Physica A 389 (2010) 179–186

Contents lists available at ScienceDirect

Physica A

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

Personalized recommendation via integrated diffusion on user–item–tag tripartite graphs

Zi-Ke Zhang [a], Tao Zhou [a,b,*], Yi-Cheng Zhang [a]

[a] Department of Physics, University of Fribourg,Chemin du Musée 3, 1700 Fribourg, Switzerland

[b] Department of Modern Physics, University of Science and Technology of China, Hefei 230026, PR China

ARTICLE INFO

Article history:
Received 24 April 2009
Received in revised form 24 July 2009
Available online 12 September 2009

Keywords:
Complex networks
Personalized recommendation
Diffusion
Infophysics
Folksonomy

ABSTRACT

Personalized recommender systems are confronting great challenges of accuracy, diversification and novelty, especially when the data set is sparse and lacks accessorial information, such as user profiles, item attributes and explicit ratings. Collaborative tags contain rich information about personalized preferences and item contents, and are therefore potential to help in providing better recommendations. In this article, we propose a recommendation algorithm based on an integrated diffusion on user–item–tag tripartite graphs. We use three benchmark data sets, *Del.icio.us*, *MovieLens* and *BibSonomy*, to evaluate our algorithm. Experimental results demonstrate that the usage of tag information can significantly improve accuracy, diversification and novelty of recommendations.

1. Introduction

The last few years have witnessed an explosion of information that the exponential growth of the Internet [1] and World Wide Web [2] confronts us with an information overload: we face too much data and sources to be able to find out those most relevant for us. Indeed, we have to make choices from thousands of movies, millions of books, billions of web pages, and so on. Evaluating all these alternatives by ourselves is not feasible at all. As a consequence, a pressing problem is how to automatically find out the relevant items for us. Internet search engine [3], with the help of keyword-based queries, is an essential tool in getting what we want from the web. However, the search engine does not take into account personalization and returns the same results for people with far different habits. In addition, not all needs or tastes can be easily presented by keywords. Comparatively, *recommender system* [4], which adopts knowledge discovery techniques to provide personalized recommendations, is now considered to be the most promising way to efficiently filter out the overload of information. Thus far, recommender systems have successfully found applications in e-commerce [5], such as book recommendations in *Amazon.com* [6], movie recommendations in *Netflix.com* [7], video recommendations in *TiVo.com* [8], and so on.

A recommender system is able to automatically provide personalized recommendations based on the historical record of users' activities. These activities are usually represented by the connections in a user–item bipartite graph [9,10]. Fig. 1 illustrates such a graph consisted of five users and four books, where users can give ratings to those books. So far, collaborative filtering (CF) is the most successful technique in the design of recommender systems [11], where a user will be recommended with items that people with similar tastes and preferences liked in the past. Despite its success, the performance of CF is strongly limited by the sparsity of data resulted from: (i) the huge number of items that is far beyond user's ability to evaluate even a small fraction of them; (ii) users do not incentively wish to rate the purchased/viewed items [12]. As shown in Fig. 1, besides the fundamental user–item relations, some accessorial information can be exploited

* Corresponding author at: Department of Physics, University of Fribourg,Chemin du Musée 3, 1700 Fribourg, Switzerland.
E-mail address: zhutou@ustc.edu (T. Zhou).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2009.08.036

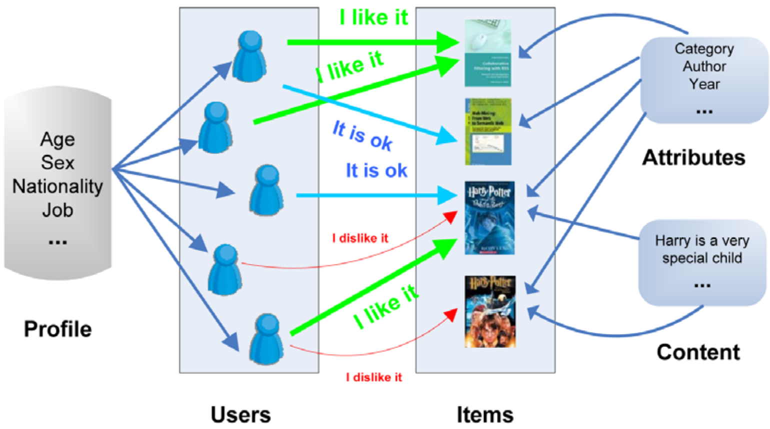

Fig. 1. (Color online) Illustration of a recommender system consisted of five users and four books. The basic information contained by every recommender system is the relations between users and items that can be represented by a bipartite graph. This illustration also exhibits some additional information that are frequently exploited in the design of recommendation algorithms, including user profiles, item attributes and item content.

to improve the algorithmic accuracy [13]. User profiles, usually including age, sex, nationality, job, etc., can be treated as prior known information to filter out possibly irrelevant recommendations [14], however, the applications are mostly forbidden or strongly restricted to respect personal privacy. Attribute-aware method [15] takes into account item attributes, which are defined by domain experts. Yet, it is limited to the attribute vocabulary, and, on the other hand, attributes describe global properties of items which are essentially not helpful to generate personalized recommendations. In addition, content-based algorithms can provide very accurate recommendations [16], however, they are only effective if the items contain rich content of information that can be automatically extracted out. For example, these methods are suitable for recommending books and articles, but not for videos or pictures.

Collaborative tagging systems (CTSes), allowing users to freely assign tags to their collections, provide promising possibility to better address the above issues. CTSes require no specific skills for user to participate, thus can overcome the limitation of vocabulary domains and size, widen the semantic relations among items and eventually facilitate the emergence of *folksonomy* [17].

The CTSes have recently attracted much attention from scientific communities. Firstly, many efforts have been addressed in understanding the structure and evolution of CTSes [18,19], as well as the usage patterns in folksonomies [20]. For example, Cattuto et al. [18] investigated the evolution of an open-ended system *Del.icio.us*, while Zhang et al. [19] empirically studied the semiotic structure of a number of seriously edited scientific journals. An interesting result is that both the open-ended and seriously edited systems share many common features [18,19].

Secondly, some studies concern the applications of CTSes. A considerable number of algorithms are designed to recommend tags to users, which may be helpful for better organizing, discovering and retrieving items [17,21,22]. The current work focuses on a relevant yet different application of CTSes, that is, to provide personalized item recommendations with the help of tag information. Schenkel et al. [23] proposed an incremental threshold algorithm taking into account both the social ties among users and semantic relatedness of different tags, which performs remarkably better than the algorithm without tag expansion. Nakamoto et al. [24] created a tag-based contextual CF model, where the tag information is treated as the users' profiles. Tso-Sutter et al. [25] proposed a generic method that allows tags to be incorporated to the standard CF, via reducing the ternary correlations to three binary correlations and then applying a fusion method to re-associate these correlations.

In this article, instead of the well-known CF, we will apply a diffusion-based recommendation algorithm on user–item–tag tripartite graphs. Huang et al. [9] proposed a *spreading activation-based CF*, which is essentially an iterative diffusion process. This algorithm can provide relatively accurate recommendations for very sparse systems that are hard to be managed by the standard CF. Zhou et al. [10] proposed an extremely fast algorithm considering only a two-step diffusion in user–item bipartite networks, which can still give slightly more accurate recommendations than the standard CF. Further more, some refinement taking into account the initial difference of activations for different items can simultaneously enhance the recommendation accuracy and diversification [26]. Zhang et al. [27] proposed an iterative opinion diffusion algorithm to predict ratings in *Netflix.com*. All the above diffusion-based algorithms obey the *conservation law*, in contrast to which Zhang et al. [28] proposed a non-conservation diffusion-based recommendation algorithm mimicking the heat conduction in networks, which is very efficient to dig out the unpopular yet relevant items. Song et al. [29] proposed a so-called *DiffusionRank* algorithm, where the prediction score is given by the the likelihood that information can propagate from a given user to a given item within a certain time period. Liu et al. [30] proposed a *diffusion-based top-k collaborative filtering*, which performs better than the pure top-*k* collaborative filtering and pure diffusion-based algorithm.

This article makes the following three contributions: (i) We propose an integrated diffusion-based algorithm making use of both the user–item relations and the collaborative tagging information. (ii) We test our algorithm in three real