Physica A 396 (2014) 66–76

Contents lists available at ScienceDirect

Physica A

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

Collaborative filtering recommendation algorithm based on user preference derived from item domain features

Jing Zhang, Qinke Peng *, Shiquan Sun, Che Liu

Systems Engineering Institute, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi Province, China

HIGHLIGHTS

- We use item domain features to construct user preference models.
- We combine user preference models with CF for personalized recommendation.
- We use the multi-attribute decision making method to calculate the user preference.
- Our method integrates domain characteristics into a personalized recommendation.
- Our method aids to detecting the implicit relationships (missed by CF) among users.

ARTICLE INFO

Article history:
Received 14 December 2012
Received in revised form 1 October 2013
Available online 21 November 2013

Keywords:
Personalized recommendation
Collaborative filtering
User preference model
Item domain feature

ABSTRACT

Personalized recommendation is an effective method for fighting "information overload". However, its performance is often limited by several factors, such as sparsity and cold-start. Some researchers utilize user-created tags of social tagging system to depict user preferences for personalized recommendation, but it is difficult to identify users with similar interests due to the differences between users' descriptive habits and the diversity of language expression. In order to find a better way to depict user preferences to make it more suitable for personalized recommendation, we introduce a framework that utilizes item domain features to construct user preference models and combines these models with collaborative filtering (CF). The framework not only integrates domain characteristics into a personalized recommendation, but also aids to detecting the implicit relationships among users, which are missed by the conventional CF method. The experimental results show our method achieves the better result, and prove the user preference model is more effective for recommendation.

1. Introduction

The rapid development of information technology and the current growth and popularity of the Internet have facilitated an explosion of information that has exacerbated the information overload problem [1]. As one of the most useful methods, personalized recommendation, which was first proposed in the 1990s [2,3], adopts knowledge discovery techniques such as data mining and machine learning to discover user interests according to user behavior and then to make recommendations [4–6]. A typical application of personalized recommendation is in electronic commerce, such as book recommendations in Amazon.com [7], movie recommendations in Netflix.com [8], video recommendations in TiVo.com [9], and so on. An efficient recommendation system not only provides appropriate recommendations for users, but also helps the service provider gain substantial profits.

Mainstream recommendation algorithms can be divided into four categories [10]: content-based (CB), collaborative filtering (CF), network-based (NB), and hybrid recommendation (HR). The CB method recommends objects that are similar

* Corresponding author. Tel.: +86 29 82667964.
E-mail addresses: zjhappy@stu.xjtu.edu.cn (J. Zhang), qkpeng@mail.xjtu.edu.cn (Q. Peng).

0378-4371/$ – see front matter
http://dx.doi.org/10.1016/j.physa.2013.11.013

to those previously preferred by the target user. However, this method cannot filter audio, image, or video information [10]. CF has been the most successful recommendation system technology [11]. In CF, we make recommendation according to the assumption that users who have the similar performances would like to choose the similar items. However, the performance of CF is significantly limited by the sparsity of data [10]. NB recommendation utilizes relationships between users and items or relationships among users to construct a network, and then analyzes the network to determine recommendations for users indirectly. However, the "cold-start" problem could not be solved [10]. Finally, HR is currently the most popular approach and it combines at least two recommendation algorithms to determine a recommendation [10].

Many scholars have recently integrated various kinds of information into the recommendation system to improve performance. Such information includes tags, time, trust relationships, browse records, social networks, and so on. For instance, Zheng and Li investigated the importance and usefulness of tags and time information when predicting user preferences and consequently examined how such information could be exploited to build an effective resource–recommendation model [12]. Yin et al. considered the latent value of trust relationships to construct a trust preference network to make recommendations [13]. Kardan et al. introduced an innovative architecture for a recommendation system that took advantage of collaborative tagging and concept maps [14]. Zhang et al. proposed a recommendation approach that combined content and relation analyses in a single model to estimate the relations among users, tags, and resources for tag, item, and user recommendations [15]. Adding information to the recommendation system not only improves the performance, but also enhances the understanding of which factors influence recommendations.

The social tagging system is currently popular with scholars who utilize user-created tags to depict user preferences for personalized recommendation. Kim et al. proposed a CF method to provide enhanced recommendation quality with user-created tags, which were employed to identify and filter user preferences for certain items [16]. Shang et al. studied a personalized recommendation model that used the ternary relations among users, objects, and tags to propose a similarity based on preferences and tagging information [17]. Zhang et al. proposed a recommendation algorithm based on an integrated diffusion on user–item–tag tripartite graphs [18]. Schenkel et al. proposed an incremental threshold algorithm that considered both social ties among users and semantic relations among different tags [19]. Nakamoto et al. created a tag-based contextual CF model, where tag information was taken as user profiles [20]. Tso-Sutter et al. proposed a generic method that enabled tags to be incorporated to the standard CF by reducing ternary correlations into three binary correlations and then applying a fusion method to re-associate these correlations [21]. However, the user-created tag data is very sparse because of human descriptive habits and the diversity of language expression. For instance, a number of users prefer "happy" to express their delight, whereas others prefer "pleasure". Likewise, some users are accustomed to using "awful" to show their dissatisfaction, whereas others prefer "terrible". This characteristic hinders the identification of users who have similar interests through a social tagging system.

Additionally, different domain items often exhibit different characteristics. For instance, performance and quality are important for electronics; genres, directors, and actors compose the main information for movies; style of music and the scope of service are other primary factors that people concern. These characteristics are called domain characteristics in this paper. Given the diversity of domain characteristics, traditional personalized recommendations do not adapt well to all domains. Thus, methods to combine domain characteristics and personalized recommendation are required.

Recently, scholars focus on domain recommendation system. For example, Chen et al. presented a diabetes medication recommendation system based on domain ontology about drug attributes and patient symptoms [22]. García-Crespo et al. presented a semantic expert hotel recommendation system based on consumer experience and hotel characteristics [23]. Xin et al. presented a financial information recommendation algorithm that combined fuzzy clustering and CF [24]. Carrer-Neto et al. presented a social-based content recommender system in the movie domain, which used Semantic Web principles to help users find content that was relevant to their preferences, and likewise proposed social networks for building a novel CF [25]. Wei et al. introduced a news recommender system in which each user was considered as a node of the network. Users could post and recommend news to others, and they also could receive news from others at the same time [26]. However, these recommendation systems only adapt to one special domain very well. Thus, developing the general method that can adapt to every domain is more meaningful. Besides, although domain characteristics are always used to enhance item feature descriptions, there are few studies that have focused on expanding user preferences based on domain characteristics of items.

In this paper, in order to find a better way to depict user preferences and integrate domain characteristics into recommendation system, we propose a framework in which we use item domain features to construct user preference models at first, and then combine these models with CF for domain recommendations. Moreover, this framework could make recommendations to users who have not selected any common items with others. Finally, the framework is applied to the movie domain and exhibits good performance.

2. Methods

2.1. Comparison of the capacity in finding neighbors by using tags and ratings

Given the differences in users describing habits and the diversity of language expression, different users use different words or phrases to describe the same feelings. This characteristic hinders the identification of users who have similar