Artificial Life and Robotics (2024) 29:107–113
https://doi.org/10.1007/s10015-023-00924-5

ORIGINAL ARTICLE

Double-ConvMF: probabilistic matrix factorization with user and item characteristic text

Takuya Tamada[1] · **Ryosuke Saga**[1]

Received: 30 March 2023 / Accepted: 23 October 2023 / Published online: 23 January 2024

Abstract
In today's information-rich society, the importance of recommender systems for matching items and customers is increasing day by day. The development of e-commerce sites and review sites has made it possible to access a large amount of product descriptions and user reviews, and it is believed that more advanced recommendation models can be proposed by efficiently utilizing this text information. ConvMF is the first model that integrates text and probabilistic matrix factorization(PMF) which is one of the matrix factorization methods. In this method, features are extracted from item text such as item descriptions using CNN architecture and integrated into PMF. However they focus only on the item text and not on the user factor. As a result, this method can not reflect user characteristics. Therefore, this paper proposes a new recommender system to extract both item and user features from item and user text using CNN and integrate them into matrix factorization.

Keywords Recommender system · Machine learning · Matrix factorization · Text mining · e-commerce

1 Introduction

Recommender systems are useful tools for presenting appropriate information from a large amount of information available to users in the age of information explosion. In particular, recommender systems are being used increasingly in web services such as e-commerce, news sites, and music streaming platforms to increase the revenue of these sites [1, 2].

One of the most successful methods of recommender system is matrix factorization [3]. Matrix factorization is a technique for decomposing a user's evaluation information into a user latent matrix and an item latent matrix with an implicit factor. Probabilistic matrix factorization (PMF) [3], which employs a probabilistic algorithm that works well on sparse datasets, was invented. To solve the cold-start problem, a model that adds various auxiliary information such as tags [4], text [5, 6], images [1, 7], and social relations [8, 9] to the matrix decomposition has been proposed.

With the advent of news sites and web media, and the development of review functions in services, such as Twitter and Amazon, people have access to a large amount of text data. However, as the amount of data increases, users have difficulty accessing the information appropriately. Therefore, among all the supplementary information available, the value of effectively using text data to recommend suitable items to users is increasing.

ConvMF [5] is a typical example of research that uses text information in matrix factorization to improve recommendation accuracy. In this method, features are extracted from the text representing the features of the item using CNN and incorporated into PMF. Here, the text that represents the features of an item is the item description written by each store and reviews posted for the item. On the other side, user reviews truly reflect user characteristics. For example, when a user posts "This shirt is comfortable to wear", we can understand that this user is looking for clothes that are comfortable to the touch. However, ConvMF cannot capture user characteristics. Therefore, we propose Double-ConvMF, which extracts features not only

This work was presented in part at the joint symposium of the 28th International Symposium on Artificial Life and Robotics, the 8th International Symposium on BioComplexity, and the 6th International Symposium on Swarm Behavior and Bio-Inspired Robotics (Beppu, Oita and Online, January 25–27, 2023).

✉ Ryosuke Saga
r.saga@omu.ac.jp

Takuya Tamada
se22364w@st.omu.ac.jp

[1] Graduate School of Informatics, Osaka Metropolitan University, Sakai, Osaka, Japan

from item description text but also from user description text and integrates them into PMF. In this paper we use the text that represents the characteristics of both users and items, and incorporate it into matrix factorization using CNN. The contributions of this paper are as follows.

- Double-ConvMF, in which user reviews are considered as text representing user characteristics and integrated into ConvMF, showed high accuracy on three real-world datasets.
- The effectiveness of item and user description texts was examined by comparing the results of them not only the ConvMF, which uses only item description text, but also a new model that uses only user description text.

2 Related work

The most famous text-based recommendation method is the one that extracts keywords from text describing items and user characteristics, and uses them for recommendation. Beel et al. proposed a recommendation system using TF-IDF, which is a measure of the frequency of word occurrence in a document [10]. Musto et al. proposed a recommendation system using word2vec, a word embedding model [11].

In addition, with the recent improvement in computing technology, many methods using deep learning have been introduced. Zhang et al. [12] extracted information from users' purchase histories and reviews. Additionally, they used hierarchical RNNs to learn from each other to improve the accuracy.

Moreover, a new method applies convolutional neural networks (CNN) to text to capture the back-and-forth relationships between sentences and improve the recommendation accuracy. Wu et al. [13] used CNN to capture the context of news article titles to improve the recommendation accuracy. Kim et al. [5] proposed ConvMF, which captures contextual information and incorporates it into PMF. ConvMF achieves improved accuracy by applying CNN to the item feature text and appropriately adjusting the prior distribution of items in PMF. In this paper, we use user description text, which is not considered in ConvMF. By appropriately adjusting the prior distribution of the user in PMF, we can improve the recommendation accuracy and examine the effectiveness of using the text of the item and the user.

A method similar to ours is BiConvMF by Liu et al. [14]. However, their application of the dataset is limited, and furthermore, no experiments were conducted focusing only on user text. In our paper, a method that applies document features only to the user side is also proposed, and comparative experiments are conducted on three real-world datasets.

3 Method

In this section, we explain our proposed method called Double-ConvMF. First, we introduce the probabilistic model of PMF [3]. Then, we describe the stochastic model and optimization method for simultaneously incorporating item and user description text into PMF.

3.1 PMF (probabilistic matrix factorization)

Salakhutdinov et al. [3] proposed a recommendation method called PMF, which is a kind of matrix factorization method. PMF assumes users N, items M, an arbitrary integer D, and a rating matrix $R \in \mathbb{R}^{N \times M}$ obtained from the user's rating information. The PMF is a matrix factorization R into a user matrix $U \in \mathbb{R}^{D \times N} = \{u_1, u_2, ..., u_N\}$ and an item matrix $V \in \mathbb{R}^{D \times M} = \{v_1, v_2, ..., v_M\}$. The measured score r_{ij} is made by user i for item j. In this case, R is expressed by the following equation:

$$p(R|U, V, \sigma^2) = \prod_{i}^{N} \prod_{j}^{M} N(r_{ij}|u_i^T v_j, \sigma^2)^{I_{ij}}, \tag{1}$$

where $N(x|u, \sigma^2)$ is the probability density function of the Gaussian normal distribution with mean u and variance σ^2. σ is Gaussian noise of R. I_{ij} is the indicator function that is equal 1 if user i rated item j and equal to 0 otherwise.

The optimal matrix U, V minimizes the loss function ε, as shown in the following:

$$\begin{aligned} min\ \varepsilon(U, V) = &\sum_{i}^{N} \sum_{j}^{M} \frac{I_{ij}}{2} (r_{ij} - u_i^T v_j)^2 \\ &+ \frac{\lambda_U}{2} \sum_{i}^{N} \|u_i\|^2 + \frac{\lambda_V}{2} \sum_{j}^{M} \|v_j\|^2, \end{aligned} \tag{2}$$

where λ_U and λ_V are the L_2 regularization terms derived from the Gaussian noise of R, U, and V.

3.2 Double-ConvMF

Figure 1 shows an overview of the probabilistic model for Double-ConvMF. $X = \{x_1, x_2, ..., x_M\}$ is the set of description documents of items, and W is the weight vector of the CNN architecture of items. $X^+ = \{x_1^+, x_2^+, ..., x_N^+\}$ is the set of description documents of users, and W^+ is the weight vector of the CNN architecture of users. In PMF, R is generated from U, V, and σ. In ConvMF, V is generated from X,W and σ_V representing the Gaussian noise. In this way, the prior distribution in PMF is adjusted appropriately.

In our proposed Double-ConvMF, we consider the effectiveness of incorporating X^+, W^+, σ_U to represent the

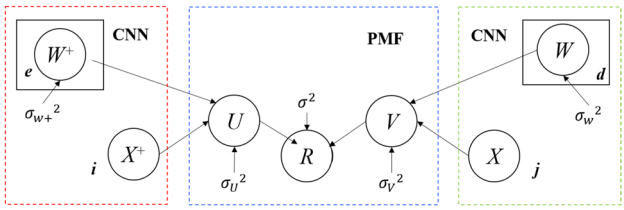

Fig. 1 Graphical model of Double-ConvMF: PMF part in center (blue), ConvMF part in right (green), and Double-ConvMF added to left part (red)

Gaussian noise in matrix factorization and improving the recommendation accuracy by appropriately adjusting the prior distribution by using the user description text for U.

In this paper, we use the CNN architecture proposed in ConvMF, which consists of four layers: embedding, convolution, pooling, and output. The following $cnn(W, x_j)$ is the feature vector of item j obtained by using CNN architecture from the document vector x_j of item j, and $cnn(W^+, x_i^+)$ is the feature vector of user i obtained by using CNN architecture from the document vector x_i^+ of user i.

When $cnn(W, x_j)$ and $cnn(W^+, x_i^+)$ are used, V and U in the PMF probability model can be expressed by the following prior distribution equations:

$$p(V|W, X, \sigma_V^2) = \prod_j^M N\left(v_j|cnn\left(W, x_j\right), \sigma_V^2 I_K\right) \tag{3}$$

$$p(U|W^+, X^+, \sigma_U^2) = \prod_i^N N\left(u_i|cnn\left(W^+, x_i^+\right), \sigma_U^2 I_K\right), \tag{4}$$

where I_K represents the identification matrix. Equations (1),(3) and (4) can be rewritten as follows:

$$\begin{aligned} &p\left(U, V, W, W^+|R, X, X^+, \sigma^2, \sigma_U^2, \sigma_V^2, \sigma_W^2, \sigma_{W^+}^2\right) \\ &\propto p\left(R|U, V, \sigma^2\right) p\left(U|X^+, W^+, \sigma_U^2\right) p\left(V|X, W, \sigma_V^2\right) \\ &\quad p\left(W|\sigma_W^2\right) p\left(W^+|\sigma_{W^+}^2\right). \end{aligned} \tag{5}$$

Now, to optimize (5), we use the following maximum a posteriori estimation:

$$\begin{aligned} &\max_{U,V,W,W^+} p\left(U, V, W, W^+|R, X, X^+, \sigma^2, \sigma_V^2, \sigma_U^2, \sigma_W^2, \sigma_{W^+}^2\right) \\ &= \max_{U,V,W,W^+} [p\left(R|U, V, \sigma^2\right) p\left(U|X^+, W^+, \sigma_U^2\right) p\left(V|X, W, \sigma_V^2\right) \\ &\qquad p\left(W^+|\sigma_{W^+}^2\right) p\left(W|\sigma_W^2\right)]. \end{aligned} \tag{6}$$

By taking the negative logarithm in (6), we can reformulate it as follows:

$$\begin{aligned} min\ \varepsilon\left(U, V, W, W^+\right) = &\sum_i^N \sum_j^M \frac{I_{ij}}{2} \left(r_{ij} - u_i^T v_j\right)^2 \\ &+ \frac{\lambda_U}{2} \sum_i^N \|u_i - cnn\left(W^+, x_i^+\right)\|^2 \\ &+ \frac{\lambda_V}{2} \sum_j^M \|v_j - cnn\left(W, x_j\right)\|^2 \\ &+ \frac{\lambda_{W^+}}{2} \sum_e^{|W_e^+|} \|W_e^+\|^2 + \frac{\lambda_W}{2} \sum_d^{|W_d|} \|W_d\|^2, \end{aligned} \tag{7}$$

where λ_U,λ_V,λ_W, and λ_{W^+} are the regularization terms derived from the Gaussian noise in U,V,W, and W^+, respectively. Partial differentiation of (7) by U and V respectively yields the following equation:

$$u_i = \left(VI_i V^T + \lambda_U I_K\right)^{-1} \left(VR_i + \lambda_U cnn\left(W^+, x_i^+\right)\right) \tag{8}$$

$$v_j = \left(UI_j U^T + \lambda_V I_K\right)^{-1} \left(UR_j + \lambda_V cnn\left(W, x_j\right)\right). \tag{9}$$

In these equation, we treat u_i as a variable and treat others as constants in (8) and we treat v_j as a variable and treat others as constants in (9) where I_i is a diagonal matrix whose diagonal components are the indicator vector $\{I_{i1}, I_{i2}, ..., I_{iM}\}$ that indicate whether user i evaluated each item. Similarly, I_j is a diagonal matrix whose diagonal components are the indicator vector $\{I_{1j}, I_{2j}, ..., I_{Nj}\}$. R_i is a rating vector $\{r_{i1}, r_{i2}, ..., r_{iM}\}$. Similarly, R_j is a rating vector $\{r_{1j}, r_{2j}, ..., r_{Nj}\}$. Based on (8) and (9), U and V are updated by stochastic gradient descent to obtain the optimal user matrix U and item matrix V.

However, W and W^+ can not be optimized in the same way as U and V because they are closely related to the features of CNN architecture, such as the max pooling layer and nonlinear activation function. Therefore, we temporarily fix U and V and use the error back-propagation method to estimate W and W^+.

4 Experiment

4.1 Goal, dataset, and experiment configuration

We compare the performance of the proposed and existing methods using three different real-world datasets to check the feasibility of our proposed method. We first explain the data set, comparison method, and evaluation metrics. Then, we discuss the experimental results.

Table 1 Dataset details

Dataset	users	items	ratings	Density (%)
amazon	12,062	14,182	29,547	0.0173
rakuten	6163	3026	23,129	0.124
yelp	53,391	2634	224,461	0.160

This experiment uses the Amazon dataset [7] for clothes, the Rakuten dataset for women's [15] accessories,[1] and the Yelp dataset for restaurants in British Columbia.[2] These datasets are believed that user preferences are more likely to be reflected in reviews. The user description text is the reviews posted by each user, the item description text for Rakuten is the item description text prepared by each store, and for Amazon and Yelp, the item description text is a concatenated document of the reviews posted for the item.

The evaluation value for each dataset takes a value from 1 to 5. As this experiment deals with text data, items without text representing the item and users without text representing the user were excluded from the data set. In addition, users who rated only one item were excluded because we could not split the data into training and test data. As a result of these processes, the statistics for each data set are shown in Table 1. In the actual experiments, the data set was randomly divided into training, validation, and test data in a ratio of 8:1:1.

For each text, the following preprocessing was performed as in ConvMF [5]: (1) set the maximum length of raw documents to 300 words, (2) removed stop words, (3) calculated the TF-IDF score for each word, (4) removed corpus-specific stop words that have a document frequency higher than 0.5, (5) selected top 8000 distinct words as vocabulary, and (6) removed all non-vocabulary words from raw documents.

In this experiment, we adopted root mean square error (RMSE) as evaluation index, and took an average of five trials to ensure reliability.

$$RMSE = \sqrt{\frac{\sum_{i,j}^{N,M} \left(r_{ij} - \hat{r_{ij}}^2\right)}{ratings}}, \tag{10}$$

where $\hat{r_{ij}}$ is the predicted score of user i for item j, and ratings is the total number of scores.

We compare Double-ConvMF with the following base lines:

Table 2 Overall test RMSE

Model	Amazon	Rakuten	Yelp
PMF	1.995	1.446	1.838
ConvMF	1.501	1.019	1.610
ConvMF+	1.489	1.0127	1.608
Left-ConvMF	1.761	1.299	1.853
Left-ConvMF+	1.791	1.298	1.778
Double-ConvMF	**1.316**	0.9308	1.471
Double-ConvMF+	1.333	**0.9128**	**1.467**
Improve	11.6%	9.9%	8.8%

- PMF [3]: Probabilistic matrix factorization is a standard method of matrix factorization that only uses user's ratings.
- ConvMF [5]: Convolutional matrix factorization is a method that extracts features from item description text using CNN and incorporates them into PMF.
- Left-ConvMF:Left-ConvMF is a method that extracts features from user description text using CNN and incorporates them into PMF.
- Double-ConvMF:Double-ConvMF is our proposed method that extracts features from item and user description text using CNN and incorporates them into PMF.

In addition to the above methods, ConvMF+, Left-ConvMF+, and Double-ConvMF+, in which a pre-trained model called Globe [16], was applied to each method, were also used as competitors.

To find the best values for λ_U and λ_V, a grid search was conducted in the range of [1,25,50,75,100].

The parameters of the other experiments were set as follows. These parameters are followed to ConvMF [5] The dimensionality D of the user matrix U and the item matrix V is set to 50. The maximum number of words in each document is set to 300. The dimensionality of the pre-trained word embedding model is set to 300. The dropout rate used in training the CNN architecture is set to 0.2.

4.2 Experimental results

Table 2 shows the RMSE of each model. Here, the bold value means the best value of methods including our proposed methods, "Improve" stands for the percentage improvement between the best value of Double-ConvMF or Double-ConvMF+ and the best value of the compared methods. From this table, we can see that the best accuracy is obtained by the proposed method Double-ConvMF or Double-ConvMF+ in any dataset. This result suggests that using the user and item description text and incorporating it into the matrix factorization is effective. Then, we compare

[1] https://rit.rakuten.com/data_release/

[2] https://www.yelp.com/dataset

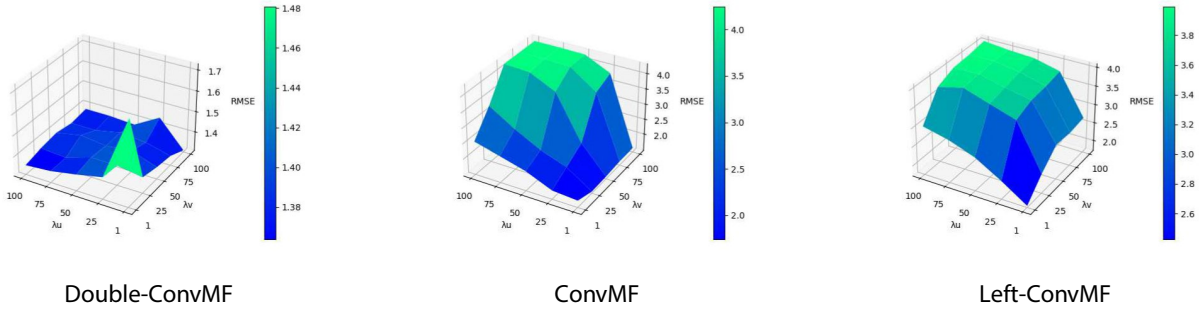

Fig. 2 Parameter analysis λ_U and λ_V on Amazon Dataset

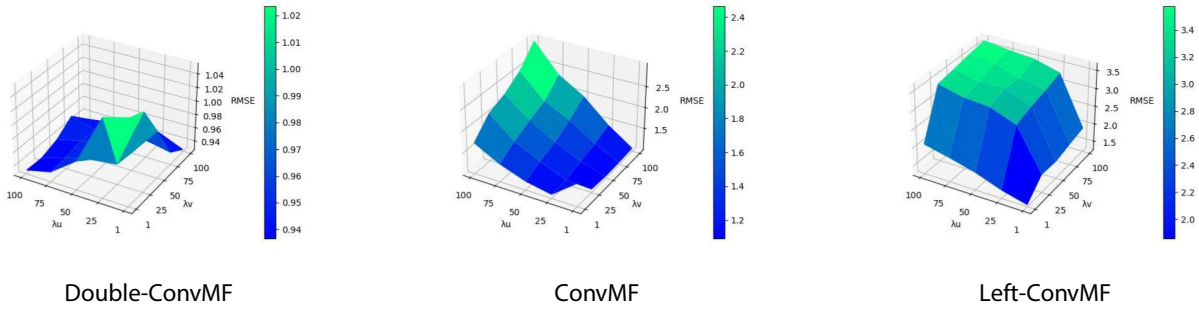

Fig. 3 Parameter analysis λ_U and λ_V on Rakuten Dataset

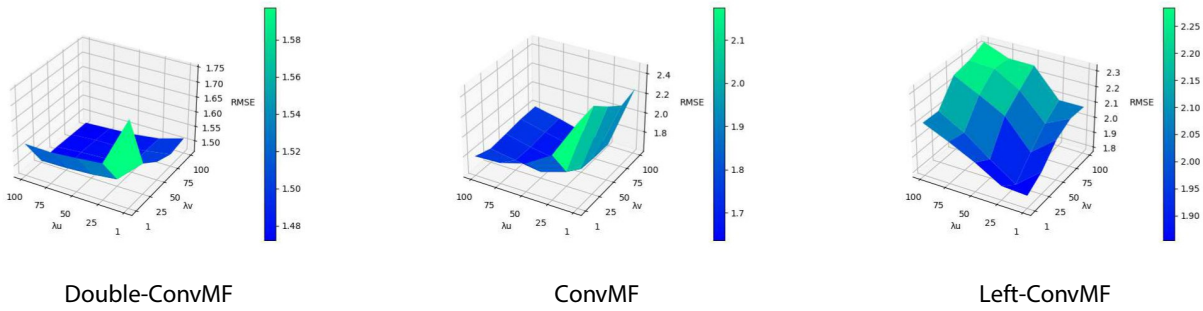

Fig. 4 Parameter analysis λ_U and λ_V on Yelp Dataset

the Improve values for each dataset: 12% for the Amazon dataset, 9.9% for the Rakuten dataset, and 8.8% for the Yelp dataset, indicating that the improvement rate increases as the dataset becomes more sparse. This result shows that our proposed method works particularly well on a sparse dataset and is effective in improving the cold-start problem.

4.2.1 Influence of user and item description text

To examine the effectiveness of the CNN architecture given to users and items respectively, we calculated the RMSE of ConvMF and Left-ConvMF for each dataset in terms of improvement over PMF, a plain matrix factorization model. In the Amazon dataset, ConvMF was 24.5% and Left-ConvMF was 11.8%, in the Rakuten dataset, ConvMF was 29.5% and Left-ConvMF was 10.2%, and in the Yelp dataset, ConvMF was 12.5% and Left-ConvMF was −0.82%. The overall trend showed a higher improvement rate in ConvMF using CNN architecture on the item side than in Left-ConvMF using CNN architecture on the user side. Therefore, this result indicates that capturing item relationships more powerfully is effective. However, in all datasets, including

Table 3 RMSE by number of reviews

Dataset	Under 2	3 Reviews	4 Reviews	Over 5
Amazon	1.356	1.356	1.309	1.143
Rakuten	–	0.9368	0.8773	0.9687
Yelp	1.510	1.591	1.089	1.322

the Yelp dataset where the Left-ConvMF value was worse than the PMF value, our proposed method, Double-ConvMF, gave the best value, suggesting that the use of both item and user description text complementarily enhanced the model.

4.2.2 Impact of pre-training model

We discuss the effectiveness of the pre-training model in Double-ConvMF. We use Glove [16] as the pretraining-model. Glove is used for initializing embedding layer of the CNN. As shown in Table 2, the improvement rate when changing from Double-ConvMF to Double-ConvMF+ is − 1.3% for the Amazon dataset, 1.9% for the Rakuten dataset, and 0.27% for the Yelp dataset. This result indicates that the pre-training model is more effective in datasets with a large number of users than items. The reason is that the context-supplementing ability of the prior learning model compensates for the small amount of user data.

4.2.3 Parameter analysis

Figures 2, 3, and 4 show the relationship between λ_U and λ_V and RMSE for each method in each data set. The overall trend is that for Double-ConvMF, RMSE increases as both λ_U and λ_V become smaller. This result suggests that U and V may have fallen into the local optimum solution when λ_U and λ_V were small. However, in the case of ConvMF and Left-ConvMF except for ConvMF on Yelp, RMSE improved only when λ_U and λ_V were small, while RMSE deteriorated in many other cases.These results show that our proposed method is more robust than existing methods. We believe that the reason is that applying the CNN architecture to both the item and user sides optimizes U and V in a balanced manner. However, we need to be careful not to fall into the trap of locally optimal solutions.

4.2.4 Impact of number of reviews

Table 3 illustrates RMSE values for the proposed Double-ConvMF method when users are categorized by the number of reviews in each dataset. Specifically, the column "under2" represents the group of users with two or fewer reviews, while the column "over5" represents users with five or more reviews. It is important to note that in the Rakuten dataset, users with two or fewer reviews were preprocessed and subsequently excluded from the dataset, resulting in blank entries for this category.

Overall, there is a noticeable trend in groups with a review count of 4 or more, particularly the over5 group, tend to exhibit improved RMSE values compared to groups with fewer reviews. This suggests that as users contribute more reviews, our proposed model can better capture their preferences, leading to enhanced accuracy.

However, it is worth highlighting that the Rakuten dataset exhibits results that deviate from this general trend. This suggests the possibility of dataset-specific dependencies, warranting further investigation and careful scrutiny to gain a deeper understanding of the underlying factors contributing to these variations.

5 Conclusion

This paper proposed a method to capture contextual information from the text of user and item description using CNN and incorporate this information into the matrix factorization. Then, we explained the algorithm for optimizing the equation obtained from the prior distribution equation. The experimental results using the three datasets showed an improvement of 8.8% - 12%. This suggests that using the user and item description text at the same time is more effective than using the item description text alone.

In the future, we would like to develop models which use not only text information but also other supportive information such as images, social networks into PMF.

References

1. Saga R, Duan Y (2018) Apparel goods recommender system based on image shape features extracted by a CNN. In: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp 2365–2369
2. Saga R, Hayashi Y, Tsuji H (2008) Hotel recommender system based on user's preference transition. In: 2008 IEEE International Conference on Systems, Man and Cybernetics, pp 2437–2442
3. Salakhutdinov R, Mnih A (2007) Probabilistic matrix factorization. Proceedings of the 20th international conference on neural information processing systems, ser NIPS 07. Curran Associates Inc, pp 1257–1264
4. Shi Y, Larson M, Hanjalic A (2010) Mining mood-specific movie similarity with matrix factorization for context-aware recommendation. Proceedings of the workshop on context-aware movie recommendation ser CAMRa 10. Association for Computing Machinery, pp 34–40. https://doi.org/10.1145/1869652.1869658
5. Kim D, Park C, Oh J, Lee S, Yu H (2016) Convolutional matrix factorization for document context-aware recommendation. Proceedings of the 10th ACM conference on recommender systems, ser RecSys 16. Association for Computing Machinery, Cham, pp 233–240. https://doi.org/10.1145/2959100.2959165
6. McAuley J, Leskovec J (2013) Hidden factors and hidden topics: understanding rating dimensions with review text. Proceedings

of the 7th ACM conference on recommender systems, ser RecSys 13. Association for Computing Machinery, Cham, pp 165–172. https://doi.org/10.1145/2507157.2507163
7. McAuley J, Targett C, Shi Q, van den Hengel A (2015) Image-based recommendations on styles and substitutes. Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval, ser SIGIR 15. Association for Computing Machinery, Cham, pp 43–52. https://doi.org/10.1145/2766462.2767755
8. Ma H, Yang H, Lyu MR, King I (2008) SoRec: social recommendation using probabilistic matrix factorization. Proceedings of the 17th ACM conference on information and knowledge management, ser CIKM 08. Association for Computing Machinery, Cham, pp 931–940. https://doi.org/10.1145/1458082.1458205
9. Wang X, Yang X, Guo L, Han Y, Liu F, Gao B (2019) Exploiting social review-enhanced convolutional matrix factorization for social recommendation. IEEE Access 7(82):826–837
10. Beel J, Langer S, Gipp B (2017) TF-IDuF: a novel term-weighting scheme for user modeling based on users personal document collections, pp 452–459. https://kops.uni-konstanz.de/handle/123456789/41879. Accessed 21 March 2018
11. Musto C, Semeraro G, Degemmis M, Lops P (2015) Word embedding techniques for content-based recommender systems: an empirical evaluation. In: RecSys Posters
12. Zhang J-D, Chow C-Y (2018) Sema: deeply learning semantic meanings and temporal dynamics for recommendations. IEEE Access 6(54):106–116
13. Wu C, Wu F, An M, Huang Y, Xie X (2019) Neural news recommendation with topic-aware news representation. Proceedings of the 57th annual meeting of the association for computational linguistics. Association for Computational Linguistics, pp 1154–1159
14. Liu P, Du J, Xue Z, Li A (2022) Bi-convolution matrix factorization algorithm based on improved convmf. In: Zhang L, Yu W, Jiang H, Laili Y (eds) Intelligent networked things. Springer Nature, Singapore, pp 122–134
15. Rakuten group Inc (2021) Rakuten ichiba data. Informatics research data repository, national institute of informatics (dataset). Rakuten group Inc. https://doi.org/10.32130/idr.2.1
16. Pennington J, Socher R, Manning C (2014) GloVe: global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). Association for Computational Linguistics, pp 1532–1543