



Personalized recommendation of stories for commenting in forum-based social media[☆]



Ngo Xuan Bach^{a,c,*}, Nguyen Do Hai^b, Tu Minh Phuong^{a,c}

^a Department of Computer Science, Posts and Telecommunications Institute of Technology, Vietnam

^b FPT Software Research Lab, Vietnam

^c Machine Learning and Applications Lab, Posts and Telecommunications Institute of Technology, Vietnam

ARTICLE INFO

Article history:

Received 6 March 2015

Revised 21 January 2016

Accepted 4 March 2016

Available online 8 March 2016

Keywords:

Internet forums

Online news services

Content-based filtering

Collaborative filtering

Hybrid recommendation

Learning-to-rank

ABSTRACT

Web 2.0 platforms such as blogs, online news, social networks, and Internet forums allow users to write comments to express their interests and opinions about the content of news articles, videos, blogs or forum posts, etc. Users' comments contain additional information about the content of Web documents as well as provide important means for user interactions. In this paper, we present a study on the task of recommending, for a given user, a short list of suitable stories for commenting. We propose an efficient collaborative filtering method which exploits co-commenting patterns of users to generate recommendations. To further improve the accuracy, we also introduce a novel hybrid recommendation method that combines the proposed collaborative features and content based features in a learning-to-rank framework. We verify the effectiveness of the proposed methods on two datasets including samples of user comments from an online forum and a forum-based news service. Experimental results show that the proposed collaborative filtering method substantially outperforms traditional content-based approaches in terms of accuracy. Furthermore, the proposed hybrid approach leads to additional improvements over individual recommendation methods and achieves higher accuracy than a baseline hybrid approach. The results also demonstrate the stability of our methods in handling newly posted stories with a small number of comments.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

An important development of the World Wide Web, known as the Web 2.0, is characterized by the explosion of user-generated content such as blogs, wikis, social networking sites, video sharing, online news, and more traditional Internet forums. In such self-publication settings, a user can publish articles or submit posts to share with others on any topic. Other users can view and comment on the posts and these comments can subsequently be viewed and commented on. Apparent benefits from self-publication media are that they facilitate more user engagement and enable social interactions between users. The success of self-publication media has been witnessed by their increasing popularity and the fast growing volume of information generated. At the same time, the increasing volume of self-published content presents new challenges, one of them is the additional information overload for users. Due to the large amount of posts and comments, finding valuable

[☆] Part of this work was done when the first author was at School of Information Science, Japan Advanced Institute of Technology.

* Corresponding author. Tel.: +84 126 317 2610.

E-mail addresses: ngoxuanbachkcn@gmail.com, bachnx@ptit.edu.vn (N.X. Bach), haind2@soft.com.vn (N.D. Hai), phuongtm@ptit.edu.vn (T.M. Phuong).

posts to read and comment has become a difficult and time-consuming task. Recently, predicting which posts are suitable for a specific user to read and comment, followed by personalized recommendations, has attracted significant research interest [2,3,5,28,29,37,38,45]. This kind of recommendations can alleviate the problem of information overload in self-publication media.

In this paper, we consider the task of recommending stories for users of forum-based media. Specifically, the problem is to provide, for each user, a personalized ordered list of new posts or stories that the user will likely to comment on (in this paper, “post” and “story” will be used interchangeably). The most popular and successful technique for generating such kind recommendation is collaborative filtering (CF) [3,20,36], in which recommendations are decided based on behaviors of users with similar tastes. There are several considerations when applying collaborative filtering to forum post recommendation. First, in forum-based media, comments appear frequently and the appearance of each comment requires the collaborative filtering algorithm to re-run on the updated data. In such a dynamic environment, it is desired to develop efficient algorithms able to make small online updates as new comments arrive [29]. Second, in many forum-based media types such as online news services, it is critical to recommend stories to users as soon as they are posted, as the freshness of stories is important for user interest. This requires the collaborative filtering algorithm to be able to handle new stories with a small number of comments. In the collaborative filtering literature, this is known as the “new item” problem and present challenges to most collaborative filtering algorithms [6,14,33]. Third, in addition to comments, there are other types of information available in a forum-based environments such as the content of stories, user ratings on comments, social attributes of users, etc. It is desired to combine such different types of information and different types of recommendation algorithms to make more accurate predictions about user preferences [20,37,39]. A successful recommender system for forum-based media must be able to address these considerations.

We address these problems with two solutions, which are two main contributions of this paper. As the first contribution, we propose a collaborative filtering method that exploits the comment behavior of users and between users to recommend posts. The proposed method is simple to implement and does not require full re-runs as new comments arrive, which makes it suitable for dynamic forum-based environments such as forums (the first problem mentioned above). We compare the proposed method with popular content-based filtering (CBF) methods that represent stories by vector space models such as TF-IDF scores [8] and topic distributions generated by a topic modeling technique [9]. We perform experiments to verify the ability of the proposed CF algorithm to recommend fresh stories with a small number of comments (the second problem highlighted above), when used individually and in combination with CBF algorithms.

As the second contribution, we propose a novel hybrid recommendation framework to combine different types of CBF and the proposed CF features to achieve improved prediction accuracy (the third problem highlighted above). The idea of hybrid recommendation is not new [1]. The simplest hybrid method performs CBF and CF separately and then combines the results, e.g. by taking the sum of CBF and CF ratings. Other hybrid methods use content-based information to augment the user-item matrix before applying CF algorithms [32], or integrate all content-based and collaborative data to learn prediction models [7,35]. Recently, Sun et al. [39] propose to use learning-to-rank for hybrid recommendation. They transform user and item data to a new, content-comparable representation before applying RankSVM to the new transformed features. In this work, we also propose using learning-to-rank to combine CBF and CF. However, our method is different from the method of Sun et al. and other hybrid methods in that we apply learning-to-rank on CBF and CF scores, not on user-item or content-item matrices. By doing so, our method can be used to integrate multiple scores generated by different CBF and CF algorithms. Furthermore, the combination weights are learned automatically using learning-to-rank algorithms. We present experimental results to show that our proposed learning-to-rank hybrid method performs better than traditional CBF methods and the new hybrid approach performs much better than both CBF and CF methods. When used with our CF method, the proposed hybrid method provides accurate recommendations while remaining efficient. Because the main aim of this work is to provide efficient and accurate recommendations of fresh stories in forum-based media, the two proposed methods are the two components of a comprehensive solution for that application.

The rest of this paper is organized as follows. Section 2 describes research background and related work. Section 3 presents recommendation methods, including traditional content-based filtering methods, our proposed collaborative filtering method, and a novel hybrid method using learning-to-rank. Section 4 describes our two datasets, experimental setup, and experimental results. Finally, Section 5 concludes the paper and discusses future work.

2. Related work and background

2.1. Basis of recommender systems

In the broader context, our solution is a specific case of recommender systems which have been studied for years with many successful deployments (Melville and Sindhvani [33]). A recommender system provides recommendations in the form of a short list of movies, news, books, products etc. that the user is likely interested in [4,11,13]. Recommender systems do not only help users in choosing products or services but also help manufactures in personalizing their advertising efforts by automating the generation of recommendations based on data analysis. The methods of designing recommender systems and their applications to real-world problems have been an active area of research [26,33,44,45]. Recommender systems depend on the domain and characteristics of the data available, and recommendation methods differ in the way they exploit these data sources to generate recommendations.

Download English Version:

<https://daneshyari.com/en/article/392220>

Download Persian Version:

<https://daneshyari.com/article/392220>

[Daneshyari.com](https://daneshyari.com)