Neurocomputing 000 (2017) 1–8

Contents lists available at ScienceDirect

Neurocomputing

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

Hashtag recommendation for multimodal microblog posts

Yeyun Gong, Qi Zhang*, Xuanjing Huang

School of Computer Science, Fudan University, Shanghai Key Laboratory of Data Science, 825 Zhangheng Road, Shanghai, PR China

ARTICLE INFO

Article history:
Received 17 April 2015
Revised 8 March 2017
Accepted 26 June 2017
Available online xxx

Communicated by Dr. Nianyin Zeng

Keywords:
Hashtag recommendation
Topical model
Social media

ABSTRACT

In microblogging services, authors can use hashtags to mark keywords or topics in microblogs. Many live social media applications (e.g., microblog retrieval, classification) can gain great benefit from these manually labeled tags. However, only a small partition of microblogs contains hashtags. Moreover, many microblog posts contain not only textual content but also images. These visual resources also provide valuable information that is not included in the textual content. To recommend hashtags for these multimodal microblogs, in this work, we propose a novel generative method incorporating textual and visual information to solve the task. Experimental results on the data collected from real world microblogging services demonstrate that the proposed method outperforms state-of-the-art methods using either textual or visual information. The relative improvement of the proposed method over the textual only method is more than 17.1% in F1-score.

1. Introduction

With the rapid growth in social media, about 72% of adult internet users are also members of at least one social networking site.[1] Over the past few years, microblogging has become one of the most popular services among all the social media. Hence, microblogs have also been widely used as sources for public opinion analyzation [1], prediction [2], reputation management [3], and many other applications [4–6]. In addition to the limited number of characters in the content, microblogs also contain a form of metadata tag (hashtag), which is a string of characters preceded by the symbol (#). Hashtags are used to mark the keywords or topics of a microblog. They can occur anywhere in a microblog at the beginning, middle, or end. Hashtags have been proven to be useful for many applications, including microblog retrieval [7], query expansion [8], sentiment analysis [9]. However, only a few microblogs include hashtags labeled by their users. Hence, automatically recommending hashtags has become an important research topic and has received considerable attention in recent years.

Although, some research has been done on this task, most of the studies have focused only on the use of textual information with discriminative models [10,11], or generative models [12–14]. However, from a dataset that contains 282.2 million microblogs crawled from Sina Weibo,[2] we observe that more than 34.8% of microblog posts contain not only text but also images. A large number of microblogs with images do not contain relevant textual annotations for these images. Hence, it is not easy to correctly recommend hashtags through approaches that use only textual information. Fig. 1 illustrates a multimodal microblog with the hash tag *#SuperBowl*. We can see that with only textual information, the correct hashtag is hardly to be identified.

The task of tag suggestion/recommendation for images is also related to this work and has been studied from various aspects [12,15,16]. Most of these works pay much attention to the tags annotated by users through social media services such as Flickr, Zooomr, and so on. The tags annotated in these services are like labels that are added to a photo to make it easier to find later. Previous works mainly focused on recommending tags that are good descriptors of the photo itself, whereas hashtags usually refer to more abstract concepts. The common content-based tags in Flickr (e.g., sky, water, car, etc.) rarely occur in the hashtags of microblogs. This means that previous methods using only visual features cannot be directly used for this task.

To overcome these problems, in this work, we propose a generative model to combine textual and visual information together. Because of the limited length of microblogs, hashtags may not even appear in the microblog textual content. Motivated by the methods proposed to handle the vocabulary gap problem for keyphrase extraction [17] and hashtag suggestion [14], we assume that the

* Corresponding author.
E-mail addresses: qz@fudan.edu.cn, qi_zhang@fudan.edu.cn (Q. Zhang).

[1] It is reported by the Pew Research Center's Internet & American Life Project in Aug 5, 2013.

[2] http://www.weibo.com. It is one of the most popular microblog services in China.

http://dx.doi.org/10.1016/j.neucom.2017.06.056
0925-2312/

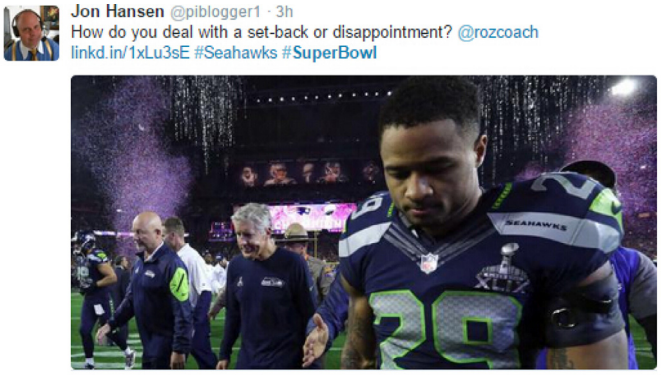

Fig. 1. An example of multimodal microblog.

hashtags, textual content, and visual words in a microblog are parallel descriptions of the same thing in different languages. Hence, we can regard the hashtag suggestion task as a translation problem. A bag-of-words (BoW) framework [18] is used to represent images. Local invariant image descriptors (e.g., SIFT) are extracted and quantized based on a set of visual words. To model the document themes, in this paper, we also propose the use of a topical translation model to facilitate the translation process. Topic-specific word triggers are used to bridge the gap between the visual words of images and hashtags. To demonstrate the effectiveness of our model, we carry out experiments on a large data set collected from a microblog service. Experimental results demonstrate that the proposed method can achieve better performance than state-of-the-art methods using textual or visual information only.

The main contributions of this work are summarized as follows:

- We propose and study the task of recommending hashtags for multimodal microblog posts. According to the best of our knowledge, it is the first work on this task. Existing works on hashtag recommendation task usually focused on textual information only.
- We propose a novel topical translation model to combine both textual and visual information. We think that it can also be easily adopted to achieve other multimodal tasks.
- We construct a large collection of multimodal microblogs from a real microblogging service. All the microblogs in the collection contain textual content, images, and hashtags labeled by their authors. This can benefit other researchers investigating the same task or other topics using multimodal social medium data.

2. Related works

Despite the fact that a lot of works have been conducted to study the problem of image tag suggestion and hashtag suggestion using textual information, to the best of our knowledge, this is the first work addressing the problem of hashtag suggestion for easy categorization and retrieval of multimodal microblogs. In this section, we give some brief descriptions about the following related research areas: tag recommendation and latent Dirichlet allocation.

2.1. Tag recommendation

Due to the usefulness of tag recommendation, many methods have been proposed from different perspectives [11,12,14,17,19].

Heymann et al. [11] investigated the tag recommendation problem using the data collected from social bookmarking system. They introduced an entropy-based metric to capture the generality of a particular tag. In [20], a Poisson Mixture Model based method is introduced to achieve the tag recommendation task. Krestel et al. [12] introduced a latent Dirichlet allocation to elicit a shared topical structure from the collaborative tagging effort of multiple users for recommending tags. Based on the observation that similar webpages tend to have the same tags, Lu et al. proposed a method taking both tag information and page content into account to achieve the task [21]. Ding et al. proposed to use translation process to model this task [14]. They extended the translation based method and introduced a topic-specific translation model to process the various meanings of words in different topics. In [22], discriminative-term-weights were used to establish topic-term relationships, of which users' perception were learned to suggest suitable hashtags for users. To handle the vocabulary problem in keyphrase extraction task, Liu et al. [17] proposed a topical word trigger model, which treated the keyphrase extraction problem as a translation process with latent topics.

Most of the works mentioned above are based on textual information. Besides these methods, visual information has also been paid lots of attentions [15,23–26]. Sigurbjörnsson and van Zwol studied the tag recommendation for images [15]. Their approach was based on the statistics of Flickr annotation patterns and tag co-occurrence statics. When a user submits a photo and enters some tags, an ordered list of candidate tags were derived for each of those entered tags. Hence, it cannot be directly applied to images only. In [23], the problem of personalized, interactive tag recommendation was also studied based on the statistics of the tags co-occurrence. Liu et al. proposed a tag ranking scheme to rank the tags associated with a given image according to their relevance to the image content [24]. They first calculated the initial relevance scores for the tags based on probability density estimation. Then a random walk is performed over a tag similarity graph to refine the relevance scores. In [26], ensembles of Support Vector Machines per tag was used to classify tag relevance.

From the brief descriptions given above, we can observe that most of the previous works focused on either textual information or visual features. In this work, the proposed method incorporates both textual and visual information using generative methods.

2.2. Latent Dirichlet allocation

Due to the capability of representing document themes, latent Dirichlet allocation (LDA) has been successfully used in natural language processing, image annotation, and various other applications [17,27–31]. In this paper, we also incorporate the topical translation model with textual and visual information based on it. In [27], multimodal LDA (mmLDA) extended LDA to describe the joint distributions of image and caption words in multimedia documents. It was based on the assumption that existence of a small set of shared latent variables that are the common causes of correlations between the 2 modalities. Putthividhy et al. also studied this task and believed that words from the two data modalities were potentially generated from non-overlapping sets of hidden topics. They proposed a topic-regression multi-modal LDA [31]. Instead of sharing the hidden topics directly between the visual and textual modalities, they used two sets of hidden topics and incorporated a linear regression module to correlate them. In this work, we also use LDA to model the latent topics. The hashtag selection algorithm takes the coverage of topics into consideration.

3. The proposed methods

In this section, we first introduce the notation and generation process of the model. Then, we describe the method used for learning models. Finally, we present the methods of how do we apply the learned model to achieve the hash tag recommendation task.

Please cite this article as: Y. Gong et al., Hashtag recommendation for multimodal microblog posts, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.06.056