



ICAC3'15

Matrix Factorization Model in Collaborative Filtering Algorithms: A Survey

Dheeraj Bokde^a, Sheetal Girase^b, Debajyoti Mukhopadhyay^{c*}

^{a, b, c} Department of Information Technology
Maharashtra Institute of Technology, Pune 411038, India

Abstract

Recommendation Systems (RSs) are becoming tools of choice to select the online information relevant to a given user. *Collaborative Filtering* (CF) is the most popular approach to build Recommendation System and has been successfully employed in many applications. Collaborative Filtering algorithms are much explored technique in the field of Data Mining and Information Retrieval. In CF, past user behavior are analyzed in order to establish connections between users and items to recommend an item to a user based on opinions of other users. Those customers, who had similar likings in the past, will have similar likings in the future. In the past decades due to the rapid growth of Internet usage, vast amount of data is generated and it has become a challenge for CF algorithms. So, CF faces issues with sparsity of rating matrix and growing nature of data. These challenges are well taken care of by *Matrix Factorization* (MF). In this paper we are going to discuss different Matrix Factorization models such as Singular Value Decomposition (SVD), Principal Component Analysis (PCA) and Probabilistic Matrix Factorization (PMF). This paper attempts to present a comprehensive survey of MF model like SVD to address the challenges of CF algorithms, which can be served as a roadmap for research and practice in this area.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of organizing committee of the 4th International Conference on Advances in Computing, Communication and Control (ICAC3'15)

Keywords: Collaborative Filtering ; Matrix Factorization ; Recommendation System

* Corresponding author.

E-mail address: debajyoti.mukhopadhyay@gmail.com

1. Introduction

In these days E-Commerce industry is growing at the exponential rate. As the users are using these systems their changing needs and variety of products is making business more and more complex. At the same time they are providing ease and flexibility to users'. In such environment, customers have difficulty to find optimal information about products from the tremendous amount of information. To help the buyers, the major e-business companies are going to develop their Recommender System to help their customers to choose items more efficiently; this serves win-win strategy in E-Commerce.

Recommendation Systems are becoming tools of choice to select the online information relevant to a given user. Recommendation System can be classified into: Content-Based (CB), Collaborative Filtering (CF) and Hybrid Recommendation System^{1,2}. CF is the most popular approach to build recommender system and has been successfully employed in many applications. CF is much explored technique in the field of Data Mining and Information Retrieval. In *Collaborative Filtering* (CF), past user behavior are analyzed in order to establish connections between users and items to recommend an item to a user based on opinions of other users'. Customers who had similar likings in the past, will have similar likings in the future. Many E-Commerce companies have already incorporated RS with their services. Examples for such recommendation systems include Product and Book recommendation by Amazon, Movie recommendations by Netflix, Yahoo!Movies and MovieLens, Product advertisements shown by Google based on the search history.

For large and complex data, CF methods frequently give better performance and accuracy than other RS techniques. Early CF algorithms for recommendation systems utilize the association inferences, which have a very high time complexity and a very poor scalability. Recent methods that use matrix operations are more scalable and efficient. The implementations and algorithms of CF for the applications of recommendation systems face several challenges. First is the size of processed datasets. The second one comes from the sparseness of rating matrix, which means for each user only a relatively small number of items are rated. So, these challenges are been well taken care by Matrix Factorization^{2,3}.

Matrix Factorization (MF) methods have recently received greater exposure, mainly as an unsupervised learning method for latent variable decomposition and dimensionality reduction³. It has successfully applied in spectral data analysis and text mining. Most of the MF models are based on the latent factor model. In a latent factor model^{2,3}, rating matrix is modeled as the product of a user factor matrix and an item factor matrix. The Matrix Factorization approach is found to be most accurate approach to reduce the problem from high levels of sparsity in RS database, certain studies have used dimensionality reduction techniques.

MF is specially used for processing large RSs databases and providing scalable approaches. In the model-based technique Latent Semantic Index (LSI) and the dimensionality reduction method Singular Value Decomposition (SVD) are typically combined^{2,3,14}. SVD and PCA are well-established technique for identifying latent semantic factors in the field of Information Retrieval to deal with CF challenges.

1.1. Related Work

This paper mainly study the Matrix Factorization models like SVD and PCA, with CF algorithms such as user-based and item-based CF. As we know that from past two decades lots of research work is going in the field of CF. CF is a promising research field in Information Retrieval so many researchers have contributed to this area.

Many CF researchers have recognized the problem of sparseness i.e., many values in the ratings matrix are null since all users do not rate all items. Computing distances between users is complicated by the fact that the number of items users have rated in common is not constant. An alternative to inserting global means for null values or significance weighting is Singular Value Decomposition (SVD), which reduces the dimensionality of the ratings matrix and identifies latent factors in the data⁸.

In 2006, the online DVD rental company Netflix announced the Netflix Prize contest with a \$1 million reward to the first team who can improve its recommender system's root mean square error (RMSE) performance by 10 percent or more^{3,19}. Contestants were allowed to build model based on released training set consisting of about 100 million movie ratings, on a scale of 1 to 5 stars, submitted by 500,000 anonymous users on more than 17,000 movies. The participating teams need to submit their predicted ratings for a test set consisting of approximately 3

Download English Version:

<https://daneshyari.com/en/article/486117>

Download Persian Version:

<https://daneshyari.com/article/486117>

[Daneshyari.com](https://daneshyari.com)