# A Matrix Factorization Technique with Word Embedding for Recommendation

Shoujian YU<sup>+</sup>, Yuewen Zhang, Dehua Chen

School of Computer Science and Technology, DongHua University, Shanghai, China

**Abstract.** In recent years, machine learning has achieved great success in computer vision and natural language processing. However, little attention has been paid to how to apply these explorations in the field of natural language processing to recommendation systems. In this paper, inspired by word embedding techniques, we take advantage of word embedding techniques and propose a word embedding based recommendation method. Specifically, word2vec is first used to train a corpus composed of item name and item attributes, the implicit item information is obtained, and then incorporate them into a matrix factorization model, which combines both latent and pre-learned features for recommendation. We have performed experiments on the MovieLens dataset, and the results show that our algorithm performs better than traditional matrix factorization algorithms and has a higher prediction rate.

**Keywords:** Recommendation System, Matrix Factorization, Collaborative filtering, Word2vec, Word Embedding

### 1. Introduction

With the increasing popularity of the Internet and the ever-increasing amount of information, the recommendation system is widely used. From traditional advertisements, book and movie recommendations to currently popular short videos, and real-time recommendation based on geographical coordinates, the recommendation system has a very wide range of applications, and in the field of e-commerce, recommendation systems occupy a pivotal position.

Typically, in a recommender system, we have a set of *users* and a set of *items*. Each user u rates a set of items by some values. The recommender has the task to predict the rating for user u on a non-rated item i or to generally recommend some items for the given user u based on the ratings that already exist. In the existing personalized recommendation models, collaborative filtering, which make recommendations based on ratings that already exist, are widely used due to their effectiveness and high scalability. Collaborative filtering is mainly divided into two categories: neighbourhood methods and model-based methods [1]. Neighbourhood methods is divided into user-based or item-based method [2]. The main idea of user-based CF is to use the user purchase record to obtain the set of users that is most similar to the active user, so as to recommend the items for the other user in this set. While item-based approaches predict the ratings of active users based on the computed information of items similar to those chosen by the active user. Model-based CF used the observed user-item ratings to train a compact model that explains the given data, so that ratings could be predicted via the model instead of directly manipulating the rating database, which learn the parameters of a model and store only those parameters, such as the latent factor approach, probabilistic matrix factorization [3], regularized singular value decomposition [4]. Among them, the regularized SVD is effective and high scalability, which is widely used in practical recommendation systems [5].

Motivated by the success of word embedding techniques [6,7,8], we first pre-train the word embedding model word2vec to incorporate additional attributes of items, and then incorporate them into the MF based

<sup>&</sup>lt;sup>+</sup> Corresponding author. Tel.: + 86 21 67792809.

E-mail address: jackyysj@dhu.edu.cn.

model. By combing the latent and pre-learned network features together, our method not only can make use of the rating of use on items, but also incorporate additional attributes of items. Experimental results on MovieLens dataset demonstrate the effectiveness of our proposed approach.

### 2. Recommendation Method

### 2.1. Matrix Factorization

In this paper, we consider an  $m \times n$  rating matrix R explaining n user's ratings on m items, where  $r_{ui}$  denotes the ratings of user u on item i. The task of the recommendation system is similar to a matrix-filling game. We use observed  $r_{ui}$  values to predict the unknown ratings data. A low-dimensional matrix factorization approach seeks to approximate the ratings matrix R by a multiplication of l-dimensional factors,

$$R \approx U_{\mu}^{T} I_{i}, \tag{1}$$

Where  $U_u$  and  $I_i$  are *l*-dimensional latent vectors representing user *u* preference and item *i* profile, respectively, and *l* is the number of latent factors for both users and items, which is also the rank in the latent factor model.

The Singular Value Decomposition (SVD) method seeks the optimal implicit vector representation of users and items by minimizing the square of the loss function. At the same time, in order to avoid overfitting, two regularization terms are added into Equation 1. Hence, we have

$$L = \min \frac{1}{2} \sum_{(u,i)} I_{ij} (r_{ui} - U_u^T I_i)^2 + \frac{\lambda_U}{2} \|U\|^2 + \frac{\lambda_i}{2} \|I\|^2$$
(2)

Where  $I_{ij}$  is the indicator function that is equal 1 if user  $U_u$  rated item  $I_i$  and equal to 0 otherwise, and  $\lambda_U$  and  $\lambda_I$  are two non-negative tuning parameters. The optimization problem in Equation 2 minimizes the sumof-squared-errors objective function with quadratic regularization terms.

#### **2.2.** Proposed Method

The Regularized Singular Value Decomposition (RSVD) method in recommender systems normally seek to factorize the user-item rating matrix into two low rank user-specific and item-specific matrices, and then utilize the factorized matrices to make further predictions. The Regularized SVD algorithm is only based on working on the user-item rating matrix and ignore the abundant attributes of item. But in our real life, when two items have many of the same attributes, the latent characteristics of the items are also very close. For movie recommendation, when the attributes of the two movies are the same, such as the genres, the release date and other attributes, the latent characteristics of the movies are also similar.

Motivated by the success of word embedding techniques, one key strategy of the proposed method is to pre-train the corpus consisting of item name and item attributes to learn features, and then incorporate them into SVD to improve prediction accuracy through word embedding technique. We first pre-train corpus consisting of item name and attributes of items, to obtain the word vector corresponding to the item name through the word2vec model. When two items have the same attributes, such as genres, the release date and other attributes, the two item names often appear in the same contextual context, the vectors corresponding to the items are similar. Therefore, we can recognize that the latent factors of the two items are also similar. After obtaining the word vector corresponding to each item through the word2vec model and fusing these pre-trained features with the latent features from collaborative filtering model linearly, we can arrive at our word embedding based recommendation method SVD\_WE:

$$L = \min \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij} \left( r_{ij} - U_i^T I_j - W_j^T X_j \right)^2 + \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_I}{2} \|I\|_F^2 + \frac{\lambda_W}{2} \|W\|_F^2$$
(3)

where  $W_j \in \mathbb{R}^d$  is the weight vector that indicates how much the pre-trained item attributes should contribute to item *j*.

We apply to stochastic gradient descent method to find a local minimum of Eq. 3, and for every observed rating  $r_{ij}$ , we have the following stochastic updating rules to learn the all the latent parameters:

$$U_{i} \leftarrow U_{i} + \gamma_{1} (\Delta_{ij} I_{j} - \lambda_{U} U_{i})$$

$$I_{j} \leftarrow I_{j} + \gamma_{2} (\Delta_{ij} U_{i} - \lambda_{I} I_{j})$$

$$W_{j} \leftarrow W_{j} + \gamma_{3} (\Delta_{ij} X_{j} - \lambda_{W} W_{j})$$
(4)

where

$$\Delta_{ij} = r_{ij} - U_i^T V_j - W_j^T X_j.$$
<sup>(5)</sup>

### 3. Experiments

#### **3.1.** Experimental Setup

In all the experiments conducted in this paper, the dimension of latent factors U, I is set as  $\kappa = 30$ . The dimension of the weight matrixes W and X for extracted features is set as d = 20. The regularization parameters of U, I, W are set as  $\lambda_U = \lambda_I = 0.02$ ,  $\lambda_W = 0.001$ . At the time, the learning rates is set  $\gamma_1 = \gamma_2 = \gamma_3 = 0.01$ . In this paper, we use two popular metrics, the Mean Absolute Error (MAR) and the Root Mean Square Error (RMSE), to measure the prediction quality of our proposed approach in comparison with other collaborative filtering methods. The matric MAE is defined as:

$$MAE = \frac{\sum_{i,j} |r_{ij} - \hat{r_{ij}}|}{N} \tag{6}$$

where  $\hat{r}_{ui}$  denotes the rating user *i* gave to item *j* as predicted by a method,  $r_{ui}$  denotes the rating user *i* gave to item *j*, and *N* denotes the number of tested ratings.

RMSE provides intuitive indicators of recommendation quality and is one of the most commonly used recommendation quality indicators. The formula of the RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i,j} (r_{ij} - \hat{r}_{ij})^2}$$
(7)

From the definitions, the smaller the RMSE, the fewer the errors between the true value and the predicted value, which means that the prediction is more accurate.

To illustrate the effectiveness of our SVD\_WE model, we compare it with the following baselines:

- SVD: The singular value decomposition (SVD) method. this is the baseline matrix factorization approach. It only used user-item matrix for recommendation. The details of this method are also introduced in Section 2.
- NMF: this method is originally proposed in [9] for image analysis. However, it is widely used in collaborative filtering. It only uses user-item matrix for recommendations.
- PMF: this method is proposed by Salakhutdinov and Minh. It also only uses user-item matrix for recommendations.

#### **3.2.** Experimental Results

Table 1 reports the RMSE and MAE values of all comparison partners on the MovieLens data set. We note that we do not compare our models with Social MF method, because our dataset does not include the information of social networks [10]. From the results are shown in Table 1, we can observe that our methods can generate better results than the others recommendation algorithm. Table 1 shows that SVD\_WE significantly outperforms its competitor with a RMSE of 0.7096 for the dataset. As NMF, PMF and SVD only uses the rating information for recommendation, it does worse than our methods. Since our method incorporates the attribute information of items to provide more item features, which are not from rating matrix, it can generate better results than the state-of-the-art recommendation algorithm SVD.

Metrics	NMF	PMF	SVD	SVD_WE
MAE	0.6992	0.6963	0.5639	0.5602
RMSE	0.8924	0.8879	0.7142	0.7096

Table. 1: Performance Comparison

In this experiment, in order to further compare the recommendation accuracy of the SVD\_WE model with the traditional SVD model, the size of latent dimensions was in a range of  $\{10,20,30,40,50\}$ , which was setting to verify the influence of different K values on predicting accuracy. In all case, from the results are shown in Table 2, we can observe that our proposed method SVD\_WE can perform better than SVD and reach the better performance, which indicates that fusing the pre-trained embedding with latent factors is helpful, and can effectively model both the latent features of users and items. The results show that the larger the K value, the

higher the prediction accuracy, which is understandable, because when the K value reaches a certain level, it is equivalent to the original scoring matrix without matrix decomposition. However, when the K value is very large, it conflicts with the purpose of matrix decomposition. Therefore, in actual experiments, we should choose the appropriate K value through multiple experiments.

Dimension	Metrics	SVD	SVD_WE
10	MAE	0.6369	0.6364
10	RMSE	0.8064	0.8053
20	MAE	0.5958	0.5955
20	RMSE	0.7544	0.7541
20	MAE	0.5639	0.5602
50	RMSE	0.7142	0.7096
40	MAE	0.5375	0.5331
40	RMSE	0.6815	0.6763
50	MAE	0.5131	0.5069
50	RMSE	0.6509	0.6435

Table. 2: Impact of the size of latent dimension K

### 4. Conclusion

In this paper, we applied word embedding techniques to learn latent representations of items. In this work, additional sources of information were incorporated to improve collaborative filter-based recommender systems (e.g., item title, item genres). We first pre-train the pre-processed corpus to learn latent representations of the items in corpus, and then incorporate them into the MF based model. From the result, we can also find that our proposed method SVD\_WE can perform better than other method, and reach the best performance in experiments, which indicates that fusing the pre-trained word embedding with latent factors is helpful and can effectively model the feature of items. This model can not only integrate the item attribute information, but also provide a new and effective method from the field of natural language processing to how to introduce other auxiliary information to improve the recommendation efficiency. If there is an accessible user or item relationship network, this kind of word embedding and fusion information to improve the recommendation rate may solve the cold start problem to a certain extent.

## 5. Acknowledgements

This paper is supported by supported by the National Key R&D Program of China under Grant 2019YFE0190500.

## 6. References

- Y. Koren, R. Bell, C. Volinsky. Matrix Factorization Techniques for Recommender Systems[J]. *Computer*. 2009, 42(8): 30-37.
- [2] X. Su, T. M. Khoshgoftaar. A Survey of Collaborative Filtering Techniques[J]. *Advances in Artificial Intelligence*, 2009, :1-199.
- [3] R. Salakhutdinov and A. Mnih. Probabilistic Matrix Factorization. in Proceedings of the 20th International Conference on Neural Information Processing Systems, USA, 2007, pp. 1257–1264.
- [4] Toroslu Ġ H. A singular value decomposition approach for recommendation systems[J]. The graduate school of natural and applied sciences of middle east technical university, Ph. D. Dissertation, 2010.
- [5] Ben Dai, Junhui Wang, Xiaotong Shen. A smooth collaborative recommender system[J]. *The Journal of Machine Learning Research (JMLR)*, 2019, 20(16):1-24.

- [6] M. Ailem, A. Salah, and M. Nadif, Non-negative Matrix Factorization Meets Word Embedding, in Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, 2017, pp. 1081–1084.
- [7] Barkan O, Koenigstein N. Item2Vec: Neural Item Embedding for Collaborative Filtering[J]. 2016.
- [8] Mikolov, Tomas; et al. Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781
- [9] Lee, D. D., and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. Nature 401.6755(1999):788.
- [10] Y. Wen, L. Guo, Z. Chen, and J. Ma, Network Embedding Based Recommendation Method in Social Networks, 2018, pp. 11–12.