F

Foundational Recommender Systems for Business

Prageet Aeron
https://orcid.org/0000-0003-3957-5912
Management Development Institute, Gurgaon, India

INTRODUCTION

The role of the Internet in commerce has steadily risen with a large number of consumers and firms now completely transacting online. As a result, the need for better exploration and search for appropriate products or rather the most suitable products is an important activity. Starting out as simple catalogues and moving to advanced search, today's recommender systems are advanced prediction systems helping users identify the right products and helping firms improve their revenues and profitability. Recommender systems are also touted as efficient as they help in faster search and lesser transaction times for everyone involved. Since the recommender systems are such a primary component of today's online organizations, it is no surprise that it has been the subject of constant scrutiny and research with state of the art technologies and strategies being developed by the firms to get an edge over their competitors through better and better recommender systems. Right from Amazon to Netflix to Alibaba, no e-commerce website can hope to make it big without recommender systems. Therefore, students of business, executives involved in business and practitioners of technology all have stakes in learning and helping develop better recommender systems.

The present work briefly discusses the basic concepts and explores the state of art techniques in this important techno-commercial space. Author starts with neighborhood based collaborative filtering, subsequently the author discusses model based collaborative filtering, and then discussion proceeds to dimension reduction and latent factor models. Finally, the author touches upon the need and basic concepts of integrating various models and concludes the article with a brief discussion.

BACKGROUND

Any recommender system is likely to either identify top-k users or items or identify missing user-item ratings and both the formulations and all have their own advantages and disadvantages depending upon the business context (Aggarwal, 2016; Zanker, Feidrich & Jannach, 2011). Before delving deeper, it is important to understand the user-rating matrix which is the most important construct in the study of recommender systems. The users could be denoted by the rows and items by the columns with items being marked as per the obtained rating from a user. Therefore, for, m "user" and n "item" scenarios, we have a m x n matrix as the user-rating matrix. In other words, r(i,j) belongs to the m x n matrix, it represents the rating given to jth item by the ith user. A word on rating types is also necessary for the readers to fully comprehend the richness and at times appreciate the complexity of the problem being handled. Ratings in practical systems could be binary (Yes/No or Positive/Negative), interval-based

DOI: 10.4018/978-1-7998-9220-5.ch167

(Likert scale) or continuous ratings. Next we start with discussing neighbourhood based collaborative methods in detail.

NEIGHBOURHOOD BASED COLLABORATIVE FILTERING

Neighbourhood based collaborative filtering algorithms are amongst the earliest algorithms for the purpose of developing recommendations. The foundational idea of these algorithms is that similar consumers are likely to rate on similar lines and at the same time similar items have a higher probability of achieving similar ratings. As has been discussed before, the process of computing similarity is a primary activity among these algorithms. Let us discuss the steps involved in the execution of the neighborhood-based collaborative filtering algorithm and simultaneously develop an understanding about these algorithms. For assessing similarity between users we can utilize various similarity indices such as raw cosine similarity, adjusted cosine similarity or the Pearson's coefficient. Pearson's coefficient is considered better for the algorithm as it has a bias-adjustment effect as compared to other cosine indices. Another variation to similarity index is achieved by using a discount factor in cases when the common ratings are below a threshold value.

User-Based Model

Assuming we utilize the Pearson's coefficient; the first step is calculating the mean rating for each user given the ratings of items excluding the items for which rating is absent. Next we identify users in the user-rating matrix who have rated the same items and Pearson's coefficient is calculated taking into account the mean ratings. The Pearson coefficient is computed between the target user and all the other users and the k highest Pearson coefficient with respect to the target user are identified, and weighted average of the ratings of the missing item for target user are then calculated. Pearson coefficient for assessing similarity, however, can be calculated only across the intersection of the set of item ratings between the two users or in other words only if an item has been rated by both the users. Mathematically speaking, for the ith user,

$$\mu_i = \frac{\sum_j r_{ij}}{\left(numberofitemswithratings\right)} \forall i \left\{1...m\right\} \tag{1}$$

where j is an index representing items. Assuming intersection of set of items rated by user "u" and "v" by $u \cap v$,

$$Pearson's Coefficient\left(u,v\right) = \frac{\sum_{j \epsilon u \cap v}\left(r_{uj} - \mu_u\right) * \left(r_{vj} - \mu_v\right)}{\sqrt{\sum_{j \epsilon u \cap v}\left(r_{uj} - \mu_u\right)^2}\sqrt{\sum_{j \epsilon u \cap v}\left(r_{vj} - \mu_v\right)^2}} \tag{2}$$

To further improve the results one may employ the mean centered rating of items for each user and subsequently compute the weighted mean rating for missing user-item rating. Many authors have also suggested usage of normalized or standardized rating as predicted output rather than mean centered