Physica A 458 (2016) 364–377

Contents lists available at ScienceDirect

Physica A

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

SibRank: Signed bipartite network analysis for neighbor-based collaborative ranking

Bita Shams, Saman Haratizadeh *

Faculty of New Sciences and Technologies, University of Tehran, North Kargar Street, 1439957131, Tehran, Iran

HIGHLIGHTS

- SibRank is a new recommendation algorithm based on similarity among users' ranking.
- SibRank models users' ranking as a novel signed bipartite network structure.
- SibRank exploits signed multiplicative rank propagation for similarity calculation.
- SibRank is able to calculate similarity between users without any common ranking.
- SibRank improves NDCG@10 up to 5% compared to other collaborative ranking methods.

ARTICLE INFO

Article history:
Received 31 December 2015
Received in revised form 10 April 2016
Available online 22 April 2016

Keywords:
Recommender system
Collaborative ranking
Signed network
Similarity measure
Preference data
Personalized ranking

ABSTRACT

Collaborative ranking is an emerging field of recommender systems that utilizes users' preference data rather than rating values. Unfortunately, neighbor-based collaborative ranking has gained little attention despite its more flexibility and justifiability. This paper proposes a novel framework, called SibRank that seeks to improve the state of the art neighbor-based collaborative ranking methods. SibRank represents users' preferences as a signed bipartite network, and finds similar users, through a novel personalized ranking algorithm in signed networks.

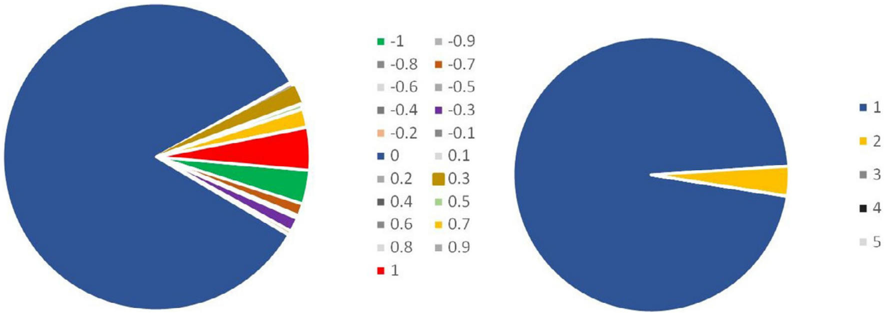
1. Introduction

Recommendation systems exploit users' historical data in order to suggest a small set of relevant services among a large volume of irrelevant ones. Traditionally, these systems seek to predict user's interests through gathering and prediction of ratings given by him to services or products. In recent years a variation of recommender systems, called collaborative ranking, has emerged that focuses on ranking data instead of ratings [1,2] to find the best items to recommend.

Rankings seem to be more informative and reliable source of information to reflect users' preferences and interests [2,3]: users' ratings commonly follow a baseline that is a function of time [4]. For instance, a user that assigns a rate of 4 to an average movie "*A*" now, might tend to assign 3 to that movie in the future. However, in both times, he probably would not prefer that average movie over an excellent movie "*B*". So, he persistently ranks "*B*" over "*A*" while the rates he gives to those movies may change over time. Furthermore, the goal of a recommender system is to find and recommend the most relevant items [3,5,6]. Therefore, it is more important to accurately predict the user's priorities rather than the absolute ratings he would give to items. For example, consider a case that a target user would rate the movies "*A*" and "*B*" with 4 and

* Corresponding author.
E-mail address: haratizadeh@ut.ac.ir (S. Haratizadeh).

http://dx.doi.org/10.1016/j.physa.2016.04.025



(a) Distribution of Kendall correlation of users.

(b) Distribution of number of common pairwise comparisons between the most similar users.

Fig. 1. Statistical properties of Kendall Correlation in a sample dataset of movielense100 K containing 10 ratings for each user. (a) Distribution of Kendall correlation of users. (b) Distribution of the number of common pairwise preferences among users whose Kendall Correlation is one.

5 respectively, that is $r(A) = 4$ and $r(B) = 5$. Suppose that some rate prediction algorithm, predicts ratings $r(A) = 2$ and $r(B) = 3$, while another algorithm predicts $r(A) = 5$ and $r(B) = 4$. Although, in rating-oriented frameworks, the latter one is considered more accurate in the sense of rate prediction, it is clear that the former one would correctly rank "*B*" higher than "*A*" and will probably make a better recommendation.

Recently, several researchers have approached recommendation from a ranking-oriented perspective. These approaches can be categorized into two groups: model-based collaborative ranking methods (MCR) and neighbor-based collaborative ranking methods (NCR). MCR techniques try to learn the latent factors of users and items while optimizing a ranking criterion (e.g. Normalized Discounted Cumulative Gain (NDCG)) [1,5,7] while NCR algorithms exploit the concept of users' similarity to infer a ranking over the user's uncollected items [2,6,8–11].

Neighbor-based collaborative ranking has not been investigated as much as model-based ranking. The reason is possibly that it is not straightforward to calculate the similarity among users when each user's profile is a set of preferences [12]. Most of the current NCR algorithms have exploited Kendall correlation that takes into account users' agreements and disagreements over pairwise comparisons [2,10,11]. However, Kendall measure, has originally been proposed to compare similarity between total orders (i.e. ranking over a set of distinct items) [12]. Therefore, it is not directly applicable to recommender systems where each user has partially ranked a different subset of items. To resolve this issue, current NCR techniques calculate users' similarity based on their common comparisons and ignore the information available through the uncommon rankings. This approach suffers from some shortcomings: First, Kendall correlation does not take into account the confidence or reliability of the source of information: the similarity calculated over a small set of common comparisons is not as reliable as one calculated over a large set of comparisons. Second, Kendall correlation between users with no common comparison is equal to zero, simply because this measure ignores all other available information in the data that can be used to discriminate between such users. These problems are more serious in sparse datasets where users rarely have a common pairwise comparison. As illustrated in Fig. 1(a), in a sparse dataset, Kendall correlation and its variants would be zero for more than 87% of pairs of users. Also, in average, more than 96% of the users among which the Kendall correlation is one, have only one common pairwise comparison (see Fig. 1(b)). That clearly indicates that when one uses Kendall correlation for finding similar users, not only all users with no common comparison with the target user are ignored, but also the recommendation is highly affected by those users who have only one common pairwise comparison with the target user.

This paper proposes a novel NCR framework, called SibRank, to resolve these issues. SibRank first constructs a signed bipartite network, called SiBreNet, to represent users' preferences. Then, it exploits a novel similarity measure, called SRank, to calculate similarity among the target user and all other users in SiBreNet.

The main contributions of this paper can be summarized as below:

- To our knowledge, this is the first work that takes advantages of signed networks to capture the similarity among users' ranking. This paper represents ranking data as a signed bipartite graph to comprise different kinds of information about agreement and disagreement of users over their preferences.
- We propose a novel similarity measure in signed networks that reflects a global view of agreements and disagreements between users.
- We conducted set of experiments to assess the performance of SibRank. The results show some improvements compared to one of the best current algorithms called EigenRank.

2. Related works

2.1. Neighbor-based collaborative ranking

The first neighbor-based collaborative ranking algorithm was proposed in Ref. [2]. The paper introduced a general framework, called EigenRank, which is now widely followed by other NCR techniques [8–10,13]. It finds similar users to