Journal of Database Management
Volume 30 • Issue 2 • April-June 2019

Collaboration Matrix Factorization on Rate and Review for Recommendation

Zhicheng Wu, Beijing Jiaotong University, Beijing, China
Huafeng Liu, Beijing Jiaotong University, Beijing, China
Yanyan Xu, School of Information Science and Technology, Beijing Forestry University, Beijing, China
https://orcid.org/0000-0001-7174-6588
Liping Jing, Beijing Jiaotong University, Beijing, China

ABSTRACT

According to the sparseness of rating information, the quality of recommender systems has been greatly restricted. In order to solve this problem, much auxiliary information has been used, such as social networks, review information, and item description. Convolutional neural networks (CNNs) have been widely employed by recommender systems, it greatly improved the rating prediction's accuracy especially when combined with traditional recommendation methods. However, a large amount of research focuses on the consistency between the rating-based latent factor and review-based latent factor. But in fact, these two parts are completely different. In this article, the authors propose a model named collaboration matrix factorization (CMF) that combines a projection method with a convolutional matrix factorization (ConvMF) to extract the collaboration between rating-based latent factors and review-based latent factors that comes from the results of the CNN process. Extensive experiments on three real-world datasets show that the projection method achieves significant improvements over the existing baseline.

KEYWORDS

Collaboration Matrix Factorization (CMF), Convolutional Neural Network, Probabilistic Matrix Factorization, Projection Method

INTRODUCTION

With the increasing amount of data in real life, recommendation systems (Davidson 2010) play a more and more important role. It is very important to model the user's preference and some work (Zhou 2018, Lu 2018) had been published. Recommender systems are most often based on collaborative filtering and the widely used approach is the use of latent factor. Matrix factorization (MF) (Lee 2001) is the most popular model to derive the latent factor. In addition, one classical method is probabilistic matrix factorization (PMF) (Salakhutdinov, 2007).

However, with the explosive growth of users and items on the Internet, the rating matrix becomes more and more sparse which destroys the performance of recommendation system. To fix the sparse problem, deep learning was employed into Matrix factorization. As it can effectively capture the non-linear and non-trivial user-item relationships, and enable the codification of more complex abstractions as data representations in the higher layers, it is not surprising that even though deep learning based recommender system is a new comer, it achieves high recommendation quality such as

DOI: 10.4018/JDM.2019040102

Neural Collaborative Filtering (NCF), Information Retrieval GAN (Generative Adversarial Networks), CCCFNet (Content-Boosted Collaborative Filtering Neural Network) model (He 2017, Wang 2017, Lian 2017) and Deep Matrix Factorization models (Xue 2017).

Although deep learning models can extract the high-level features, it still cannot fix the sparse problem perfectly because when compared to the enormous quantity of users and items, the quantity of ratings is too small. Therefore several recommendation techniques had been proposed that consider not only rating information but also auxiliary information such as demography of users, social networks, content information and review information (Kawale, 2015, Ling, 2014, Mcauley, 2013, Purushotham, 2012, Wang, 2011, Wang, 2015). On the one hand, based on social network (Yang 2017) proposed a model that consists truster part and trustee part and (Guo 2015) proposed a model that combines SVD++ with social trust information. On the other hand, based on content information, (Wang, 2011) proposed collaborative topic regression (CTR) that combines topic modeling (LDA) (Blei, 2016) and collaborative filtering in a probabilistic approach. Most recently, (Wang, 2015) proposed collaborative deep learning (CDL) that integrates Stacked Denoising Auto-Encoder (SDAE) (Vincent, 2010) into PMF. In order to fully capture document information and take contextual information into consideration, CNN (Krizhevsky, 2012) is used in processing of text information. Based on review information, DeepCoNN (Yu, 2017) adopts two parallel convolutional neural networks to model user behaviors and item properties from review texts. (Shen, 2016) built an e-learning resources recommendation model. It uses CNN to extract item features from text information of learning resources such as introduction and content of learning material, and follows the same procedure of (Dieleman, 2013) to perform recommendation. However, the use of review in both items and users would lead to the reuse of this information. Therefore ConvMF (Yu, 2016) combines CNN with PMF in a similar way as CDL and only use review in item's part.

ConvMF and other content-based models have been proved to improve the recommendation system's performance, but they focus on the consistency between the rating-based latent factors and the review-based latent factors. It is true to consider the consistency because rating information and review information talk about the same set of items, but they are from different sources. Thus, rating information and review information should provide different knowledge and such individual knowledge could complement to each other. Thus, the authors combine the consistency and complementarity and propose a new conception that is called collaboration. In this paper, the authors propose a model called CMF (Collaboration Matrix Factorization) to make use of these two resources.

To demonstrate the effectiveness of CMF, a series of experiments are conducted on three real-world datasets. Our experimental results have shown that CMF has ability to achieve perfect recommendation result. In addition, the long tail items are also taken into consideration and the result on the long tail items also has great improvement.

Our contributions are summarized as follows:

- An architecture named CMF is proposed, which integrates mapping method, matrix factorization and deep feature learning. It models the mappings between the rating-based latent factors used in PMF and the review-based latent factors in CNN models;
- Many experiments are done on three real-world datasets to evaluate the performance of our model. Our model outperforms the existing baselines;
- Apart from the whole datasets, the authors also take the long tail items into consideration because the long tail items make more profits and be more meaningful than other items. The experimental result shows that our model is better.

The remainder of the paper is organized as follows. Section 2 gives an overview of the related models. Section 3 proposes the CMF model. Section 4 experimentally evaluates CMF and discuss the evaluation results. Section 5 summarizes our contributions and gives future work.