Neural Networks 58 (2014) 98–110

Contents lists available at ScienceDirect

Neural Networks

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

2014 Special Issue

Exploring personalized searches using tag-based user profiles and resource profiles in folksonomy

Yi Cai [a,*], Qing Li [b], Haoran Xie [b], Huaqin Min [a]

[a] School of Software Engineering, South China University of Technology, Guangzhou, China

[b] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong Special Administrative Region

ARTICLE INFO

Article history:
Available online 4 June 2014

Keywords:
Social media
Personalized search
User profile
Folksonomy
Big data

ABSTRACT

With the increase in resource-sharing websites such as YouTube and Flickr, many shared resources have arisen on the Web. Personalized searches have become more important and challenging since users demand higher retrieval quality. To achieve this goal, personalized searches need to take users' personalized profiles and information needs into consideration. Collaborative tagging (also known as folksonomy) systems allow users to annotate resources with their own tags, which provides a simple but powerful way for organizing, retrieving and sharing different types of social resources. In this article, we examine the limitations of previous tag-based personalized searches. To handle these limitations, we propose a new method to model user profiles and resource profiles in collaborative tagging systems. We use a normalized term frequency to indicate the preference degree of a user on a tag. A novel search method using such profiles of users and resources is proposed to facilitate the desired personalization in resource searches. In our framework, instead of the keyword matching or similarity measurement used in previous works, the relevance measurement between a resource and a user query (termed the query relevance) is treated as a fuzzy satisfaction problem of a user's query requirements. We implement a prototype system called the Folksonomy-based Multimedia Retrieval System (FMRS). Experiments using the FMRS data set and the MovieLens data set show that our proposed method outperforms baseline methods.

1. Introduction

The advent of resource-sharing websites such as Flickr[1] and YouTube[2] allows users to share multimedia resources with each other and has resulted in huge amounts of resources on the Web. Resource searches have become more important and challenging since users demand higher retrieval quality. Current resource search methods primarily depend on the relevant match of the query and resource descriptions. Although different users input the same query terms, they may have different information needs corresponding to their own preferences. Thus, it is necessary to implement personalized resource retrieval based on users' profiles to ensure more pertinent and useful search results.

Currently, collaborative tagging systems have become more and more popular, and many social resource sites support tagging mechanisms. For example, bookmarks on Del.icio.us[3] may be tagged in terms of topics interesting to the user; on Flickr, users can upload and annotate their own photos; on Last.fm,[4] users can annotate their favorite songs. The resources and the tags posted by Web users on these systems are supposed to be highly dependent on their interests and the tags given by users provide rich information for building more accurate and specific user profiles (Al-Khalifa & Davis, 2006). Furthermore, the tags applied by different users to a resource are useful for describing the resource, which presents a collaborative viewpoint on resource description; such a description is considered to be more meaningful and acceptable from users' perspectives (Al-Khalifa & Davis, 2007; Cattuto, Benz, Hotho, & Stumme, 2008).

According to these characteristics of collaborative tagging systems, researchers consider that constructing user and resource profiles from collaborative tags is instrumental for personalized resource searches. Some research work such as (Kim & Ha, 2007) has

* Corresponding author. Tel.: +86 15511324868.
E-mail addresses: yicai.scut@gmail.com, ycai@scut.edu.cn (Y. Cai).

[1] http://www.flickr.com.

[2] http://www.youtube.com.

[3] http://delicious.com.

[4] http://www.last.fm.

http://dx.doi.org/10.1016/j.neunet.2014.05.017
0893-6080/

been conducted to construct user and resource profiles from tags in collaborative tagging systems; there also some personalized search methods proposed based on tags such as (Noll & Meinel, 2007; Vallet, Cantador, & Jose, 2010; Xu, Bao, Fei, Su, & Yu, 2008). However, limitations exist in current tag-based personalized search methods, which includes the following:

- In previous works, the weights of tags in a user profile or resource profile were based on term frequency (TF), TF inverse document frequency (TF-IDF[5]) or the BM25 method. By using absolute TF as tag weights in user profiles, the weights of active users who frequently annotate resources are increased; the weights of other users who rarely annotate resources are decreased. For TF-IDF, IDF is used to reflect how well a tag can distinguish a user (or resource) from others; it is not necessary for, nor useful for, indicating a user's preferences on tags or how a resource is relevant to tags. Since TF and IDF are variables of the BM25 function, the latter method suffers the same limitations by nature. Section 2.4 provides a more detailed discussion of this situation.
- In current works on tag-based personalized search, the match of a user and a resource is based on the similarity between the corresponding user's profile and the resource's profile. However, the resources that are relevant to a user's interests actually need to have as many of the user's favorite tags as possible (instead of simply being as similar as possible). The weight of a tag in a resource profile is typically the degree relevancy of the resource to the tag; the weight of a tag in a user profile is the favor degree of the user on the tag. If the weight of a tag in a user profile is more similar to the weight of the corresponding tag in a resource profile, a better match is not necessarily implied. For instance, the best match between a resource and a user is not returned when all of the matching tags' weights in both the user profile and the resource profile are small. This fact is true because a small weight of a tag x in a user profile means that the user is interested in x to a small extent instead of being not interested in x; a small weight of a tag y in a resource profile may mean that the resource is not as relevant to y. Therefore, although the similarity value between x and y is high, the tagged resource may not be relevant to the user's interests. In our viewpoint, the problem of how a resource satisfies a user's interest requirement is more of a fuzzy satisfaction problem, instead of a similarity measure problem. For example, Bob likes "spicy" taste to a degree 0.2, which does not mean that Bob dislikes spicy tastes. Furthermore, recipes that are not spicy are Bob's favorite. Additional discussions and illustrations of this subject are given in Section 2.5.

In this article, we discuss and elaborate on the limitations of current works related to tag-based personalized search. To handle these limitations, we propose a new tag-based model for constructing user and resource profiles. Based on the tag-based user profiles and resource profiles, we furthermore propose a novel personalized search framework in collaborative tagging systems. The new features of the proposed framework and the contributions of our work are as follows.

- For constructing user profiles, we use a normalized TF (NTF) to indicate the preference degree of a user on a tag. For a given tag x, the NTF is the possibility or proportion of a user using x, and we regard the NTF as being a more appropriate reflection of how much the user is interested in x.
- In our framework, instead of the keyword matching or similarity measurements used in previous works, the relevance measurement between a resource and a user query (termed *query relevance*) is treated as a fuzzy satisfaction problem of a user's query requirements. If a resource can satisfy as many of a user's query requirements as possible (in terms of the all query terms), then the resource will be more (content) relevant. Based on an observation of user query behavior, we present a query relevance function that takes into consideration the number of matching tags between the query and a resource profile.
- Similar to the query relevance measurement introduced above, the relevance measurement between a resource and a user's interest requirement (termed *user (interest) relevance*) is considered to be a fuzzy satisfaction problem. If a resource can satisfy as many of the user's interests (i.e., tags in user profiles) as possible, then the resource will be more relevant.
- We explore the relationships between query relevance and user interest relevance and treat them as explicit and implicit information needs, respectively. The final resources (relevance) ranking score measurement is also considered to be a fuzzy satisfaction problem of both query relevance and user interest relevance in our work.
- To illustrate and validate our proposed approach, we implement a prototype personalized recipe retrieval system called the Folksonomy-based Multimedia Retrieval System (FMRS). In FMRS, a recipe can be introduced to users through multimedia resources such as texts of recipe introduction, videos demonstrating the cooking procedure of a recipe, or photos of dishes corresponding to the recipes. Experimental studies are conducted on the FMRS to evaluate the effectiveness of our approach.
- In addition to FMRS, we also conduct experiments with real data from MovieLens.[6] The experimental results show that our method outperforms the state-of-the-art methods of personalized resource searches.

To the best of our knowledge, this work is the first study to model query relevance measurements, user (interest) relevance measurements and resource relevance ranking as fuzzy satisfaction problems.

The rest of this article is organized as follows. Section 2 introduces the background and related work, where we also reveal and discuss the limitations of existing studies. In Section 3, we model user and resource profiles in collaborative tagging systems. In Section 4, we propose a personalized search method using tag-based user and resource profiles. In Section 5, we conduct experiments to compare our proposed method with previous methods on the FMRS and MovieLens data sets. We discuss integrating sentiment filtering and concept-based models for user profiling in Section 6. Section 7 concludes the article and introduces potential future studies.

2. Background and related works

In this section, we first survey several existing studies on collaborative tagging and personalized searches. Next, we examine and discuss the limitations of these works in terms of user and resource profiling, as well as user (interest) relevance measurements.

[5] IDF refers to inverse user frequency for the user profile construction and inverse resource frequency for the resource profile construction.

[6] http://www.grouplens.org/node/73.