Extracting Core Users Based on Features of Users and Their Relationships in Recommender Systems

Li Kuang, School of Software, Central South University, Changsha, China
Gaofeng Cao, School of Software, Central South University, Changsha, China
Liang Chen, School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China

ABSTRACT

As an effective way to solve information overload, recommender system has drawn attention of scholars from various fields. However, existing works mainly focus on improving the accuracy of recommendation by designing new algorithms, while the different importance of individual users has not been well addressed. In this paper, the authors propose new approaches to identifying core users based on trust relationships and interest similarity between users, and the popular degree, trust influence and resource of individual users. First, the trust degree and interest similarity between all user pairs, as well as the three attributes of individuals are calculated. Second, a global core user set is constructed based on three strategies, which are frequency-based, rank-based, and fusion-sorting-based. Finally, the authors compare their proposed methods with other existing methods from accuracy, novelty, long-tail distribution and user degree distribution. Experiments show the effectiveness of the authors’ core user extraction methods.

KEYWORDS

Core Users Extraction, Interest Similarity, Popular Degree, Recommender System, Resource, Trust Influence, Trust Relationships

1. INTRODUCTION

With the development of Internet technology, various kinds of information online provides colorful and convenient life to people, and they are now used to reading news and updates of friends from Internet, buying books, commodities and various things online, watching movies and listening to music through Internet. However, on the other hand, with the explosion of information, more and more people find it hard to get the information that they really require and are interested in quickly and effectively. In such a background, recommender system is proposed and applied in many online platforms as an effective way to solve the problem of “information overload”. It can provide more personalized services by predicting potential interests according to users’ historical choices. Recommender systems have already been applied widely. For example, Amazon.com uses one’s purchase records to recommend books, Adaptive Info.com uses one’s reading history to recommend news, and the TiVo digital video system recommends TV shows and movies on the basis of users’ viewing patterns and ratings.

The recommender system is very helpful for filtering information, and the core of recommender system is the recommendation algorithm. Recommender system helps users make choices by the way
of information filtering, and a successful recommendation to one people may influence subsequent recommendation to other people. The influence is expanded greatly with successive recommendations. From such perspective, it can also explain the evolution of a popular movie.

Recommender systems typically produce a list of recommendations in one of the three ways: content-based (Peng, 2010), collaborative filtering (Chen, 2012; Li, 2010) and context-aware approaches (Adomavicius, 2011; Panniello, 2014). Content-based filtering is also referred to cognitive filtering, which recommends items based on a comparison between the content of the items and the user profiles. Collaborative filtering approaches are based on collecting and analyzing a large amount of information on users’ behaviors, activities or preferences and predicting what users will like based on their similarity to other users. Whereas context-aware approaches consider the influence of context factors, such as natural situation (time, location and etc.) and user’s profile (age, gender, profession and etc.), on users’ demand, preferences, and the selection and definition of neighbors. Based on the basic algorithms, many extensions have been made (Zheng, 2014; Habegger, 2014; Chen, 2010; Zhang, 2014; Xu, 2013; Zhang, 2015).

Different from the previous studies and approaches, the perspective of physics have also found applications in designing recommendation algorithms (Lü, 2011; Zhang, 2008; Zhou, 2010; Xu, 2014), and the most representative ones are mass diffusion and heat conduction. The mass diffusion (MD) algorithm is based on user-item bipartite networks, and it considers user’s interests as limited resources. If a user rates a lot of items, the user’s weight for each item will be reduced, but the user will have more opportunities to establish a connection with other users. The heat conduction (HC) method recommends items to an individual user through a process motivated by heat diffusion: items which are liked and disliked by a user are represented as hot and cold spots respectively, and recommendation is made according to the equilibrium “temperature” of the nodes in the user-item bipartite networks. The hybridization of the two propagation-based methods can effectively solve the diversity-accuracy dilemma in recommendation (Zhou, 2010).

Related work mainly focus on designing new recommendation algorithms, while the different importance of individual users in the recommendation process has not yet been well addressed. Some latest study has proved that there exists a group of core users who carry most of the information for recommendation. With them, the recommender system can already generate satisfactory recommendation. It is meaningful to identify core users in recommendation since it can improve the efficiency as well as the robustness of recommender system by excluding irrelevant and unreliable users. However, the proposed approach to extracting core users are mostly based on out-degrees of nodes in graph theory, while the semantic meaning between nodes is not considered. Moreover, there exists data sparsity problem for the proposed similarity-based approach, since the similarity of two users can be calculated if and only if they have rated common items.

In order to solve the problems, in this paper, we propose new approaches to identifying core users based on interest similarity and trust relationship between users, as well as three important attributes of individuals including one’s popularity degree, trust influence and resource. The identification of core users is realized through four steps: First, the trust degree and interest similarity between all user pairs are calculated and sorted from high to low. In particular, for interest similarity, we propose to cluster items based on K-means and then calculate the interest similarity of two users based on similar items. Therefore, we can get the nearest neighbor list of each user from the perspective of trust relationship and interest similarity. Second, the popularity degree, trust influence and resource of each individual user are calculated. Third, based on the calculated neighbor lists and values, we propose three strategies to select core users. Fourth, we validate the effect of the extracted core users by various methods in recommendation. The contributions of this paper include:

- The paper proposes two strategies to extract core users based on trust relationships and interest similarity between users. By considering the semantic meaning between items, we improve the calculation of user similarity. We also consider to improving users’ interaction data in social
Exploring Information Security Governance in Cloud Computing Organisation
www.igi-global.com/chapter/exploring-information-security-governance-in-cloud-computing-organisation/217927?camid=4v1a

Web Service Discovery and Composition for Virtual Enterprises
www.igi-global.com/article/web-service-discovery-composition-virtual/3093?camid=4v1a