Information Sciences 378 (2017) 59–74

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

DiABlO: Optimization based design for improving diversity in recommender system

Anupriya Gogna*, Angshul Majumdar

Indraprastha Institute of Information Technology – Delhi, Okhla Phase – III, New Delhi,110020, INDIA

ARTICLE INFO

Article history:
Received 30 September 2015
Revised 10 October 2016
Accepted 22 October 2016
Available online 24 October 2016

Keywords:
Collaborative filtering
Recommender systems
Diversity
Matrix factorization
Blind compressive sensing

ABSTRACT

Primary task of a recommender system is to improve user's experience by recommending relevant and interesting items to the users. To this effect, diversity in item suggestion is as important as the accuracy of recommendations. Existing literature aimed at improving diversity primarily suggests a 2-stage mechanism – an existing CF scheme for rating prediction, followed by a modified ranking strategy. This approach requires heuristic selection of parameters and ranking strategies. Also most works focus on diversity from either the user or system's perspective. In this work, we propose a single stage optimization based solution to achieve high diversity while maintaining requisite levels of accuracy. We propose to incorporate additional diversity enhancing constraints, in the matrix factorization model for collaborative filtering. However, unlike traditional MF scheme generating dense user and item latent factor matrices, our base MF model recovers a dense user and a sparse item latent factor matrix; based on a recent work. The idea is motivated by the fact that although a user will demonstrate some affinity towards all latent factors, an item will never possess all features; thereby yielding a sparse structure. We also propose an algorithm for our formulation. The superiority of our model over existing state of the art techniques is demonstrated by the results of experiments conducted on real world movie database.

1. Introduction

Explosive growth in the number of e-commerce (business-to-client (B2C)) portals has powered deep interest in recommender systems (RS) [7,33]. The primary task of RS is to suggest personalized products and services (from a large repository) to users. Good (accurate) suggestions results in customer satisfaction which translates to certain tangible benefits for the e-commerce portal, namely higher profits, increased customer satisfaction and goodwill.

Most works in RS design [3,6,26,27,30,44] focus on improving prediction accuracy; quantified using measures such as Mean absolute error (MAE) and precision-recall curves. Recommender systems based on improving the said metrics suggest products and services that are highly aligned with a user's past preferences i.e. recommended items share a high degree of similarity with those liked by the user in the past. For example, if a user has preferred horror movies in the past, it is likely that the recommender system will suggest only horror movies. Such an accuracy centric design, with emphasis on recommending only similar items will compromise on diversity, and over time customers may lose interest in the suggested item list. It has been reported that heterogeneity or variety in the recommendations can improve customers' interest in the recommender system [2].

* Corresponding author.
E-mail addresses: anupriyag@iiitd.ac.in (A. Gogna), angshul@iiitd.ac.in (A. Majumdar).

http://dx.doi.org/10.1016/j.ins.2016.10.043
0020-0255/

High customer satisfaction translates into higher profits for online portals; however, this is not the only driver for the e-commerce portal. Most portals (e.g. Amazon in US and India, Flipkart in India) act as 'marketplace' / 'network' where retailers (sources) sell their products online. If only common items based on popularity (albeit customized) are suggested to clients, sellers of niche products and services would not be recommended. Eventually, this would lead such sellers to stop procuring the portal's services.[1] Therefore, it is the responsibility of the e-commerce portal to improve the visibility of such niche products by improving diversity in recommendations by suggesting a broader spectrum of products (and services) to the users.

Even for e-commerce portals selling their own items (e.g. Netflix, Apple Music, Gaana) the trade-off between accuracy and diversity is apparent as well. There is a financial cost in acquiring an item (however niche it might be); in order to break even (and eventually profit) the e-commerce portal must suggest it to the users. Recommendations solely based on accuracy would fail to suggest niche items. It is imperative that in order to suggest such items, some amount of diversity needs to be in-built into the recommendation algorithm.

In light of the above arguments, it can be safely summarized that characteristics of a good RS extend beyond conventional accuracy measures, placing emphasis on comprehensive recommendation criteria.

1.1. Problem motivation

The necessity for diversified recommendations, as highlighted above, and corroborated in existing studies [28,46] suggests the need for a broader measure of recommendation quality, than captured by the usual accuracy centric metrics. Motivated by such goals works in the recent past [34,32,39] concentrated on improving novelty, diversity, and visibility for long tail items. However, there is an inherent trade-off between accuracy and novelty / diversity. Accuracy of RS focuses on recommending items highly similar to a user's past preference whereas, diversity promotes digression from the users' past choices; these are conflicting traits. In consequence, the challenge in design of an effective recommendation process is to maintain sufficient relevance to a user's past choice (i.e. high accuracy) while making novel and diversified recommendations.

Most current studies [5,37] aimed at improving diversity, propose a two-stage approach – first stage involves prediction of missing ratings and the second is centered on generating a recommendation list. Such studies use off-the-shelf collaborative methods [10,12,14,21,23,24] for the first stage (prediction). Steps are taken to improve diversity in the second stage during generation of the recommendation list; once, the ratings on previously unrated items are predicted (in the first stage via collaborative filtering), the items are ranked using a diversity-promoting ranking strategy. This differs from conventional RS, where the predicted ratings are ranked in decreasing order – aimed at maximizing accuracy (albeit at the cost of diversity); and the top-N items are recommended. Studies focused on augmenting diversity, adopt a modified ranking strategy. It involves selection of a heuristic measure of relevance (ranking threshold), items rated above which (based on predicted ratings/score) are considered as candidates for revised ranking scheme.

There are several existing works that propose strategies to rank the predicted ratings to enhance diversity. Authors in [48] use topic diversification to generate a hybrid recommendation list which has weighted contributions from accuracy as well as diversity promoting measures. In [32], items are segregated into multiple clusters such that items in a cluster are highly dissimilar. Subsequently, attention is given to each cluster in the final recommendation list such that due diversity can be provided. Others, like [5], have used notions such as item popularity, average ratings and variance of ratings to devise ranking strategies that provide desired diversification.

The key drawback of all the said approaches is that they are ad hoc two-stage models which entail heuristic selection of ranking thresholds and strategies. Such heuristic approach without cross-talk between the two stages is unable to ensure optimality of solution. In this work, we propose a mathematically/theoretically sound optimization based solution for addressing the accuracy-diversity predicament.

1.2. Our contribution

In this work, we present a modified latent factor model to provide diverse, yet sufficiently accurate, recommendations. Our model is based on incorporating two conflicting concepts, one promoting accuracy and other diversity, in a unified optimization framework. The combined framework is used to predict the missing ratings. The predicted rating values, hence obtained, are sorted in decreasing order (similar to conventional RS model) thus eliminating need for empirically derived ranking strategies.

Our design provides improved diversity from both user's as well as system's perspective while maintaining due coherence with a user's past preference (i.e. sufficiently high accuracy). We are also able to substantially augment the visibility of long tail (less popular) items. This is credited to our optimization centric design which ensures optimality of solution, improving upon the existing works, which is reflected in the results highlighted in subsequent sections.

Some existing works promoting diversity require secondary information such as item metadata [46,47] (required to classify them into distinct/diverse clusters), or require extracting secondary information from ratings [5] (computing item's

[1] This trend is visible in developing countries like India. Large e-commerce portals are only being used for common items. Niche products and services such as traditional craftsmen are moving away from these large corporate e-commerce portals and starting their own website.