Web Semantics: Science, Services and Agents on the World Wide Web 34 (2015) 13–26

Contents lists available at ScienceDirect

Web Semantics: Science, Services and Agents on the World Wide Web

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

Neighborhood-user profiling based on perception relationship in the micro-blog scenario

Jianxing Zheng [a], Bofeng Zhang [a,*], Xiaodong Yue [a], Guobing Zou [a], Jianhua Ma [b], Keyuan Jiang [c]

[a] School of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China

[b] Faculty of Computer and Information Sciences, Hosei University, Japan

[c] Department of Computer Information Technology & Graphics, Purdue University Calumet, IN, USA

ARTICLE INFO

Article history:
Received 7 September 2014
Received in revised form
25 June 2015
Accepted 26 June 2015
Available online 3 July 2015

Keywords:
Micro-blog
Ontology
Resource perception relationship
Follow perception relationship
Neighborhood-user profile

ABSTRACT

In the micro-blog scenario, personal user profiling relying on content is limited for recommending desired diverse subjects due to its shortcomings of short text, often leading to a poor recall. Currently, many methods only utilized the personal knowledge from each individual user to represent user profile without considering the neighborhood information. However, resource information related to neighboring friends play an important role in improving the performance of recommender systems. In this paper, we present the personalized expanded user profiling for micro-blog subject recommendation via ontology semantics structure. Next, taking into account diffusion ability of followee friends, we discuss resource perception relationship (RPR) and follow perception relationship (FPR). Finally, we discuss how, by adjusting the importance of RPR and FPR, the neighborhood is selected to construct neighborhood-user profile, which can mine new relevant subjects for target user. Our experimental results demonstrate the effectiveness of our neighborhood-user profiling in comparison to the existing collaborative filtering and personal user profile recommendation approaches on Sina micro-blog platform datasets.

1. Introduction

Nowadays, as a newly emerged communication tool and public medium on the Internet, micro-blog spreads popular hot topics from one user to millions of individuals just in a few minutes, which allows the user to receive desired information anytime and anywhere. Meanwhile, searching for personalized interests and feelings posted to the multi-source information platforms, such as micro-blog systems like Twitter, social network sites including Facebook and LinkedIn, and personal homepages and blogs as well as many others [1,2], is an interesting yet challenging task. Especially, in the micro-blog system platforms, people repost a lot of short messages about their daily activities and feelings so as to maintain latest interests or friendship.

Many researchers have successfully tested feasibility of applications in many areas including interesting topics [3] and micro-blog environments [4]. However, personal User Profile (UP) is a customized model of interest representing and reasoning for a user, which is implicitly contained and generated from one's behaviors, browsing contents, or feedbacks [5–7]. That is, how to fulfill personalized activities and information requirements with one's micro-blog user profile is an important yet challenging issue. Very little research, however, has been done on this issue.

In the micro-blog scenario, each micro-blog is short and lacks sufficient information for user profile construction. As is expected, a user profile is not only generated from individual short messages, but also profits from existing interactions of friends [8–10]. With more than 215 million users and more than 175 million postings per day in 2012, Twitter is one of the most prominent micro-blog services on the web [11]. In particular, most of users are used to forwarding tweets for communication, instead of directly posting. Hence, followees make an important role in the propagation and spread of personalized interests. Traditional user profiles capture personal interests over one's own knowledge [12–15], which are not holistic for discovering diverse information. In this case, items and products in user profiles could not reflect currently concerned subjects and socially propagative topics thoroughly.

In many scenarios, traditional collaborative filtering (CF) strategy provides users a lot of valuable information on the basis of mutual understanding and knowing. In social communities such as

* Corresponding author.
E-mail addresses: jxzheng185@gmail.com (J. Zheng), bfzhang@shu.edu.cn (B. Zhang), yswantfly@shu.edu.cn (X. Yue), guobingzou@gmail.com (G. Zou), jianhua@hosei.ac.jp (J. Ma), jiang@purduecal.edu (K. Jiang).

http://dx.doi.org/10.1016/j.websem.2015.06.005
1570-8268/

Facebook, LinkedIn and Twitter, the solution of CF is challenging. First, sparse data derived from short text is insufficient to capture enough similar users to recommend desired items, which hurt both the precision and recall of recommender systems. Second, pluralistic society makes people generate diverse interests. Not only are users restricted to daily monotonous interesting item, but they may be interested in diverse items posted by their friends [16]. A vast amount of diverse data enables similarity between users is small, which also leads the capture of similar users is hard. Meanwhile, a small amount of similar users is difficult to discover the items of the high correlation, which cannot effectively be applied into CF strategy [9]. Actually, when we follow what the followees have written, we can reflect our interests in a tracking way; and when we glimpse other followees or communities, we can realize where the interests come from. Follow friends' knowledge is a kind of effective collective wisdom, which could extend personal interest to other latent but relevant subjects. Additionally, follow relation is a new back-to-back linkage, which can supply the target user diverse interests from collaborative users [9,17]. It is reasonable that these follow friends contain a group of intimate interest users, named as neighborhood. Therefore, neighborhood with sufficient knowledge could help an individual user build the Neighborhood User Profile (NUP), addressing the problem of information shortage in representing personal interest.

In this work, we, using Sina micro-blog data source, constructed novel neighborhood user profile based on the collective knowledge. First, taking into account roles of followee friends in the interest propagation of the target user, we investigated the follow perception relationship and resource perception relationship. Furthermore, by adjusting the importance of two kinds of relationships, we discovered the neighborhood of a user. Lastly, the NUP relying on neighbor interests is proposed. In addition, the proposed NUP is evaluated by comparing against the existing personal UP and CF recommendation methods through experiments on large amounts of data from the Sina micro-blog platform.

Our experimental results show that the proposed NUP approach outperforms other methods in both precision and recall but with a relatively higher time complexity. By analyzing the expanded interests by NUP, we have observed that the recommendations based on NUP can accelerate the diffusion of the user interest, especially some semantically related interest between friends. We introduce the idea of neighborhood to solve the problem of acquiring behavior interest of social users. In particular, with the consideration of both the roles of followees' friends and resource perception relationship equally, the selected neighborhood could expand semantic interest efficiently. When the neighborhood only includes oneself, the NUP becomes a conventional individual user profile. However, the zooming size of neighborhood is an important issue for interest supplement related to social networks and social Webs, which needs to leverage the adaptive diversification fusion algorithm for zooming-in and zooming-out of the neighborhood.

The remaining part of this paper is organized as follows. Section 2 briefly discusses the works related to user profile. In Section 3, an overview of our recommendation framework based on neighborhood user profile is presented. In Section 4, we introduce personal interest acquiring method. Concepts of neighborhood and detailed descriptions of interest extending in neighborhood user profile are presented in Section 5. In Section 6, we demonstrate the application of our system as well as our experiment results along with discussions on strength and limitations. Finally, we conclude in Section 7 directions of our future work.

2. A brief review of UP works

In the scenario of user profile construction, how to convert the raw micro-blog documents into user's interesting subjects is usually challenging. To exactly recommend appropriate products to the user, many researchers have published their works in discovering demonstrated ways to build user profiles [18–20,12–14,10,21,22]. In this section, we will briefly review some popular works related to user profile.

2.1. Content-based UP

Content-based user profile focuses on document content analysis to classify the categorization of browsing historical records for deriving the hot interest and meaningful subjects for a target user. Many researches use individual information from the current search session or personal information to construct UP [23]. Via the key word set from one's comments and article, Meguebli et al. [13] built a user profile and article profile for each user, and computed the similarity between the article profile and user profile to sort the recommendation article list. By taking full advantage of informal and unstructured labeled data in Tweets, Lim [8] proposed a LDA-based Twitter Opinion Topic Model (TOTM) to aggregate or summarize opinions of a product, which can discover target specific opinion words and improve opinion prediction. Instead of employing a human-generated ontology, Harvey [14] proposed a novel latent topic models to describe both the clicked URLs and the interests of users from click log data. Considering the characteristics of short text messages expressing user's opinions and interests, Esparza [24] described users and products from the terms used in abbreviated and highly personalized commentary and studied how Twitter-like short-form messages could be leveraged as a source of indexing and retrieval information. Lin [15] utilized a semi-supervised variant of LDA that accounts for both text and metadata to characterize version features into a set of latent topics for exploring UP modeling in the app domain. By adopting features including user interest match, content-dependent user relationship and user influence, Wang [9] proposed a machine learned ranking function to find a new interest group of users for newly twitter information diffusion. Although these works have obtained the remarkable achievements, the existing user profiling methods are not fit for micro-blog users. In the micro-blog scenario, traditional UP methods capture too sparse interests to permit robust personalization and only recommend limited subjects for user's diverse needs. In this work we focus on the roles of social relationships between users for user profiling as they provide a richer source of information about the user's sufficient interests and preferences.

2.2. Semantics-based UP

Semantics-based user profile focuses on researching the semantic linkages of blogs or articles to discover a user's interest. Based on clustering the keywords from the electronic academic publications in online service, Tang [10] focuses on extending scientific subject ontology to refine user interest profiling. Varga [25] introduced a new semantic graph, called category metagraph, to extract a more fine grained categorization of concepts to provide a set of novel semantic features from short text messages. As the cosine similarity and TF–IDF weighting scheme for terms occurring in news messages are used in most user profiles, Hogenboom [26] extended semantics based weighting techniques, Bing–SF–IDF+, by considering the synset semantic relationships and by employing named entity similarities using Bing page counts, to perform better of F1 than TF–IDF and SF–IDF methods. For the hierarchical semantic structure embedded in the query and the document, Huang [27] used a deep neural network (DNN) to rank a set of documents for a given query and proposed a series of Deep Structured Semantic Models (DSSM) for Web search. Tao [21] constructed user profiles based on the ontology with world knowledge and user local knowledge, and utilized semantic specificity of concept in each ontology layer to mine user's