Neurocomputing 129 (2014) 409–420

Contents lists available at ScienceDirect

Neurocomputing

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

Mining user-contributed photos for personalized product recommendation ☆

He Feng, Xueming Qian *

Department of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an 710049, China

ARTICLE INFO

Article history:
Received 13 February 2013
Received in revised form
4 September 2013
Accepted 7 September 2013
Communicated by Jinhui Tang
Available online 7 November 2013

Keywords:
Products recommender
Personalized recommendation
Hierarchical topic space
Social media

ABSTRACT

With the advent and popularity of social media, users are willing to share their experiences by photos, reviews, blogs, and so on. The social media contents shared by these users reveal potential shopping needs. Product recommender is not limited to just e-commerce sites, it can also be expanded to social media sites. In this paper, we propose a novel hierarchical user interest mining (*Huim*) approach for personalized products recommendation. The input of our approach consists of user-contributed photos and user generated content (UGC), which include user-annotated photo tags and the comments from others in a social site. The proposed approach consists of four steps. First, we make full use of the visual information and UGC of its photos to mine user's interest. Second, we represent user interest by a topic distribution vector, and apply our proposed *Huim* to enhance interest-related topics. Third, we also represent each product by a topic distribution vector. Then, we measure the relevance of user and product in the topic space and determine the rank of each product for the user. We conduct a series of experiments on Flickr users and the products from Bing Shopping. Experimental results show the effectiveness of the proposed approach.

1. Introduction

With the rapid development of e-commerce, products recommender system has been exploited to suggest attractive and useful products' information to facilitate user's decision-making process. The intelligent of products recommendation can help users to deal with information overload and provide them personalized services [39]. Product recommendation is popular in e-commerce sites. Some e-commerce sites such as Amazon and Bingshopping recommend products to users based on previous buys as well as what others have been bought when they bought the same product. They keep tracks of users spending and analyze their interests by collaborative filtering [3]. However, in collaborative filtering based products recommendation approaches, only the relevance of users is considered. Thus, they are not personalized to user's interest.

With the booming of social networks, more and more people are will to share their personal affairs, new things and their favorite photos with their friends. For example, Facebook has about one billion users. Flickr is photo sharing website, it also have a very large amount of users. The total number of photos shared by users in Flickr had reached 6 billion by August 2011. There are about 3 million photos uploaded by users each day. The user contributed photos and user generated content can reveal the user's interests very well [44–46,50,51]. Thus the social media websites are the ideal platforms to facilitate the personalized products recommendation.

To improve user experience by making ads relevant to the webpage content, Broder et al. proposed a system for contextual ad matching based on a combination of semantic and syntactic features [40]. They used the semantic phrase to classify the webpage and the ads into taxonomy, and then ranked ads by the proximity of the ads and webpage categories. Although they classified both ads and page content within a large taxonomy, they ignored that some ads are relevant to several topics. For instance, the tag *canon* is relevant to digital camera and also relevant to the bags of camera. Thus, it is better to represent user interests by topic distribution vector rather than taxonomy.

Taking above mentions into consideration, we propose a novel hierarchical user interest mining (*Huim*) method to explore user's potential shopping needs based on user-contributed photos in her/his social media sites. We recommend personalized products according to the mined user interests. There are three main problems needed to be solved: (1) the gap between appearances of user-contributed photos and their textual descriptions (i.e. UGC). For example, when a user uploaded some images of her new iphone, she may label images by the words "*the amazing*

☆This work was supported in part by National Natural Science Foundation of China (NSFC) Project No. 60903121 and No. 61173109, and Microsoft Research Asia.
* Corresponding author. Tel.: +86 29 82668772.
E-mail addresses: fenghe7658@stu.xjtu.edu.cn (H. Feng), qianxm@mail.xjtu.edu.cn (X. Qian).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.09.018

apple". In this circumstance, the tag of *apple* has two meanings: fruit and electronic product. So we need to mine more content related information from both the visual information and the UGC of user contributed photos. (2) The noise and ambiguous tags in UGC have negative effect to mine user's interests. The textual descriptions contain informal expressions with noise tags (such as the preposition and other non-topic tags) and some ambiguous tags which generated by users. Thus, to recommend personalized products, we need to suppress the noise and ambiguous tags, and to enhanced user interest topics for the user contributed photos. (3) How to measure the relevance of user and product. If a user shared some photos of basketball game, then it is reasonable to suggest some products about basketball like knee pad, and it is also acceptable to recommend basketball video games of 360 Xbox. Among the above three problems measuring the relevance is the core problem in product recommendation.

To fulfill personalized product recommendation, we propose a hierarchical topic vector representation approach to represent user interest and product. Our approach is carried out as follows: (1) tag enrichment for the user contributed photos, (2) introduce a public topic space and map user interest and product descriptions to it and get their topic distribution vectors, (3) enhance user interested topics by a hierarchical approach is proposed to suppress noise and ambiguous textual descriptions, (4) measure the relevance of user and product in the hierarchical public topic space and rank the products for the user.

The main contributions of this paper are summarized as follows: (1) propose a personalized product recommendation system which mining users' interests from their contributed photos; (2) propose an effective user and product relevant measurement approach by introducing a hierarchical public topic space; (3) propose an effective hierarchical user interest representation approach which is robust to suppress noise and ambiguous textual description and enhance user interested topics.

Compared to our preliminary version [50], several improvements are made: (1) the detailed steps of the proposed hierarchical user interest mining approach are provided; (2) we extend the approach from brand recommendation to more general product recommendation, and (3) more experimental results and discussions are provided.

The remainder of this paper is organized as follows. In Section 2, we present the related works on products recommendation, user interest mining and multimedia advertising. In Section 3, our personalized products recommendation based on hierachical user interest mining approach is introduced in detail. Experiments and discussions are given in Section 4 and conclusions are drawn in Section 5.

2. Related work

In this section, we briefly review the related works on products recommendation, user interest mining from social media, and multimedia advertising.

2.1. Products recommendation

Products recommendation has been emphasized with the advent of E-commerce [4,7,12]. The previous works [28–31] has been mainly focused on collaborative filtering based recommender system since it was first implemented by Goldberg et al. [27]. Ricci et al. designed a hybrid collaborative/content based recommendation approach [28]. This approach has the advantage to overcome the shortcomings of content- and collaborative-filtering based recommendation approaches. And then, various clustering and data mining technologies were proposed in product and services recommendation [49–51]. Qian et al. propose a user preferred vocabulary mining approach to recommend user preferred vocabularies for the user newly shared photos [51]. Wesley et al. developed a mining association rules procedure from a datasets to support on-line products recommendation [32]. They proposed clustering module and extracting module to mine user's hidden habits from the datasets. The clustering module is based on a self-organized map neural-network to carry out data grouping. The extracting module uses rough set theory to determine rules for the clusters and their relationships. They used analytic hierarchy process to determine the related weights of customer and lifetime value, based on which products are recommended to group of customers using an associated rule mining approach. Liu and Shih suggested a novel recommendation methodology that combined group decision-making and data mining to address the lifetime value of a customer to a firm [33]. Cao and Li proposed fuzzy-based system to retrieve optimal products based on the customer's interactions with their system [34]. Zheng et al. explore location information for personalized travel recommendation and friends recommendation [25,26].

2.2. User interest mining

The success and popularity of social media sites have generated many interesting and challenging problems to the research community. Mining users' interests from their contributed information have attracted much attention [20,35–38,49–51]. User interest mining approaches can be classified into one of the following two categories: text based, and visual content and text combined approaches. In text-based approach, Li et al. discovered social interest only by user-generated tags [35]. Their research results on a large scale real-world traces shown that user generated tags are highly consistent with the web content. User interests were represented by the patterns of several high co-occurrence and frequently appeared tags. Feng and Qian represent user interest by the nature category constrained topic vector distribution. Services/products/items and user interests can be represented by the topic vector distribution [49,50]. Banerjee et al. used text mining technology to extract users' interests from their micro-blogs [36]. They classified keywords into two types: content-indicative and usage-indicative. Then they discovered the usage statistics of the co-occurrence of two types of keywords to represent users' interests. Choudhury et al. [37] and Wang et al. [20] took both the visual content of user shared images and the context of social media into consideration to mine user interest.

Usually, user generated textual descriptions including tags are noisy. So, many tag enhancement/filtering approaches are proposed to improve tag quality [44–48]. For example, Tang et al. propose a sparse based semi-supervised learning approach to infer semantic concepts from social community user-contributed photos and noisy tags [44–46]. Qian et al. proposed an effective tag filtering approach by using the similar compatible principle [47]. Li et al. propose a latent community classification and multi-kernel learning based image annotation approach to solve the tag ambiguous problem [48].

2.3. Multimedia advertising

Products recommendation has a slightly different from advertising. It is a special case of advertising. Advertising pays little attention on advertiser and bid. However, the core problems of them are the same: the relevance of ads or products and user. According to the formats of advertising, there are four types of approaches: text-based advertising [6,10,41,42], image based advertising [5,13,14,19,20,22], game based advertising [11], and video based advertising [2,9,15–18,23].