A personalized authoritative user-based recommendation for social tagging

Jianliang Wei [a], Fei Meng [b,*], N. Arunkumar [c]

[a] *School of Management Engineering and E-commerce/Contemporary Business and Trade Research Center, Zhejiang Gongshang University, Hangzhou, Zhejiang, China*
[b] *Department of Public Foundation, Zhejiang Police College, Hangzhou, Zhejiang, 310053, China*
[c] *Department of Electronics and Instrumentation Engineering, SASTRA University, Thanjavur, India*

HIGHLIGHTS

- A new method for calculate user authority in social tagging system is put forward.
- Combining user authority and user vote, weights of tags added by users are adjusted, and a resource model algorithm is optimized.
- The results show that the algorithm generates personalized information recommendations based on authoritative users performs better than traditional models.

ARTICLE INFO

ABSTRACT

Personalized information recommendation based on social tagging is a hot issue in academia nowadays, but the concept of an authoritative user has not been emphasized in the existing literature. This paper first proposes a method to determine user authority in a social tagging system, in which the quality authority and quantity authority of users are calculated from a user co-occurrence network, which is derived from users' participation in the social tagging system. Degree centrality is employed for the user authority calculations, which are taken as weights for tag voting. On this basis, a resource model is constructed by summing up the tags from each user and their corresponding weights to represent each resource in the collection. User models are then obtained based on the resource models, and cosine similarity is used for making resource recommendations to users. An experiment was conducted on a dataset crawled from *Delicious.com*. The results show that the average GP relevance of the authoritative user based algorithm reaches 0.6115 much better than two benchmark algorithms.

1. Introduction

With the development of e-commerce, personalized information recommendation is attracting much attention from internet and e-commerce companies. To be employed as an operational technique, the most important issue in personalized recommendation is to have a method to precisely describe a user profile. With the rise of user generated content and collective intelligence, recommendation based social tagging is becoming a hot issue. Using tag relationships and resource relationships among users, a number of user profile construction techniques have been developed using existing data process methods, which have produced many achievements in personalized recommendation [1].

In the existing literature [2,3], however, users are hypothesized as equal, and thus their tags are treated equally without distinction to the value a more authoritative user may offer. In other words, the differences among influential authoritative users and non-authoritative users are dismissed. In fact, in both social networks and computer networks, the concept of node authority is universally recognized. Although more users have become both the generators and organizers of information resources, the cognitive differences among users determine the quality of their efforts. These differences are reflected in the level of quality of information generated and organized by users, which can turn into a measure to evaluate user authority. The higher the authority a user has, the more trustable the information generated by that user is, while information posted by a user with low authority may be able to be ignored [4]. Thus, it is useful to differentiate user authorities in social tagging systems. Information resources saved by authoritative users are more likely to be of higher quality and to be of value

* Corresponding author.
  E-mail address: mengfei@zjjcxy.cn (F. Meng).

to others with similar interests [5]. Additionally, tags contributed by authoritative users to describe and categorize resources could better reflect the features of the resources in a collection.

Therefore, a resource profile model based on authoritative users is a more reasonable model than traditional ones, and can be used to construct authoritative user profiles which can further be used to offer personalized authoritative resource recommendations to other users. Our authority based recommendation algorithm can be used in tagging websites to improve their personalized recommendation service, and also can help the websites to identify their key users and pay more attention on them in order to raise user satisfaction.

2. Related works

The authoritative user is an age-old notion to academia. Research on celebrity and opinion leaders were a part of early stage studies. With the emergence of ranking algorithms such as PageRank and Hyperlink-Induced Topic Search algorithm (HITS), discussions of user authority turned to the large scale quantitative analysis stage. Topics were extended to new domains, such as online forums, microblogs and shopping websites, and some researchers gave their attention to social tagging websites.

Erlandsson et al. believed that user influence was similar to the importance of web pages, and put forth the idea that methods such as PageRank can be utilized for finding experts, but their findings indicate that simple measures such as centrality measures are more suitable for finding experts in social networks [3]. The HITS algorithm also has been applied frequently in order to discover authorities in question-answering communities, and has been shown to be better than degree-based methods, especially to find experts in a question–answer community [6], where a good asker asked good questions to attract answers from good answerers, and good answerers answered good questions asked by good askers. Also in the domain of question–answer communities, Sahu et al. (2016) used a HITS-like algorithm to find reliable users by incorporating features such as questions starts, total thumbs up/down, best answers and so on [7]. By combining content and network methods together, Sahu et al.'s effort was different than the work of others. Moreover, Chang et al. [8] evaluated user reputation using co-occurrence features between questions and answers, wherein the similarity between question and answers was calculated based on the collection of n-grams from a title or question and the answers. Tang & Yang [9] proposed a new approach named UserRank to study user influence within an online healthcare community, in which the relationship among users' replies, the conversation content and the response immediacy were incorporated, to assign weights to the links of the constructed social network. In the online music and video domain, researchers proposed an algorithm to identify opinion leaders in Yahoo! Music DB, based on the date of users who evaluated media contents [10]. Han, Kim, and Cha [11] used a modified PageRank algorithm to calculate user reputation by considering user activities such as subscriptions, uploads and favorites, finding that user reputation was closely related to subscriptions and the number of uploads.

A number of research papers are focused on the topic of social tagging. Some researchers hold that the popularity of an information resource is related to the number of times it has been tagged. Therefore, from that perspective, resource ranking can be conducted by adding up the number of taggers, and using that as a measure to evaluate the quality of a resource [12]. Based on this approach, Yanbe et al. [13,14] put forth an algorithm named SBrank, which is based on the number of times a resource is tagged. They found that SBrank could make up for a shortcoming in PageRank which cannot reasonable account for new webpages. Thus, they have suggested that a combination two algorithms is

needed, in order to make use of all the advantages of both. In order to make the top ranked result relevant and with diversity, Qian et al. propose a social re-ranking algorithm considerate their visual information, semantic information and social clues, and inter-user re-ranking and intra-user re-ranking are adopted [15]. Based on the ability to influence the number of view and comments on social network, Yamasaki et al. (2017) present a demo using our FolkPopularityRank algorithm, which can score and recommend text tags based on their ability to influence the popularity-related numbers [16].

Using a HITS-based algorithm is another way to explore authoritative users. Combined with revised HITS, networks were constructed where users, tags and resources were taken as nodes and their relationships as edges, in order to find expert users and authoritative resources [17]. Yao et al. (2010) choose HITS algorithm to extract the authority of a user authority based on the user network [18]. In a study of the application of social tags in enterprise, an algorithm named ExperRank has been offered. According to the algorithm, the ranking of experts within enterprise is determined by their tag contributions [19]. More recently, Špiraneca & Ivanjkob [20] conducted a study with an expert and a novice group, and found that users with more knowledge and expertise could create folksonomies of higher quality, but their effort only focused on the differences in tagging behavior between the two groups. Mao et al. use HITS to refine the weights of tags in tag co-occurrence networks, and then transform the weights of tags into recommendation scores for items [21]. Overall, these related works are mainly on expert or authoritative users identification, some research even give method to calculation the resource authority [22], but seldom have focus on the influence of authoritative users on user profile construction and personalized information recommendation.

3. Resource model construction based on authoritative user

Before resource model and user model construct, we have two basic assumptions need to be declare. Assumption one is that interest of a user can be almost totally reflected by the resources he/she has saved, and assumption two is that authority of a user can be almost totally reflected by other user's voting on tags and resources.

3.1. User authority degree

In social tagging systems, a user-centered network is one several different kinds of networks which can be deduced from relations among user–tag–resource. According to the relationship of users defined by the edges in the network, there are two types of user networks. One is a tag-based user network, in which users are connected through the use of common tags. The other is a resource-based user network, in which users who saved the same resources related. In this paper, users are translated into nodes in the network, and network analysis is employed to identify authoritative users, as well as their authority degree.

HITS is one of the most classic algorithms for computing a online pages authority. Content authority and hub authority are the two key indices for the algorithm. Authority is evaluated by the quality of the content provided by web page. A good authority is said to be a page that was linked to by many different hubs. Hub quality is determined by the quality of linked pages; a good hub is a page that pointed to many other pages. According to the theory of degree centrality in social networks, if an actor has direct relevancy with many other actors, relatively speaking, then that actor is in a central position and has greater power. Users who have higher degree centrality in a network usually represent greater authorities.