Applied Soft Computing 15 (2014) 21–29

Contents lists available at ScienceDirect

Applied Soft Computing

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

Enhancing scalability and accuracy of recommendation systems using unsupervised learning and particle swarm optimization

Sagarika Bakshi[a], Alok Kumar Jagadev[a], Satchidananda Dehuri[b,*], Gi-Nam Wang[c]

[a] Department of Computer Science and Engineering, Institute of Technical Education and Research, SOA University, Bhubaneswar, India
[b] Department of Systems Engineering, Ajou University, San 5, Woncheon-dong, Yeongtong-gu, Suwon 443-749, South Korea
[c] Department of Industrial Engineering, Ajou University, San 5, Woncheon-dong, Yeongtong-gu, Suwon 443-749, South Korea

ARTICLE INFO

Article history:
Received 14 October 2012
Received in revised form 30 August 2013
Accepted 23 October 2013
Available online 31 October 2013

Keywords:
Recommendation system
Collaborative filtering
Sparsity measures
Particle swarm optimization
Unified measures of sparsity
Alpha estimates

ABSTRACT

Recommendation system has been a rhetoric area and a topic of rigorous research owing to its application in various domains, from academics to industries through e-commerce. Recommendation system is useful in reducing information overload and improving decision making for customers in any arena. Recommending products to attract customers and meet their needs have become an important aspect in this competitive environment. Although there are many approaches to recommend items, collaborative filtering has emerged as an efficient mechanism to perform the same. Added to it there are many evolutionary methods that could be incorporated to achieve better results in terms of accuracy of prediction, handling sparsity as well as cold start problems. In this paper, we have used unsupervised learning to address the problem of scalability. The recommendation engine reduces calculation time by matching the interest profile of the user to its partitioned and even smaller training samples. Additionally, we have explored the aspect of finding global neighbours through transitive similarities and incorporating particle swarm optimization (PSO) to assign weights to various alpha estimates (including the proposed α_7) that alleviate sparsity problem. Our experimental study reveals that the particle swarm optimized alpha estimate has significantly increased the accuracy of prediction over the traditional methods of collaborative filtering and fixed alpha scheme.

1. Introduction

Recommendation system (RS) is a type of information filtering system that gives advice on products' information, or services that a user may be interested in. They assist users with the decision making process when choosing items with multiple alternatives [23,30]. RS is based on human social behaviour where opinions and tastes of known acquaintances are taken into consideration while making decisions. RS generates personalized recommendations that incorporate various mathematical, probabilistic [32], soft computing, and swarm based computing techniques [14]. For conducting research on recommendation system generally the researchers adopt to movie recommendation sites [38,7,33]. Recommendation systems have a wide range of applications in e-commerce [29,24], e-learning [43,10] and digital libraries [39,40]. Because of the information overload and its vigorous dissemination, it is a challenge for the researchers to design a RS that can take into account many demographic, contextual, and preference factors and predict useful items more efficiently. The crux of RS is a filtering algorithm involving knowledge, demographic, content, utility, collaborations or their combination, and when applied collectively are called hybrid RS. Among all the existing types of recommendation systems, those that employ collaborative filtering (CF) methods are widely used and are known to be effective methods of recommendation technique.

Generally, the quality of CF recommendation systems can be enhanced by improving the similarity measures and neighbourhood set which is the most critical component of the CF mechanism. Due to some CF limitations such as sparsity, scalability, and cold-start problems the recommendations' accuracy is extremely affected [21,15]. Although many researchers tried to address these issues, the CF systems are still in need of more improvements in overall recommendations accuracy and performance. Since CF is based on the ratings of the neighbours who have similar preferences, it is very important to select the neighbours properly. The more the similarity measure is improved, the more valuable neighbours and accurate recommendations are obtained.

Sparsity of ratings data is one of the basic challenges of CF systems. CF may not produce the desired result in the presence of sparse data where many users do not share common items.

* Corresponding author. Fax: +82 2 365 2579.
E-mail addresses: sagarikabakshiitr@gmail.com (S. Bakshi), alok.jagadev@gmail.com (A.K. Jagadev), satchi@ajou.ac.kr, satchi.lapa@gmail.com (S. Dehuri), gnwang@ajou.ac.kr (G.-N. Wang).

1568-4946/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2013.10.018

Therefore users must rate a sufficient number of items to get a reasonable profile overlap so that the system is able to establish its neighbours [21]. High sparsity data levels result in poor recommendation quality, hence they need to be taken care of. This issue is addressed in our approach by establishing transitive relationships between different neighbours.

Alongside, the crucial part of CF algorithms consists of the group of metrics used to determine the similarity between each pair of users [2,9], among which the Pearson correlation metric stands out as a reference for many models. Local similarity measures such as the Pearson similarity measure and cosine similarity measure [11], base the similarity computation step between two users, on the set of common items rated by both users.

The most important ingredient of accuracy of any filtering technique is the way it learns user models [5]. Effective prediction of ratings from a small number of examples is also important [2]. This can mitigate the problem of scalability to some extent. Hence to enhance the scalability of the recommendation systems, we have used *k*-means algorithm [27,34] for determining a small size of training set.

In this work, we propose a new α estimation scheme based on the overall sparsity in the ratings data as well as sparsity based on the active users and the items whose ratings need to be predicted. The dependency of α value on the user and item ensures that the local and global neighbours are weighed differently using PSO [31] for every user and for every prediction. The experimental results support our ideas and demonstrate that the proposed methods are superior to the fixed α scheme [36]. Our experimental results show that the prediction quality has increased significantly and even users with limited ratings are provided with a considerate number of recommendations. Further, the PSO based α estimates are comparatively better to genetic algorithm based α estimates in terms of running time and accuracy [6]. This argument is proved by experimental results and theoretically as well, because unlike genetic framework, PSO uses less number of parameters and operators.

The rest of the paper is structured as follows. In Section 2, we present the background of this work like traditional collaborative filtering approach for recommendation system and PSO. Section 3 presents our proposed system framework. Section 4 concerns with experimental study and results. Finally, Section 5 concludes this paper with future work.

2. Background

The recent widespread growths of information available on the Internet, additionally speeded up by social networks, have evoked the need for systems which can help users to find valuable information. In this context, information filtering based on CF can support users in their searches.

Collaborative filtering (CF) [20] is a method wherein opinions are obtained from people with similar preferences with respect to the user under observation. It is the phenomenon that users with similar tastes and choices in the past tend to agree in future with items not rated. The CF system represents users with their ratings on a set of items. The CF system selects a set of similar users according to a similarity or correlation measure. Then, it generates ratings predictions for items not rated by the target user. Finally, the system recommends the items with the highest predicted ratings. Due to some CF limitations such as sparsity, scalability, and cold-start problems the recommendations' accuracy is extremely affected [21,15].

Over the years, a variety of solutions to the data sparsity problem, have been proposed. However, a novel approach of combining predictions from locally similar neighbours, i.e., users who have co-rated items, and globally similar neighbours is proposed among them [36]. Two users are considered to be globally similar if they are connected through locally similar neighbours. It was established experimentally that when the data is highly sparse, globally similar neighbours provide better prediction, whereas in case of low sparsity, predictions from local neighbours are superior. A parameter alpha was utilized to balance contributions from global neighbourhood over local neighbourhood [36]. The value taken by the parameter α was static (0.5), i.e., the importance given to local/global predictions remained the same across all predictions.

Researchers have proposed several methods improving the similarity measure, since similarity measures influence the accuracy of predictions. Some of them are Proximity-Impact-Popularity (PIP) measure [3], UNION similarity [45], random walk counting [18] user-class similarity [49], and iterative message passing procedure [28]. All the proposed methods have their own lacunae and need improvements so as to overcome the challenges of collaborative filtering. Added to it, these methods are completely domain and application specific and they address a particular issue. For example, PIP is proposed to address the cold-starting problem. However, its significance is limited to the similarity calculation of traditional user-based CF. User-class similarity can only be used when class information is available and fail to achieve meaningful results in very large-scale data as the similarity update will take much more time. UNION similarity is used with sparse data; however, it does not take scalability into consideration. An iterative-clustered collaborative filtering (ICCF) recommendation system was proposed [1]. In ICCF, the spectral clustering is iteratively utilized in both user-based and item-based collaborative filtering to predict the unknown ratings. Accordingly, the ICCF successfully alleviates the sparsity and cold-starting problems. However, all users and items have an equal influence when calculating the similarity whereas the similarity measure must reflect the importance of different features. Thus it is really challenging to build such a system that addresses all the issues of CF and which is not application specific.

A number of adapted approaches generated improvements in accuracy, when features are weighted in the instance to instance distance computation [47]. In the CF, feature weighting methods assign a weight to each feature, user or item, that measures how important is the feature in the overall similarity. In this connection, Ujjin and Bently [46] have contributed a particle swarm optimization based RS to fine tune profile matching of each active user.

Diaz-Aviles et al. [16] have cast the problem of item recommendation to rank problem and then used PSO to learn ranking function by optimizing the combination of unique characteristics of users, items, and their interactions. Alam et al. [4] have built a web based implicit recommender system by proposing a hierarchical PSO clustering. They have grouped the user session by combining the best attribute of hierarchical and partitional clustering using PSO.

2.1. *Traditional collaborative filtering based recommendation system*

There are two basic entities concerned in a recommendation system. The *user* (also referred to as customer) is a person who uses the recommendation system to provide his or her opinion and receive recommendation about items. The *item* (also referred to as product) is being rated by users. The inputs of a recommendation system are usually arithmetic rating values, which express the users' opinion of items. Ratings are normally provided by the user and follow a specified numerical scale (Example, 1: bad to 5: excellent). The outputs of a recommendation system can be either predictions or recommendations. The following are the three main steps of recommendation system.