Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
February 2017, 24(1): 67–76
www.sciencedirect.com/science/journal/10058885
The Journal of China Universities of Posts and Telecommunications
http://jcupt.bupt.edu.cn

Social media mining and visualization for point-of-interest recommendation

Ren Xingyi (✉), Song Meina, E Haihong, Song Junde

1. School of Computing, Beijing University of Posts and Telecommunications, Beijing 100876, China
2. Engineering Research Center of Information Networks, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract

With the rapid growth of location-based social networks (LBSNs), point-of-interest (POI) recommendation has become an important research problem. As one of the most representative social media platforms, Twitter provides various real-life information for POI recommendation in real time. Despite that POI recommendation has been actively studied, tweet images have not been well utilized for this research problem. State-of-the-art visual features like convolutional neural network (CNN) features have shown significant performance gains over the traditional bag-of-visual-words in unveiling the image's semantics. Unfortunately, they have not been employed for POI recommendation from social websites. Hence, how to make the most of tweet images to improve the performance of POI recommendation and visualization remains open. In this paper, we thoroughly study the impact of tweet images on POI recommendation for different POI categories using various visual features. A novel topic model called social media Twitter-latent Dirichlet allocation (SM-TwitterLDA) which jointly models five Twitter features, (i.e., text, image, location, timestamp and hashtag) is designed to discover POIs from the sheer amount of tweets. Moreover, each POI is visualized by representative images selected on three predefined criteria. Extensive experiments have been conducted on a real-life tweet dataset to verify the effectiveness of our method.

Keywords social media, Twitter, POI recommendation, visualization

1 Introduction

With the ever-accelerating development of modern technologies, social networking services have become increasingly popular nowadays. Among them, Twitter plays an important role in providing a worldwide information sharing platform. All this up-to-date check-in information describes the POIs which are happening all over the world. POI recommendation has been a popular research topic which aims at discovering meaningful POIs from the sheer amount of raw tweets data.

Different from the traditional text documents such as web pages, tweets contain a large number of meaningless and polluted content and are usually short. These characteristics present two challenges to the POI recommendation task. First, the limitation that a tweet has maximum 140 characters leads to a problem that most existing topic model may not work anymore for POI recommendation. Second, tweets are often overwhelmed by meaningless words and irrelevant images. Along with these challenges, tweets data also bring opportunities which are not found in conventional media. Apart from the textual content, tweets contain other features such as image, timestamp, location and hashtag. The benefits of incorporating different features into POI recommendation are twofold. First, these features provide supplemental information that facilitates the POI recommendation from different aspects, which enables users to have a better understanding with more vivid visualization. However, existing methods have the following two main limitations. First, although a fast-growing percentage of tweets contain

Received date: 05-09-2016
Corresponding author: Ren Xingyi, E-mail: xyren@bupt.edu.cn
DOI: 10.1016/S1005-8885(17)60189-4

images now, few efforts have been devoted to study these images in tweets and utilize them in the POI recommendation task. Second, even though different Twitter features have been utilized in different models (e.g., spatio-temporal information, images, hashtag), no one has comprehensively exploited all these features in one framework.

To tackle the above challenges and overcome the limitations, we propose a novel Twitter POI recommendation approach termed SM-TwitterLDA. The framework is illustrated in Fig. 1. Five different Twitter features including text, image, timestamp, location and hashtag, are extracted from the raw tweets stream and input to our SM-TwitterLDA model. Notably, only text is compulsory and the other four features may not be necessarily present. First of all, POIs are then periodically detected. Then, our proposed the POI recommendation method takes a querying user along with a querying time and a querying region as input, and automatically combines the learnt interest of the querying user and the local preference of the region to produce the top-k recommendations. In the final step, the POIs are visualized by summarizing images of all the tweets related to the POIs, where the representative images are selected based on our defined three criteria. Different from the traditional topic models, in our SM-TwitterLDA each short tweet is generated from one topic rather than a mixture of topics. Importantly, we apply an support vector machine (SVM) classifier as a preliminary filter to remove noisy images. We also take advantage of the state-of-the-art visual feature generated by CNN and specifically design a component in our SM-TwitterLDA model to maximize the profit from tweet images.

The main contributions of this paper are summarized as follows:

1) We design a novel generative probabilistic model for effective POI recommendation in Twitter which jointly models various Twitter features. The intrinsic correlations among these features are fully explored for the performance improvements.

2) We extensively study the impact of images on the Twitter for POI recommendation performance and address the issue of the noisy tweet images.

3) We define three criteria to select the representative images for POI recommendation visualization and verify its quality by a case study.

4) We conduct comprehensive experiments on a large real-world tweet dataset. The results prove the advantages of our SM-TwitterLDA outperforms other state-of-the-art POI recommendation techniques.

The remainder of this paper is organized as follows. In Sect. 2, we review the related work on POI recommendation and visualization. Our proposed POI recommendation is presented in Sect. 3, followed by the visualization for POI recommendation approach in Sect. 4. The experimental results are shown in Sect. 5 and Sect. 6 concludes the paper.

2 Related work

2.1 POI recommendation

With the growing popularity of LBSNs, applying POI recommendation to provide better location-based service has drawn a lot of attention from the academia and industry. POI recommendation in Twitter has been extensively studied in recent years. At the earliest, some research explored textual information for POI recommendation. Ferrari et al. [1] analyzed Twitter posts and performed LDA on the data to extract urban patterns, such as hotspots and crowd behaviors. Then, some work explored geographical and text information for POI recommendation. Yin et al. [2] proposed a location-content-aware recommendation system that offered a particular user a set of venues or events by giving consideration to both personal interest and local preference. Further, some research explored geographical and time information for POI recommendation. Yuan et al. [3] proposed a time-aware POI recommendation to recommend POIs for a given user at a specified in a day, they developed a collaborative recommendation model that is able to incorporate geographical and temporal information to enhance the recommendation. Moreover, some work explored geographical and categorical information for POI recommendation. Zhang et al. [4] proposed a POI recommendation approach through exploiting geographical correlations, social correlations and categorical correlations among users and POIs to predict the relevance score of a user to an unvisited POI so as to make recommendation for users. Recently, some research exploit geographical and image information for POI recommendation. Lim [5] proposed POI recommendation framework using geo-tagged photos and a POI list to construct user visit sequences for tour