag{8}$$

$$\mathbf{w}_i \leftarrow \mathbf{w}_i + \eta(\varepsilon_{ui} \mathbf{r}_{ui} - \lambda \mathbf{w}_i) \tag{9}$$

Here, $\varepsilon_{ui} = v_{ui} - (\mathbf{w} + \mathbf{w}_i)^T \mathbf{r}_{ui}$, η is learning rate. After getting \mathbf{w} and \mathbf{w}_i , given reviews in test datasets, we can use $\hat{v}_{ui} = (\mathbf{w} + \mathbf{w}_i)^T \mathbf{r}_{ui}$ to predict the review rating.

3.4. RRP Based on User Context and Product Context

In Section 3.1, we present our research goals through a toy example. In Section 3.2, we present a RRP model based on the review content and user context. In Section 3.3, we present a RRP model that is based on the review content and product context. In this part, we propose a RRP method based on user context and product context by integrating the user context information and the product context information.

$$\hat{v}_{ui} = (\mathbf{w} + \mathbf{w}_u + \mathbf{w}_i)^T \mathbf{r}_{ui} \tag{10}$$

Here, \hat{v}_{ui} is the predicted rating of user u for item i ; \mathbf{w} , \mathbf{w}_u , \mathbf{w}_i are the common, user-specific and product-specific parameters in RRP model; and, \mathbf{r}_{ui} is the vector representation of review content.

In order to get optimum parameters \mathbf{w} , \mathbf{w}_u , \mathbf{w}_i , given training datasets available, we use the least square error loss principle to minimize the objective function.

$$\min_{\mathbf{w}, \mathbf{w}_u, \mathbf{w}_i} \sum_{\text{trainsets}} (v_{ui} - (\mathbf{w} + \mathbf{w}_u + \mathbf{w}_i)^T \mathbf{r}_{ui})^2 + \lambda (\|\mathbf{w}\|^2 + \|\mathbf{w}_u\|^2 + \|\mathbf{w}_i\|^2) \tag{11}$$

Here, λ is the regular coefficient, $\|\mathbf{w}\|$, $\|\mathbf{w}_u\|$, and $\|\mathbf{w}_i\|$ are the regular term of the parameter. To estimate the parameter \mathbf{w} , \mathbf{w}_u , and \mathbf{w}_i , we use a stochastic gradient descent algorithm to solve this optimization problem. We use the following update rules to learn the parameters \mathbf{w} , \mathbf{w}_u , and \mathbf{w}_i .

$$\mathbf{w} \leftarrow \mathbf{w} + \eta(\varepsilon_{ui} \mathbf{r}_{ui} - \lambda \mathbf{w}) \tag{12}$$

$$\mathbf{w}_u \leftarrow \mathbf{w}_u + \eta(\varepsilon_{ui} \mathbf{r}_{ui} - \lambda \mathbf{w}_u) \tag{13}$$

$$\mathbf{w}_i \leftarrow \mathbf{w}_i + \eta(\varepsilon_{ui} \mathbf{r}_{ui} - \lambda \mathbf{w}_i) \tag{14}$$

Here, $\varepsilon_{ui} = v_{ui} - (\mathbf{w} + \mathbf{w}_u + \mathbf{w}_i)^T \mathbf{r}_{ui}$, η is learning rate. After getting \mathbf{w} , \mathbf{w}_u , and \mathbf{w}_i , given reviews in test datasets, we can use $\hat{v}_{ui} = (\mathbf{w} + \mathbf{w}_u + \mathbf{w}_i)^T \mathbf{r}_{ui}$ to predict the review rating.

4. Experiments and Evaluations

We have performed extensive experiments on a variety of datasets to demonstrate the effectiveness of our methods when compared to other state-of-the-art RRP method. We first introduce the datasets and the evaluation metric used in our experiments in Section 4.1. Experimental settings and research problem are given in Section 4.2. Performance evaluation and some analysis of the model are discussed in Sections 4.3–4.5, respectively.

4.1. Datasets and Evaluation Metric

In order to evaluate the performance of our methods, we respectively implement our experiments on English data sets and Chinese data sets. We have selected two public English data sets Yelp 2013 and Yelp 2014, which is a large-scale dataset consisting of restaurant reviews (<https://www.yelp.com/dataset/challenge>). On the Chinese dataset, we constructed two Douban movie review datasets because there is no suitable public dataset in Chinese.

Douban (<http://www.douban.com/>) is a popular Chinese review site where users can post comments on movies, books and music and give a 1 to 5 star rating. Through the API interface that is provided by Douban, we download the Douban movie user information and then sort the Douban movie users according to the number of published reviews. We selected film users who have published more than 50 movie reviews as seeds. Focusing on the movie users, through the Douban API interface, grab the movie reviews published by Douban movie users. Based on the captured Douban data, two movie review data sets were constructed. The statistics of the four datasets are shown in Table 1.

Table 1. Statistical Information of Yelp 2014, Yelp 2013, and Two Douban Movie Review Datasets.

Datasets	#Users	#Reviews	#Items	#Reviews/User	#Reviews/Product
Douban 1	1476	22,593	3041	15.31	7.43
Douban 2	1079	13,858	2087	12.84	6.64
Yelp 2014	4818	231,163	4194	47.97	55.12
Yelp 2013	1631	78,966	1633	48.42	48.36

We use mean absolute error (MAE) and root mean square error (RMSE) as evaluation indicators to evaluate the performance of different RRP methods. MAE and RMSE are defined, as follows:

$$\text{MAE} = \frac{\sum_{\text{testsets}} |\hat{v}_{ui} - v_{ui}|}{N_{\text{total}}} \quad (15)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{\text{testsets}} (\hat{v}_{ui} - v_{ui})^2}{N_{\text{total}}}} \quad (16)$$

Here, \hat{v}_{ui} is the review rating predicted by various methods, v_{ui} is the true rating of the review in the test set, and N_{total} is the number of reviews in the test set.

4.2. Experimental Settings and Research Questions

We divided each dataset shown in Table 1 into two parts: training set and test set. We use 80% of each dataset as the training set and the rest is used as the test set. All of the hyper-parameters of our methods are selected based on the performance on the training set. For comparison, we summarize our proposed models and baseline methods as follows.

- LR + global: RRP method based on linear regression trained and tested on all users and products.
- LR + global + UPF: We extract user features [36] and corresponding product features (denoted as UPF) from training data, and concatenate them with the features in baseline LR + global.

- LR + individual + product: RRP method based on linear regression trained and tested on individual product.
- LR + individual + user: RRP method based on linear regression trained and tested on individual user.
- KNN: RRP method based on k-nearest neighbor.
- MF: RRP method based on matrix factorization.
- RRP + UC: RRP method based on review content and user context.
- RRP + PC: RRP method based on review content and product context.
- RRP + UC + PC: RRP method based on user context and product context.

By combining the review content, user context and product context information, we propose a RRP method based on user context and product context. To analyze the performance of our method and the factors that affect the performance of our method, on the four datasets, we did three experiments to answer the following three questions:

- (1) Whether our method performance is better than the benchmark method.
- (2) The impact of the number of reviews per user and reviews for each product on our methods
- (3) The impact of user context and product context on our methods.

4.3. Performance Comparison of Different Methods

To verify the performance of our method, we compared our method and six benchmark methods on four different datasets. The RRP results of nine different methods are shown in Table 2.

Table 2. Mean absolute error (MAE) and root mean square error (RMSE) of Six Different Methods in Four Datasets.

Datasets	Douban1	Douban1	Douban2	Douban2	Yelp2014	Yelp2014	Yelp2013	Yelp2013
Metric	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
KNN	1.0659	1.4547	1.0626	1.4271	0.7112	0.9993	0.6987	0.9856
MF	0.8341	1.0653	0.8056	1.0387	0.5132	0.8146	0.4871	0.8042
LR + global	0.8477	1.1008	0.8277	1.0741	0.5686	0.8985	0.5623	0.8931
LR + global + UPF	0.8365	1.0681	0.8154	1.0684	0.5305	0.8606	0.5225	0.8334
LR + individual+product	0.8431	1.0909	0.8231	1.0702	0.5589	0.8884	0.5526	0.8736
LR + individual + use	0.8370	1.0703	0.8124	1.0645	0.5383	0.8682	0.5324	0.8533
RRP + PC	0.8325	1.0542	0.8170	1.0513	0.5252	0.8523	0.5062	0.8114
RRP + UC	0.8216	1.0391	0.8081	1.0282	0.5158	0.8326	0.4961	0.8024
RRP + RC + PC	0.8111	0.9899	0.7803	0.9794	0.4841	0.7946	0.4652	0.7441

On four different datasets, our approach reduced the MAE and RMSE of the RRP and achieved better performance than the six benchmark methods. Both the RMSE of global methods, such as LR + global and LR + global + UPF, and those of individual methods, such as LR + individual + user and LR + individual + product, are higher than our method. This is because the global methods fail to capture the individuality of each user and product, and the individual methods suffer from data sparseness. Our approach can outperform both the global and individual methods because our method can capture individuality of user and product, and at the same time, exploit the common knowledge shared by different users and products to handle data sparseness problem.

In addition, our approach performs better than the KNN and MF method. It indicates that our method is a more appropriate way to RRP than KNN and MF. For example, the RMSE of RRP + UC + PC is reduced by 5.71% $((1.0387 - 0.9794)/1.0387)$ when compared to MF, which is the best method in baselines in Douban2. The RMSE of RRP + UC + PC is reduced by 7.47% $((0.8042 - 0.7441)/0.8042)$ compared to MF, which is the best method in baselines in Yelp 2013.

The experimental results show that RRP + UC + PC can improve the performance of the RRP. This is because RRP is not only related to the review text content, but also related to user context and

product context. By combining the user context and product context information, the performance of the RRP is improved to some extent.

From Table 2, we can find a very interesting result. In douban1 and douban2 datasets, the MAE of RRP + UC + PC is only reduced by 4.32% $((0.8477 - 0.8111)/0.8477)$ and 5.73% $((0.8277 - 0.7803)/0.8277)$ as compared to LR. However, in yelp2014 and yelp2013 datasets, the MAE of RRP + UC + PC is reduced by 14.86% $((0.5686 - 0.4841)/0.5686)$ and 17.27% $((0.5623 - 0.4652)/0.5623)$ compared to LR. This is because the number of reviews per user and reviews for each product in yelp datasets are much larger than the number of reviews per user and reviews for each product in the douban datasets.

4.4. The Impact of the Number of Reviews Per User and Reviews for Each Product on Our Methods

In order to further research the impact of the number of reviews per user and reviews for each product on our methods. We hope to answer two questions through experiments. First, increasing the number of reviews per user can benefit each user's personalized RRP. Second, increasing the number of reviews per product can benefit the personalized RRP for each product.

To study the impact of number of reviews per user on our proposed approach, in the experiment, we selected P% reviews for each user to train RRP + UC. The value of P varies from 10 to 100 at intervals of 10. The experimental results on the yelp2013 datasets are shown in Figure 1. The results of the RRP + LR + individual + user and RRP + LR + global methods are also shown in Figure 1 as a baseline method for comparison.

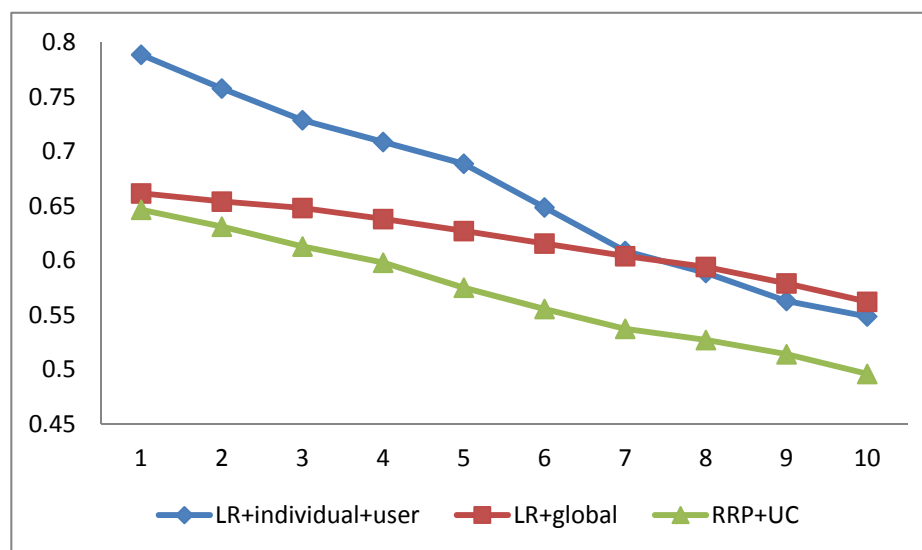


Figure 1. Impacts of Number of Reviews Per User on MAE of Our Methods in Yelp2013 Datasets.

From Figure 1, we can see that, as the number of reviews per user increases, the MAE of our proposed method and baseline methods are decreasing. This result is because as the number of reviews per user increases, more information is used in the training process of RRP, thus helping to learn a more accurate RRP method. In addition, the effects of our method proposed in this paper can be better than the LR + individual + user and LR + global methods in different training text sizes. This is because the LR + global approach does not capture the personalized sentiment expression of each user, while the LR + individual + user approach faces the challenge of data sparsity. In addition, the performance of our proposed method becomes more apparent than the LR + individual method when the number of training texts per user is reduced.

However, compared with the LR + global method, our method fluctuates more with the change of the number of reviews per user. Our method is more sensitive to the training data than, and not as stable as, the LR + global method.

To study the impact of reviews for each product on our proposed approach, in the experiment, we selected P% reviews for each product to train RRP + PC. The value of P varies from 10 to 100 at intervals of 10. The experimental results on the yelp2013 datasets are shown in Figure 2. The results of the LR + individual + product and LR + global methods are also shown in Figure 2 as a baseline method for comparison. From Figure 2, we can get a conclusion similar to Figure 1.

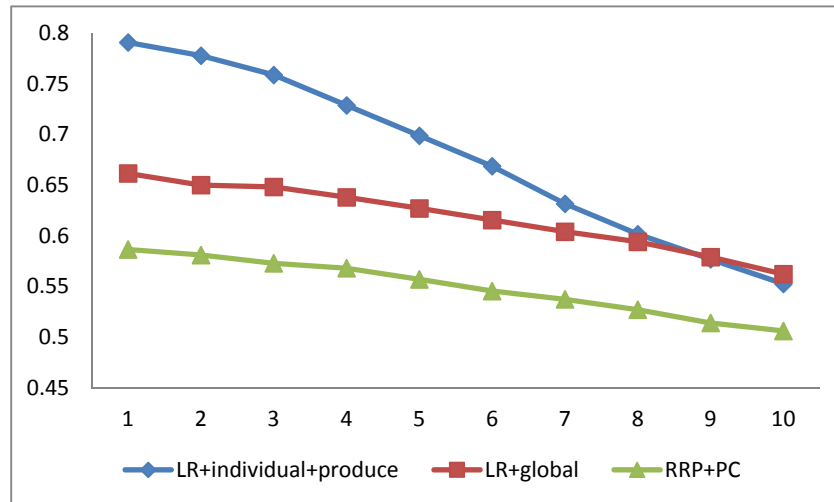


Figure 2. Impacts of Number of Reviews Per User on MAE of Our Methods in Yelp2013 Datasets.

4.5. The Impact of User Context and Product Context on Our Methods

To evaluate the impact of user context and product context on RRP, we separately remove user context and product context from RRP + UC + PC and conducted comparative experiments on four datasets. The experimental results of the four datasets are shown in Figures 3 and 4, respectively.

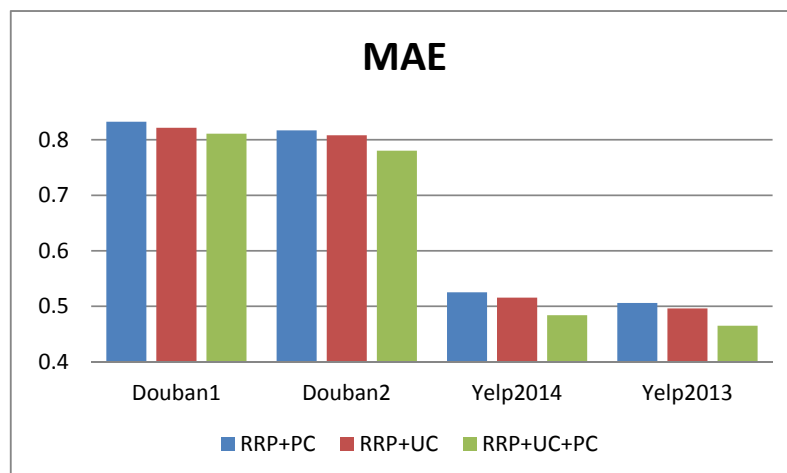


Figure 3. Impacts of User Context and Product Context on MAE of Our Methods in the Four Datasets.

From Figure 3, we can find the MAE of RRP + UC + PC is greatly improved when the user context information is removed compared to the removal of the product context. For example, when the user context information is removed, the MAE of RRP + UC + PC is increased by 4.49% $((0.8170 - 0.7803)/0.8170)$ in the douban2 datasets. However, when the product context information is removed, the MAE of RRP + UC + PC is only increased by 3.44% $((0.8081 - 0.7803)/0.8081)$ in the douban2 datasets.

From Figure 4, we can get the same result. We find the RMSE of RRP + UC + PC is greatly improved when the user context information is removed as compared to the removal of the product

context. For example, when the user context information is removed, the MAE of RRP + UC + PC is increased by 8.29% $((0.8114 - 0.7441)/0.8114)$ in the yelp 2013 datasets. However, when the product context information is removed, the MAE of RRP + UC + PC is only increased by 7.27% $((0.8024 - 0.7441)/0.8024)$ in yelp2013 datasets.

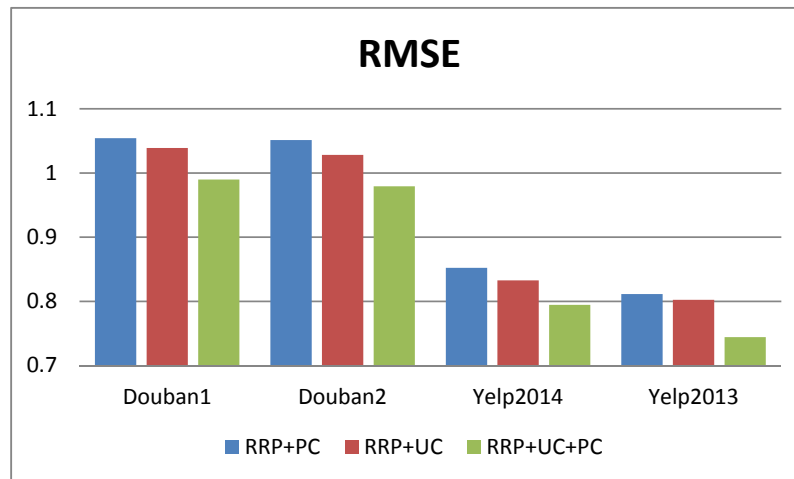


Figure 4. Impacts of User Context and Product Context on RMSE of Our Methods in the Four Datasets.

The experimental results on four different datasets show that user contexts are more effective than product contexts in RRP. This is because the user context information contains not only user-special words, but also user's personal sentiment bias.

5. Conclusions

In this paper, we present a new RRP method that is based on the user context and product context. More specifically, in order to solve problem of existing RRP methods based on review content, considering the user personalized information and product context information, which are useful to predict rating of review, we propose a RRP method based user context and product context. Experimental results on four datasets show that our proposed methods have more performance than the state-of-the-art baselines in RRP.

Existing review rating prediction methods only use a single model to capture the sentiment of review texts. Our method respectively considers affection of user and product on reviews content in capturing the sentiment of review texts. However, our method does not regarding common affection of user and product on review texts in capturing the sentiment of review texts. Regardless of existing methods or our method, only the review content information is modelled in RRP. In the future, in order to address these issues, we will firstly model together affection of user context and product context on reviews texts, incorporating the user-product-specific model into our method to further improve the performance of RRP. At the same time, we will use user and product context information as the attention mechanism of the neural network model to adjust the weighting coefficient of the global model, user-specific model, product-specific model, and user-product-specific model in our proposed future method.

Author Contributions: Conceptualization, B.W.; Data curation, Y.H.; Formal analysis, B.W.; Investigation, B.W.; Methodology, B.W.; Project administration, Y.H.; Supervision, X.L.; Writing—original draft, B.W.; Writing—review & editing, X.L.

Funding: This research was funded by Foundation of He'nan Educational Committee (19A520032) and Ph.D. Start-up Foundation of Pingdingshan University (PXY-BSQD-2018007) and National Natural Science Foundation of China (U1536201, U1405254, and 61271392).

Acknowledgments: Thanks to Zhigang Yuan for valuable discussions on the research and the manuscript.

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

References

1. Pang, B.; Lee, L. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.* **2008**, *2*, 1–135. [[CrossRef](#)]
2. Liu, B. Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* **2012**, *5*, 1–167. [[CrossRef](#)]
3. Khan, F.H.; Qamar, U.; Bashir, S. A semi-supervised approach to sentiment analysis using revised sentiment strength based on Senti Word Net. *Knowl. Inf. Syst.* **2017**, *51*, 851–872. [[CrossRef](#)]
4. Khan, F.H.; Qamar, U.; Bashir, S. e SAP: A decision support framework for enhanced sentiment analysis and polarity classification. *Inf. Sci.* **2016**, *367*, 862–873. [[CrossRef](#)]
5. Khan, F.H.; Qamar, U.; Bashir, S. Multi-objective model selection (MOMS)-based semi-supervised framework for sentiment analysis. *Cogn. Comput.* **2016**, *8*, 614–628. [[CrossRef](#)]
6. Khan, F.H.; Qamar, U.; Bashir, S. SWIMS: Semi-supervised subjective feature weighting and intelligent model selection for sentiment analysis. *Knowl. Based Syst.* **2016**, *100*, 97–111. [[CrossRef](#)]
7. Mishra, S.; Diesner, J.; Byrne, J.; Surbeck, E. Sentiment analysis with incremental human-in-the-loop learning and lexical resource customization. In Proceedings of the 26th ACM Conference on Hypertext & Social Media, Guzelyurt, Cyprus, 1–4 September 2015; pp. 323–325.
8. Horrigan, J. *Online Shopping*; Pew Internet and American Life Project Report; Pew Research Center: Washington, DC, USA, 2008.
9. Wu, Y.; Ester, M. FLAME: A probabilistic model combining aspect based opinion mining and collaborative filtering. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai, China, 2–6 February 2015; pp. 199–208.
10. Qu, L.; Ifrim, G.; Weikum, G. The bag-of-opinions method for review rating prediction from sparse text patterns. In Proceedings of the 23rd International Conference on Computational Linguistics, Beijing, China, 23–27 August 2010; pp. 913–921.
11. Zheng, L.; Zhu, F.; Mohammed, A. Attribute and Global Boosting: A Rating Prediction Method in Context-Aware Recommendation. *Comput. J.* **2018**, *60*, 957–968. [[CrossRef](#)]
12. Ganu, G.; Elhadad, N.; Marian, A. Beyond the Stars: Improving Rating Predictions using Review Text Content. In Proceedings of the International Workshop on the Web and Databases, WEBDB 2009, Providence, RI, USA, 28 June 2009.
13. Zheng, L.; Noroozi, V.; Yu, P.S. Joint Deep Modeling of Users and Items Using Reviews for Recommendation. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, Cambridge, UK, 6–10 February 2017; pp. 425–434.
14. Wang, H.; Lu, Y.; Zhai, C. Latent aspect rating analysis on review text data: A rating regression approach. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 25–28 July 2010; pp. 783–792.
15. Tang, D.; Qin, B.; Liu, T. Learning Semantic Representations of Users and Products for Document Level Sentiment Classification. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 26–31 July 2015; pp. 1014–1023.
16. Wu, F.; Huang, Y. Personalized Microblog Sentiment Classification via Multi-Task Learning. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16), Phoenix, AZ, USA, 12–17 February 2016; pp. 3059–3065.
17. Tang, D.; Qin, B.; Yang, Y.; Liu, T. User Modeling with Neural Network for Review Rating Prediction. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015), Buenos Aires, Argentina, 25–31 July 2015; pp. 1340–1346.
18. Pang, B.; Lee, L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, Ann Arbor, MI, USA, 25–30 June 2005; pp. 115–124.
19. Li, F.; Liu, N.; Jin, H.; Zhao, K.; Yang, Q.; Zhu, X. Incorporating reviewer and product information for review rating prediction. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011; pp. 1820–1825.
20. Liu, J.; Seneff, S. Review sentiment scoring via a parse-and-paraphrase paradigm. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1, Singapore, 6–7 August 2009; pp. 161–169.

21. Lee, H.C.; Lee, S.J.; Chung, Y.J. A study on the improved collaborative filtering algorithm for recommender system. In Proceedings of the 5th ACIS International Conference on Software Engineering Research, Management & Applications (SERA 2007), Busan, Korea, 20–22 August 2007; pp. 297–304.
22. Jeong, B.; Lee, J.; Cho, H. Improving memory-based collaborative filtering via similarity updating and prediction modulation. *Inf. Sci.* **2010**, *180*, 602–612. [[CrossRef](#)]
23. Shi, Y.; Larson, M.; Hanjalic, A. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Comput. Surv.* **2014**, *47*. [[CrossRef](#)]
24. Li, P.; Wang, Z.; Ren, Z.; Bing, L.; Lam, W. Neural rating regression with abstractive tips generation for recommendation. In Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, 7–11 August 2017; pp. 345–354.
25. Catherine, R.; Cohen, W. TransNets: Learning to Transform for Recommendation. In Proceedings of the Eleventh ACM Conference on Recommender Systems, Como, Italy, 27–31 August 2017; pp. 288–296.
26. Kim, D.; Park, C.; Oh, J.; Lee, S.; Yu, H. Convolutional Matrix Factorization for Document Context-Aware Recommendation. In Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, 15–19 September 2016; pp. 233–240.
27. Seo, S.; Huang, J.; Yang, H.; Liu, Y. Interpretable Convolutional Neural Networks with Dual Local and Global Attention for Review Rating Prediction. In Proceedings of the Eleventh ACM Conference on Recommender Systems, Como, Italy, 27–31 August 2017; pp. 297–305.
28. He, X.; Chen, T.; Kan, M.; Chen, X. TriRank: Review-aware explainable recommendation by modeling aspects. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 18–23 October 2015; pp. 1661–1670.
29. Ling, G.; Lyu, M.R.; King, I. Ratings meet reviews, a combined approach to recommend. In Proceedings of the 8th ACM Conference on Recommender Systems, Foster City, CA, USA, 6–10 October 2014; pp. 105–112.
30. McAuley, J.; Leskovec, J. Hidden factors and hidden topics: Understanding rating dimensions with review text. In Proceedings of the 7th ACM conference on Recommender Systems, Hong Kong, China, 12–16 October 2013; pp. 165–172.
31. Ren, Z.; Liang, S.; Li, P.; Wang, S.; de Rijke, M. Social Collaborative Viewpoint Regression with Explainable Recommendations. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, Cambridge, UK, 6–10 February 2017; pp. 485–494.
32. Bao, Y.; Fang, H.; Zhang, J. Topicmf: Simultaneously exploiting ratings and reviews for recommendation. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, QC, Canada, 27–31 July 2014; pp. 2–8.
33. Diao, Q.; Qiu, M.; Wu, C.; Smola, A.J.; Jiang, J.; Wang, C. Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 193–202.
34. Jakob, N.; Weber, S.H.; Müller, M.C.; Gurevych, I. Beyond the stars: Exploiting free-text user reviews to improve the accuracy of movie recommendations. In Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion, Hong Kong, China, 2–6 November 2009; pp. 57–64.
35. Gong, L.; Al Boni, M.; Wang, H. Modeling social norms evolution for personalized sentiment classification. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 7–12 August 2016; pp. 855–865.
36. Gao, W.; Yoshinaga, N.; Kaji, N.; Kitsuregawa, M. Modeling user leniency and product popularity for sentiment classification. In Proceedings of the Sixth International Joint Conference on Natural Language Processing, Nagoya, Japan, 14–18 October 2013; pp. 1107–1111.

