

# Scoring Prediction Model Based on Fusion of Text and Temporal Features

Haoqian Li<sup>1</sup>, Xuesong Su<sup>2</sup>, Wenguang Zheng<sup>1\*</sup>, Jiayi Song<sup>1</sup> and Yingyuan Xiao<sup>1</sup>

<sup>1</sup>*School of Computer Science and Engineering, Tianjin University of Technology, Tianjin, China*

<sup>2</sup>*Technology Inspection Center of Shengli Oilfield SINOPEC Dongying, Shandong, China*

## Abstract

Nowadays, the Internet has penetrated into every household and is inseparable from our lives. The Internet provides users with a wide range of online products and services, the utilization of online platforms is much higher than before, and users gradually become dependent on online consumption. To promote the development of e-commerce ecosystem, the main task is to increase the purchase rate of goods. In order to reduce information overload and meet the diverse needs of users, a personalized recommendation system that implements recommendation tasks using review text is an effective solution to increase the purchase rate of users, and it also solves the problems of sparse data and cold starts. Currently, many deep recommendation models based on review text have emerged in this field, but these models have some other problems, such as lack of fusion of multiple features and lack of judgment of feature importance levels.

## Keywords:

review text; recommendation system; feature fusion; rating prediction.

## 1. Introduction

With the development of the Internet, people are now entering the era of big data, and the intelligent society makes our information exist in every registered end of the application; when we log in our account to browse the website, this website also records our browsing information, when we shop online, the information of the purchased products and when we give feedback, our comments on the products are also recorded one by one. In general, a growing number of studies argue that since reviews explain users' opinions, they should help to infer potential dimensions for predicting ratings or purchases. The schemes incorporating reviews have since evolved from simple regularization methods to achieve rating prediction to neural network methods to achieve rating prediction.

One of the main directions of research in recommender systems is to improve prediction by using latent information features, especially in cold-start environments where interaction data may be sparse or noisy. It has been shown in current rating prediction algorithms that introducing review text data into the rating prediction task can greatly improve the accuracy of recommendations. To enhance the fusion of features and the importance of focusing on interaction features to make improvements in the accuracy of user-item rating prediction, this paper proposes a multi-feature fusion-based rating prediction model (2TFRS). Our proposed algorithm models the rating prediction problem as a text matching problem using user-item review text fused with temporal features. The interaction between features is enhanced by learning methods that capture important matches between user text and item text; secondly, the model still combines review timing information, user and item rating information. Finally, the prediction scores are derived through a regression layer.

---

ICBASE2022@3rd International Conference on Big Data & Artificial Intelligence & Software Engineering, October 21-23, 2022, Guangzhou, China

EMAIL: 1369227010@qq.com (Haoqian Li); wenguangz@tjut.edu.cn (Wenguang Zheng)

ORCID: 0000-0002-7604-9893 (Haoqian Li)



© 2022 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

## 2. Related works

In this section, we introduce the classical recommendation models incorporating deep learning techniques in recent years and the innovative points and construction ideas of our proposed model.

In 2016, the word2vec<sup>[1]</sup> and Item2vec algorithms, borrowed from the field of natural language processing, learned the user's behavior sequences as sentences for representation. In the same year, Google released YoutubeDNN<sup>[2]</sup> which introduced the classical recommendation system architecture, divided into two phases of recall and ranking, providing ideas for subsequent industrial-grade recommendation systems. Many other classical models were proposed in the same period, such as ConvMF<sup>[3]</sup>, PNN<sup>[4]</sup>, DNN<sup>[5]</sup>, DeepCrossing<sup>[6]</sup> etc. Many scholars have found that the accuracy of the prediction results is greatly improved by taking the attention mechanism into account when modeling the scoring prediction problem. Seo et al. proposed an attention-based convolutional neural network model (D-Attn model<sup>[7]</sup>) to extract potential representations of users and items. In 2017, L Zheng et al. proposed the DeepCoNN<sup>[8]</sup> model to mine the nature of user-behavioral goods from user-item review texts; In 2018, C. Chen et al. proposed the NARRE<sup>[9]</sup> model, which adds the attention mechanism to assign weight to each comment based on the DeepCoNN model; in 2020, Parisa Abolfath and Saeedeh Momtazi proposed the MPRS<sup>[10]</sup> model, which aims to focus on user-item interaction features; In 2022, Peilin Yang et al. proposed a deep learning-based main auxiliary network—MAN<sup>[11]</sup>, which helps the main network to generate rating prediction values by focusing on the deep meaning at the word vector level with the help of the auxiliary network; this paper is based on this model and presents new ideas. Considering the influence of time decay factor on recommendation results, combining the temporal information of reviews as the weights of embedded vectors and then doing local interactions between vector pairs, performing convolutional operations and then inputting into the regression layer to obtain predicted scores.

## 3. Our proposed model

First, we consider each data record as a tuple  $(u, v, r_{uv}, re_{uv}, t_{uv})$  to represent the review records written by user  $u$  for item  $v$ , including the rating  $r_{uv}$ , the review  $re_{uv}$  and comment time  $t_{uv}$ . To construct the user-item matching matrix, a review document-based approach is used. Firstly, all the reviews written by user  $u$  are concatenated into a single document, denoted as  $d_{1:n}^u$ , consisting of  $n$  words. Similarly, integrate all the comment texts of item  $v$  into one document as  $d_{1:m}^v$  where  $m$  is the length of the document.

Next, we use the word vector matching matrix  $M_{uv}$  for each pair of user  $u$  and item  $v$  as input to the CNN architecture, and each element in the matrix  $m_{p,q}^{u,v}$  represents the similarity of the  $p$ th word  $j_p$  in the user text document  $d_{1:n}^u$  to the  $q$ th word  $k_q$  in the item text document  $d_{1:m}^v$ . The first layer of our proposed model is the embedding layer, which serves to map each word in the review text to a  $d$ -dimensional vector. The words of each user review text and the words of each item review text are trained by a word embedding function that can be used with the pre-trained embeddings from GloVe<sup>[12]</sup> used in Wikipedia. The  $\alpha_p^u$  and  $\beta_q^v$  in Equation 1 represent the word embedding vectors of word  $j_p$  and word  $k_q$  respectively, with the user embedding vector denoted as  $\alpha$  and the item embedding vector as  $\beta$ . In this model, we consider that the user's preferences may change with time, we propose a weighting function that takes into account the time decay, the user's recent review behavior represents the user's recent preferences, and the item's recent reviews often represent the recent quality of this item, which can accurately improve the recommendation performance.

$$f(\alpha) = \alpha \cdot \exp\left(\frac{-|\Delta t|}{\theta}\right) \quad (1)$$

Where  $\Delta t$  is the time difference between the time of user  $u$  commenting on item  $v$  and the current time. As  $\Delta t$  increases, the weight value of the embedding vector becomes smaller, which means that the

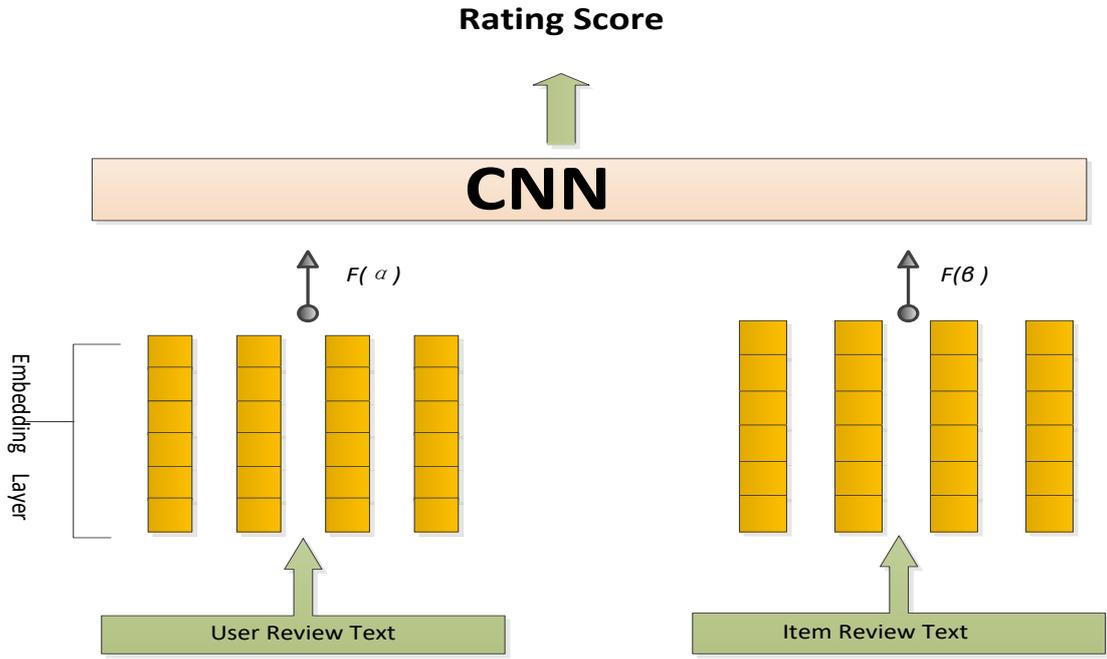
user's commenting behavior long ago will have less influence on the prediction score. Then, the user embedding vector with weights added is denoted as  $\varepsilon$  and the user embedding vector with weights added is denoted as  $\eta$ .

To better capture the word meanings between each word for matching, we calculate the similarity between two words by computing the cosine similarity of their word embedding vectors to construct a user-item matching matrix to achieve a joint representation of user-item pairs, capture the joint semantic information .

Its similarity is calculated as shown in Equation 2:

$$m_{p,q}^{u,v} = \frac{\varepsilon_p^u \cdot \eta_q^v}{\|\varepsilon_p^u\| \|\eta_q^v\|} \quad (2)$$

The above is the first layer of the architecture of our proposed model. Next, we input the user-item pair matching matrix constructed above into the convolutional neural (CNN) architecture.



**Figure 1** Scoring prediction model based on text and temporal feature fusion

The CNN framework consists of a convolutional layer, a pooling layer, and a fully connected layer, each convolutional layer uses a filter  $K_i \in \mathbb{R}^{c \times t}$  to generate a feature map  $z_i$ , defined as Equation 3:

$$z_i = ReLU(M_{uv} * K_i + b_i) \quad (3)$$

$z_i \in \mathbb{R}^{(n-c+1)(m-t+1)}$ , The symbol  $*$  is the convolution operator,  $b_i$  is the bias term.

We apply a maximum pooling operation to each feature mapping obtained from the convolutional layer and use the maximum value in each pooling window as one of the features of the corresponding kernel. By repeating the convolutional and maximum pooling layers, we allow to filter the interference and extract each feature simultaneously in a certain proportion. The result of the final maximum pooling layer is passed to a fully connected layer and then to a single neuron layer with a linear activation function - the regression layer. Concomitantly, the User-Item prediction score  $\hat{r}_{uv}$  is output, and the calculation formula 4 is as follows:

$$\hat{r}_{uv} = W \times O + g \quad (4)$$

Where  $O$  is the result of the fully connected layer,  $W$  contains the weights of the regression layer, and  $g$  is the bias term.

The method proposed in this paper is a rating prediction model based on the fusion of textual and temporal features to achieve rating prediction, and this model is suitable for regression prediction based on user-item similarity.

## 4. Experiments

We used the Amazon Review Dataset-5core public dataset to make an evaluation for each experiment. Each dataset was randomly divided into training set, validation set and test set in the ratio of 80:10:10. Classify the comment data set by category, as shown in Table 1:

All comment texts were first processed by the Stanford Core NLP word generator to obtain the sub words, then their stop words and punctuation marks are removed, and finally the word embedding vectors were trained on the Glove. 6B.50d word list used in Wikipedia. In CNN, we set the number of convolutional layers to 7, and the number of convolution kernels for each layer is 64; Set the hyperparameter batch size to 128 and the dropout rate to 0.5. The root mean square error (RMSE) is used to evaluate the performance of our proposed algorithm. Let  $N$  be the total number of data points to be tested. Then, define the RMSE (5) as

$$RMSE = \sqrt{\frac{1}{N} \sum (r_{uv} - \hat{r}_{uv})^2} \quad (5)$$

In order to better test the performance of our proposed models. For this purpose, we tested the DeepCoNN model, the NARRE model, the MPRS model, and the MAN model on a subset of Amazon on the latest version. The mean squared error values (RMSE) obtained for our models as well as for the models of the comparison experiments on five different classes of the Amazon data subset have been listed Table 1:

**Table 1**  
RMSE results of the compared methods on different datasets

Dataset	Deep-CoNN	MPRS	MAN	2TFRS(relative improvement)
Cell Phones and Accessories	1,4129	1.0129	1.1022	1.0743(2.79%)
Toy and Games	0,9024	0.8959	0.8384	0.8271(1.13%)
Office Products	0.8692	0.7887	0.7727	0.7536(1.91%)
Amazon instant video	1.1132	1.1022	0.9786	0.9471(3.15%)
Digital Music	0.9006	0.8760	0.8701	0.8502(1.99%)

As shown in the table, the performance of our proposed model is superior to the baseline experiments after testing on five data subsets. It can be seen that our model obtains the largest relative improvement (3.15%) on the Amazon dataset AZ-IV (Amazon instant video) category.

## 5. Conclusion

After drawing on a large number of excellent papers on recommendation systems, we propose a new model based on feature fusion to achieve rating prediction. The model is modeled separately by using user reviews and item reviews, embedding the word vectors, processing the temporal information as the weights of the embedded vectors, and then calculating the similarity of the user-item vector pairs with weights to form a matching matrix to The CNN architecture and regression network are used to predict the user's rating value of items.

## 6. References

- [1] Mikolov, Tomas, et al. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781(2013)
- [2] Covington, Paul, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. Proceeding of the 10th ACM conference on recommender systems. 2016
- [3] D. Kim, C. Park, J. Oh, S. Lee, H. Yu, Convolutional matrix factorization for document context-aware recommendation, in: Proceedings of the 10th ACM Conference on Recommender Systems – RecSys’6, 2016
- [4] Qu, Yuanru, et al. Product-based neural networks for user response prediction. 2016 IEEE 16th International Conference on Data Mining (ICDM)IEEE,2016
- [5] Zhang, Weinan, Tianming Du, and Jun Wang. Deep learning over multi-field categorical data. European conference on information retrieval. Springer, Cham,2017
- [6] Shan, Ying, et al. Deep crossing: Web-scale modeling without manually crafted combinatorial features. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.ACM.2016
- [7] S. Seo, J. Huang, H. Yang, Y. Liu, Representation learning of users and items for review rating prediction using attention-based convolutional neural network, in: Proceedings of the 3rd International Workshop on Machine Learning Methods for Recommender Systems – MLRec, 2017
- [8] L. Zheng, V. Noroozi, P.S. Yu, Joint deep modeling of users and items using reviews for recommendation, in: Proceedings of the 10th ACM International Conference on Web Search and Data Mining - WSDM '17,2017, pp. 425–434
- [9] C. Chen, M. Zhang, Y. Liu, S. Ma, Neural attentional rating regression with review-level explanations, in: Proceedings of the 2018 World Wide Web Conference, 2018, pp. 1583–1592
- [10] Parisa Abolfath, Beygi Dezfouli, Saeedeh Momtazi, Mehdi Dehghan. Deep neural review text interaction for recommendation systems, Applied Soft Computing Journal 100 (2021) 106985
- [11] Yang, Peilin, et al, MAN: Main-auxiliary network with attentive interactions for review-based recommendation in: Applied Intelligence-2022pp:1—16
- [12] J. Pennington, R. Socher, C. D. Manning, GloVe: Global Vectors for Word Representation, EMNLP (2014)