Signal Processing 124 (2016) 45–53

Contents lists available at ScienceDirect

Signal Processing

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

A picture tells a thousand words—About you! User interest profiling from user generated visual content

Quanzeng You [a], Sumit Bhatia [b], Jiebo Luo [a,*]

[a] Department of Computer Science, University of Rochester, Rochester NY, United States
[b] IBM Almaden Research Center, San Jose CA, United States

ARTICLE INFO

Article history:
Received 27 April 2015
Received in revised form
15 October 2015
Accepted 31 October 2015
Available online 10 November 2015

Keywords:
User profiling
Social multimedia
Multimedia analysis

ABSTRACT

Inference of online social network users' attributes and interests has been an active research topic. Accurate identification of users' attributes and interests is crucial for improving the performance of personalization and recommender systems. Most of the existing works have focused on textual content generated by the users and have successfully used it for predicting users' interests and other identifying attributes. However, little attention has been paid to user generated visual content (images) that is becoming increasingly popular and pervasive in recent times. We posit that images posted by users on online social networks are a reflection of topics they are interested in and propose an approach to infer user attributes from images posted by them. We analyze the content of individual images and then aggregate the image-level knowledge to infer user-level interest distribution. We employ image-level similarity to propagate the label information between images, as well as utilize the image category information derived from the user created organization structure to further propagate the category-level knowledge for all images. A large scale social network dataset of 1.5+ million images created from Pinterest is used for evaluation and the experimental results demonstrate the effectiveness of our proposed approach.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Online Social Networks (OSNs) such as Facebook, Twitter, Pinterest, and Instagram have become a part and parcel of modern lifestyle. A study by Pew Research centre[1] reveals that three out of every four adult internet users use at least one social networking site. Such large scale adoption of OSNs and active participation of users have led to research efforts studying relationship between users' digital behavior and their demographic attributes (such as age, interests, and preferences) that are of particular interest to social science, psychology, and marketing. A large scale study of about 58,000 Facebook users performed by Kosinski et al. [1] reveals that digital records of human activity can be used to accurately predict a range of personal attributes such as age, gender, sexual orientation, and political orientation. Likewise, there have been numerous works that study variations in language used in social media with age, gender, personality, etc. [2–4]. While most of the popular OSNs studied in literature are mostly text based, some of them (e.g., Facebook, Twitter) also allow people to post images and videos. Recently, OSNs such as Instagram and Pinterest that are majorly image based have gained popularity with almost 20 billion photos already been shared on Instagram and an average of 60 million photos being shared daily.[2]

* Corresponding author.
E-mail addresses: qyou@cs.rochester.edu (Q. You), sumit.bhatia@us.ibm.com (S. Bhatia), jluo@cs.rochester.edu (J. Luo).

[1] http://www.pewinternet.org/2013/12/30/social-media-update-2013/

[2] http://instagram.com/press/

http://dx.doi.org/10.1016/j.sigpro.2015.10.032
0165-1684/© 2015 Elsevier B.V. All rights reserved.

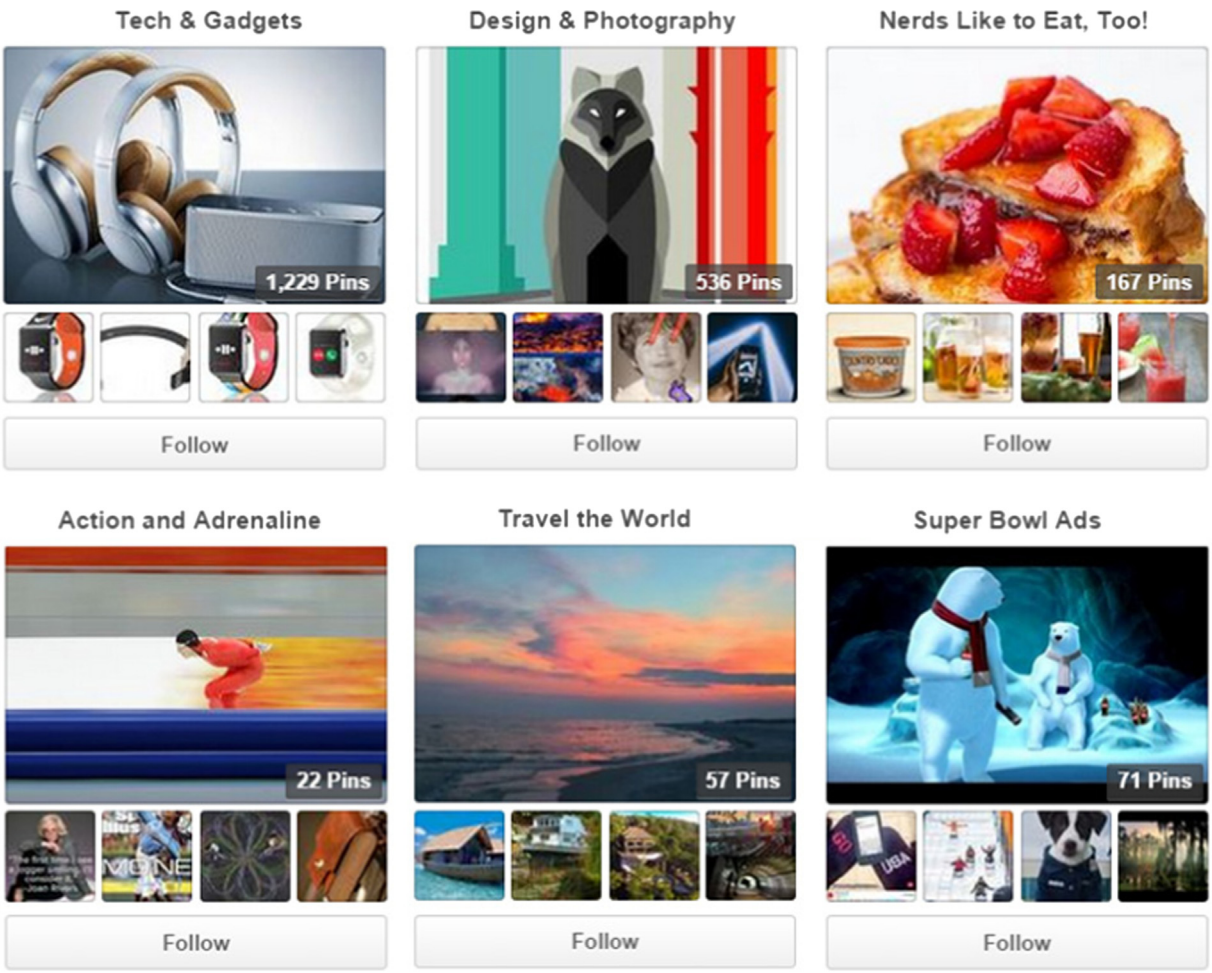

Fig. 1. Example pinboards from one typical Pinterest user.

The most appealing aspect of image based OSNs is that visual content is *universal* in nature and thus, not restricted by the *barriers of language*. Users from different cultural backgrounds, nationalizes, and speaking different languages can easily use the same *visual language* to express their feelings. Hence, analyzing the content of user posted images is an appealing idea with diverse applications. Some recent research efforts also provide support for the hypothesis that images posted by users on OSNs may prove to be useful for learning various personal and social attributes of users. Lovato et al. [5] proposed a method to learn users' latent visual preferences by extracting aesthetics and visual features from images favorited by users on Flickr. The learned models can be used to predict images likely to be favored by the user on Flickr with reasonable accuracy. Cristani et al. [6] infer personalities of users by extracting visual patterns and features from images marked as favorites by users on Flickr. Can et al. [7] utilize the visual cues of tweeted images in addition to textual and structure-based features to predict the retweet count of the posted image. Motivated by these works, we investigate *if the images posted by users on online social networks can be used to predict their fine-grained interests or preferences about different topics*. To understand this better, Fig. 1 shows several randomly selected pinboards (collection of images, as they are called in Pinterest) for a typical Pinterest user as an example. We observe that different *pins* (a pin corresponds to an image in Pinterest) and pinboards are indicative of the user's interests in different topics such as sports, art and food. We posit that the visual content of the images posted by a user in an OSN is a reflection of her interests and preferences. Therefore, an analysis of such posted images can be used to create an interest profile of the user by analyzing the content of individual images posted by the user and then aggregating the image-level knowledge to infer user-level preference distribution at a fine-grained level.

1.1. Problem formulation and overview of proposed approach

Problem Statement: Given a set $\mathcal{I}$ of images posted by the user u on an OSN, and a set $\mathcal{C}$ of interest categories, output a probability distribution over categories in $\mathcal{C}$ as the interest distribution for the user. In order to solve this problem, we first need to understand the relationships between different categories (topics) and underlying

Download English Version:

https://daneshyari.com/en/article/562229

Download Persian Version:

https://daneshyari.com/article/562229

Daneshyari.com