Expert Systems With Applications 92 (2018) 507–520

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques

Mehrbakhsh Nilashi[a,b,*], Othman Ibrahim[a], Karamollah Bagherifard[c]

[a] Faculty of Computing, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
[b] Department of Computer Engineering, Lahijan Branch, Islamic Azad University, Lahijan, Iran
[c] Young Researchers and Elite Club, Yasooj Branch, Islamic Azad University, Yasooj, Iran

ARTICLE INFO

Article history:
Received 22 July 2017
Revised 23 September 2017
Accepted 24 September 2017
Available online 29 September 2017

Keywords:
Recommender systems
Ontology
Clustering
Dimensionality reduction
Scalability
Sparsity

ABSTRACT

Improving the efficiency of methods has been a big challenge in recommender systems. It has been also important to consider the trade-off between the accuracy and the computation time in recommending the items by the recommender systems as they need to produce the recommendations accurately and meanwhile in real-time. In this regard, this research develops a new hybrid recommendation method based on Collaborative Filtering (CF) approaches. Accordingly, in this research we solve two main drawbacks of recommender systems, sparsity and scalability, using dimensionality reduction and ontology techniques. Then, we use ontology to improve the accuracy of recommendations in CF part. In the CF part, we also use a dimensionality reduction technique, Singular Value Decomposition (SVD), to find the most similar items and users in each cluster of items and users which can significantly improve the scalability of the recommendation method. We evaluate the method on two real-world datasets to show its effectiveness and compare the results with the results of methods in the literature. The results showed that our method is effective in improving the sparsity and scalability problems in CF.

1. Introduction

Finding information in large-scale websites is a difficult and time-consuming process. Artificial Intelligence (AI) approaches are appearing at the forefront of research in information retrieval and information filtering systems. Recommender systems are a good example of one such AI approach. They have emerged in the e-commerce domain and are one way to address this issue. Such systems have been developed to actively recommend relevant information to users (Jugovac, Jannach, & Lerche, 2017; Nilashi, Jannach, bin Ibrahim, Esfahani, & Ahmadi, 2016a), typically without the need for an explicit search query. The history of recommender systems dates back to 1979 in relation to cognitive science (Rich, 1979). These systems have been important tools among other application areas such as, information retrieval (Salton, 1989), tourism (Kabassi, 2010), management science (Murthi & Sarkar, 2003), approximation theory (Powell, 1981), consumer choice modeling in business and marketing (Lilien, Kotler, & Moorthy, 1992), and forecasting theories (Armstrong, 2001).

Collaborative Filtering (CF) systems are information retrieval systems that operate under the assumption a user will like the same data items that other users have liked in the past. These systems are particularly popular and have been applied in many online shopping websites (Nilashi, Jannach et al. 2016; Nilashi, Salahshour et al., 2016b). CF algorithms mainly aggregate feedback for items from different users and use the similarities between items and items (item-based) or between users and users (user-based) to provide recommendations to a target user (Nilashi, Jannach, bin Ibrahim, & Ithnin, 2015).

Basically, CF recommendation algorithms are based on two main categories which are model-based and memory-based methods (Adomavicius & Tuzhilin, 2005). Memory-based (or heuristic-based) methods, such as correlation analysis and vector similarity, search the user database for user profiles that are similar to the profile of the active user that the recommendation is made for. In this type of recommender systems, it is important that the user and item databases remain in system memory during the algorithm's runtime. Because heuristic-based approaches can make predictions based on the local neighbourhood of the active user, or can base their predictions on the similarities between items, these systems can also be classed into the user-based and item-based approaches (Sarwar et al., 2001).

* Corresponding author.
E-mail addresses: nilashi@liau.ac.ir (M. Nilashi), othmanibrahim@utm.my (O. Ibrahim), k.bagherifard@iauyasooj.ac.ir (K. Bagherifard).

https://doi.org/10.1016/j.eswa.2017.09.058
0957-4174/

Memory- and model-based approaches have some advantages and disadvantages for item recommendation. Sparsity has been one of the main difficulties associated with these approaches, whereas recommendation with high accuracy has been one of the important advantages of the memory-based approach. However, this approach is not scalable for current recommendation systems as their databases include huge numbers of items and users. In addition, memory-based methods are heuristics and the prediction and recommendations are based on the whole ratings provided by the users to the items. Hence, all ratings are required to be maintained in memory. This method is a typical approach for high recommendation accuracy based on CF, but is not scalable for large-scale websites that use the huge number of users and items in recommendation systems (Sarwar, Karypis, Konstan, & Riedl, 2002). According to Goldberg, Roeder, Gupta, and Perkins (2001), model-based methods for learning a model utilize the group selection of ratings which is then applied to provide rating predictions. In addition, model-based CF algorithms have been an alternative approach to *k*-NN to solve the scalability problem of memory-based method (Nilashi, Esfahani et al., 2016c). A probabilistic method is utilized for these systems and the unrated value of a user prediction is measured based on the ratings the user has given to other items. Model-based algorithms do not suffer from memory-based drawbacks and can create prediction over a shorter period of time compared to memory-based algorithms because model-based algorithms perform off-line computation for training. These techniques regularly make a concise rating pattern off-line. Model-based CF (e.g., Singular Value Decomposition (SVD)-based CF) improves the scalability and the efficiency problem (Koren, Bell, & Volinsky, 2009, Liu et al., Nilashi et al., 2015; Nilashi, bin Ibrahim, & Ithnin, 2014), but may lead to some problems such as decreasing the accuracy performance (Linden, Smith, & York, 2003).

Hence, in this study a new method is proposed based on CF method to overcome the sparsity and scalability problems in CF algorithms accordingly to improve the performance of recommender systems. In fact, the performance improvement is achieved using ontology (Shambour & Lu, 2012) and dimensionality reduction (Koren, 2008; Koren et al., 2009) techniques. At the moment, there is no implementation of recommender systems by the use of combining ontology and dimensionality reduction techniques to solve the scalability and sparsity issues of CF recommender systems. Accordingly, this research tries to develop a new recommendation system based on CF using ontology and dimensionality reduction techniques. In order to enhance the prediction accuracy and overcome the scalability issue of recommender systems, we propose to use ontology and SVD. Specifically, we develop the method for user- and item based CF. We use the items' ontology for the item-based semantic similarity calculation and SVD for the item- and user-based CF recommendation part. In comparison with the previous studies, in this research we:

- develop a new recommendation system using ontology and dimensionality reduction techniques.
- improve the accuracy of recommendation systems by alleviating the sparsity issue in item-based CF using ontology.
- improve the scalability of recommendation systems using dimensionality reduction techniques.

The remainder of this paper is organized as follows: In Section 2, we briefly introduce the related subjects for the development of the proposed recommender system. In Section 3, problem statement and our research contributions are presented. Section 4 presents research methodology. Section 5 provides method evaluation results. In Section 6, we provide the discussion. Finally, conclusion is provided in Section 7.

2. Background theories

In the following sub-sections, we present the related subjects for the development of the proposed recommender system. Since, the ontology, clustering, dimensionality reduction and CF are important components of the proposed method, a short introduction of them is presented.

2.1. CF recommendation methods

The recommendation systems generally are divided into three categories: CF, Content-Based Filtering (CBF) and hybrid method. CF techniques in recommender systems are particularly popular and have been applied in many online shopping websites (Liu et al., 2011; Nilashi et al., 2014). The key to successful collaborative recommendation lies in the ability to make meaningful associations between people and their product preferences, in order to assist the end-user in future transactions. Similarities between past experiences and preferences are exploited to form neighbourhoods of like-minded people from which to draw recommendations or predictions for a given individual user. Based on the genuine process of CF strategy (Schafer, Frankowski, Herlocker, & Sen, 2007), a target user in the website will receive recommendation list of items that other users, with similar tastes, liked in the past. All CF methods require the past ratings of users in order to predict and accordingly recommend items to the target user (Cheng & Wang, 2014). To do so, similarities between the users and items are calculated using the distance measures. As CF can be classified as user-based and item-based, accordingly the similarity calculation for these approaches will be different (Nilashi et al., 2014).

2.2. Clustering methods and CF

CF is one of the methods widely used in recommender systems using two different techniques, memory-based and model-based (Breese, Heckerman, & Kadie, 1998; Su & Khoshgoftaar, 2009). The memory-based depends on the entire rating which exists in the user-item matrix for forming neighbors of the active user to generate recommendation tailored to his/her preferences. In contrast, the model-based methods learn the models of recommendations from the entire ratings to generate the recommendation for the target user. The well-known machine learning techniques for this approach is clustering (Sarwar et al., 2002), probabilistic Latent Semantic Analysis (pLSA) (Hofmann & Puzicha, 2004), matrix factorization (e.g. SVD) (Koren et al., 2009) and machine learning on the graph (Zhou et al., 2008).

Since memory-based techniques are easy to understand, implement and be successfully utilized in the real world application, they are considered suitable methods in recommender systems. However, this method often fails in large-scale applications. The sparsity of user-item matrix that is resulted since the user only rates few items throughout a large database of items is one of the issues in this technique that cause this failure. Thus, calculated similarity between users/items is unreliable value because of the few overlapping ratings between them. Efficiency is another issue in memory–based CF because similarity between pairs of items or users is needed to be measured for finding their neighborhood. A line of studies has been conducted for overcoming this drawback of memory-based techniques by a model-based clustering approach for enhancing efficiency (Gong, 2010; He, Yang, & Jiao, 2011; Sadaei, Enayatifar, Lee, & Mahmud, 2016; Shinde & Kulkarni, 2012; Wang, 2012). Clustering method groups similar items or users into separate clusters to identify neighborhood Clustering techniques have been used either directly or as a preprocessing stage in recommender systems (Adomavicius & Tuzhilin, 2005; Gong, 2010; Nilashi et al., 2014; Pham, Cao, Klamma, & Jarke,