Information Processing and Management 53 (2017) 1171–1184

Contents lists available at ScienceDirect

Information Processing and Management

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

Item-network-based collaborative filtering: A personalized recommendation method based on a user's item network

Taehyun Ha, Sangwon Lee*

Department of Interaction Science, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul, 03063 Republic of Korea

ARTICLE INFO

Article history:
Received 6 October 2016
Revised 6 March 2017
Accepted 19 May 2017

Keywords:
Recommendation system
Collaborative filtering
Item network
Centrality indicators
Social network analysis

ABSTRACT

Recommendation systems are becoming important with the increased availability of online services. A typical approach used in recommendations is collaborative filtering. However, because it largely relies on external relations, such as items-to-items or users-to-users, problems occur when the relations are biased or insufficient. Focusing on that limitation, we here suggest a new method, item-network-based collaborative filtering, which recommends items through four steps. First, the system constructs item networks based on users' item usage history and calculates three types of centrality: betweenness, closeness, and degree. Next, the system secures significant items based on the betweenness centrality of the items in each user's item network. Then, by using the closeness and degree centrality of the secured items, the algorithm predicts preference scores for items and their rank orders from each user's perspective. In the last step, the system organizes a recommendation list based on the predicted scores. To evaluate the performance of our system, we applied it to a sample dataset of 196 Last.fm users' listening history and compared the results with those from existing collaborative filtering methods. The results showed that the suggested method performed better than the basic item-based and user-based collaborative filtering methods in terms of Accuracy, Recall, and F1 scores for top-*k* recommendations. This indicates that an individual user's item relations can be utilized to remedy the problems occurring when the external relations are biased or insufficient.

1. Introduction

Users interact with many items in their everyday lives, and their usage patterns become complex. Particularly since Ashton (2009) introduced the concept of the Internet of things in a presentation in 1998, increasingly various items are expected to become involved and connected in human lives. In this context, recommending items from among the overload of possibilities has become an important issue. In their early stage, recommendation systems were considered within a limited scope, such as documental information in a library, but their application areas have been extended to other information domains (Ekstrand, Riedl, & Konstan, 2011), such as music (Bogdanov et al., 2013), videos (Halvey, Vallet, Hannah, & Jose, 2014), movies (Pera & Ng, 2013), and e-commerce services (Liao & Chang, 2016). The development of a recommendation system that provides high quality items is now regarded as a key factor of successful businesses (Shi, Larson, & Hanjalic, 2014).

* Corresponding author.
E-mail address: upcircle@skku.edu (S. Lee).

http://dx.doi.org/10.1016/j.ipm.2017.05.003
0306-4573/

In general, recommendation systems are categorized into content-based and collaborative filtering methods. Content-based filtering methods recommend items by focusing on a descriptor for each item, whereas collaborative filtering methods recommend items by focusing on external relations, such as user-to-user or item-to-item relations. The performance of content-based filtering methods depends on the quality of the descriptors. Also, the performance can be biased if a user's items are only focused on specific preferences (over specialization and novelty problems) (Celma, 2010). It is known that pure collaborative filtering methods usually outperform pure content-based filtering methods (Balabanović & Shoham, 1997; Gurcan & Birturk, 2016; Lin, Yang, & Wang, 2014). To recommend items for a target user, collaborative filtering methods find similar users or items based on the target user's history (Lin et al., 2014).

Pure collaborative filtering methods also face several challenges, such as sparsity, scalability, and the gray sheep problem (Celma, 2010; Lee, Yang, & Park, 2004; Shi et al., 2014; Su & Khoshgoftaar, 2009). Because collaborative filtering methods rely primarily on external relations, problems occur when those external relations are insufficient or biased (scalability or gray sheep problems). To address those problems, several hybrid filtering methods and extended collaborative filtering methods have been suggested by researchers (Burke, 2002; Shi et al., 2014). When the external relations contain a problem, hybrid methods remedy it using an item descriptor (Burke, 2002). Some extended collaborative filtering methods incorporate a user's side information, such as social networks, tags, reviews, and comments, to remedy problems (Shi et al., 2014). However, those approaches are ineffective when a descriptor or side information cannot be secured.

The external relation dependency problem can be addressed with supplementary information. For example, if the external item-to-item relations contain a problem, an individual user's internal item-to-item relationship history could provide supplementary information to remedy the problem. By focusing on some significant items in an individual user's history and relating them to other items, rarely used items could be discovered (over specialization, novelty, and scalability problem). This method could be effective for users with unusual or inconsistent preferences because it considers only significant items for recommendations (gray sheep problem). In addition, the method mainly uses the external item-to-item relations for collaborative filtering and uses the internal item-to-item relations as a user's content descriptor. This implies that a collaborative filtering method can be extended in content-based perspectives without requiring any extra descriptors or information.

Therefore, we here suggest a new recommendation method, item-network-based collaborative filtering. We extend social network analysis methods and use them to analyze an individual user's internal item-to-item relations. The items and their relations are expressed as an *item network*. Our method analyzes an individual user's context by focusing on his/her item network. We use the network indicators of betweenness, closeness, and degree centralities to apprehend the items' characteristics from a network perspective, and we provide an application example for validation. We constructed item networks for 196 users based on a sample dataset and then created recommendations. We compared our recommendation results with those from existing collaborative filtering models and found several potential advantages in our method.

2. Background

2.1. Recommendation systems

Recommendation systems are any kind of system that provides a list of items that a user might like. In general, existing recommendation methods can be categorized into two types: content-based filtering and collaborative filtering. Content-based filtering methods retrieve items based on item descriptors. For example, music items can use the singer, genre, rhythm, and tempo as descriptors. If a user has many and various items, content-based filtering methods can perform well. However, if the items are insufficient or biased to certain types, the performance can be poor. On the other hand, collaborative filtering focuses on user-to-user or item-to-item relations. Collaborative filtering is subcategorized into user-based and item-based methods. User-based collaborative filtering methods assume that similar users will prefer similar items. The methods categorize similar users based on their histories and then retrieve items for users. Item-based collaborative filtering methods assume that users will prefer items similar to ones they have already chosen. Those methods categorize similar items based on the items' history and then retrieve them for users. Collaborative filtering methods have problems when the user-to-user or item-to-item relations are insufficient or biased.

Generally, pure collaborative filtering methods cover the shortcomings of pure content-based filtering methods (Balabanović & Shoham, 1997; Gurcan & Birturk, 2016; Lin et al., 2014). Also, collaborative filtering does not require items to have descriptors. Thus, not just for performance, but also for ease of implementation, many commercial sites such as Amazon.com and Last.fm use collaborative filtering methods, and many studies have investigated a way to improve the methods. Some of the well-known challenges include sparsity, scalability, and the gray sheep problem (Lee et al., 2004; Shi et al., 2014; Su & Khoshgoftaar, 2009). Sparsity happens when user-to-user or item-to-item relations are biased. In this situation, rarely used items are not reflected in the recommendations. Scalability problems occur when a recommendation system has to calculate similarity and preference scores for an enormous number of items. The gray sheep problem arises for users whose preferences are unusual or inconsistent with those of other users.

To address those problems, several approaches have been suggested. As for the sparsity problem, Liu, Hu, Mian, Tian, and Zhu (2014) pointed out the ineffectiveness of existing similarity measures and suggested new similarity measures to improve the accuracy of collaborative filtering. Kaleli (2014) proposed an entropy-based neighbor selection method which can remedy the sparsity problem, and Fernández-Tobías, Braunhofer, Elahi, Ricci, and Cantador (2015) considered user per-