Sentiment-Aware Multi-modal Recommendation on Tourist Attractions

Junyi Wang[1], Bing-Kun Bao[2(✉)], and Changsheng Xu[1,3,4]

[1] Hefei University of Technology, Hefei, China
jywang921@gmail.com
[2] College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China
bingkunbao@njupt.edu.cn
[3] National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing, China
csxu@nlpr.ia.ac.cn
[4] University of Chinese Academy of Sciences, Beijing, China

Abstract. For tourist attraction recommendation, there are three essential aspects to be considered: tourist preferences, attraction themes, and sentiments on themes of attraction. By utilizing vast multi-modal media available on Internet, this paper is aiming to develop an efficient solution of tourist attraction recommendation covering all these three aspects. To achieve this goal, we propose a probabilistic generative model called Sentiment-aware Multi-modal Topic Model (SMTM), whose advantages are four folds: (1) we separate tourists and attractions into two domains for better recovering tourist topics and attraction themes; (2) we investigate tourists sentiments on topics to retain the preference ones; (3) the recommended attraction is guaranteed with positive sentiment on the related attraction themes; (4) the multi-modal data are utilized to enhance the recommendation accuracy. Qualitative and quantitative evaluation results have validated the effectiveness of our method.

Keywords: Tourism recommendation · Multi-modal computing
Topic model · Sentiment analysis

1 Introduction

With the acceleration of globalization and fast development of technologies for travel needs, personalized tourism become more and more popular especially among younger generations. Nowadays, the opinions and sentiments shared on travel websites act an important role for tourists on attraction selection. TripAdvisor[1], one of the most popular ones in its kind, is being checked by many tourists for the multi-modal comments, including text and image, on the attractions they plan to visit. However, due to the fast growth of these travel websites,

[1] [online]. Available https://www.tripadvisor.in/.

I. Kompatsiaris et al. (Eds.): MMM 2019, LNCS 11295, pp. 3–16, 2019.
https://doi.org/10.1007/978-3-030-05710-7_1

the overwhelming and sometimes unorganized information become a hurdle to the tourists to find the pieces that values most to them. Therefore, it is not hard to see that a sharp increasing demand of customizable and automatic tourist attraction recommendation solutions.

To design a satisfying tourist attraction recommendation method, the mining on following three aspects from multi-modal data is crucial. The first one is "tourist preferences", which are the tourist topics that a tourist truly interested on among all those he/she visited, commented, followed, liked and so on. The second one is "attraction themes", which are the types of experience that tourists would received through visits. For example, the theme of Disney Land seems most likely to be "entertainment" over "historical site". Of course, an attraction could have multiple themes. And the last one is "sentiment on a theme of the attraction", which measures the quality of attraction on a certain theme by the view of tourists. Based on these three aspects, the task of recommending a tourist with the most suitable attractions can be decomposed into mining the preferences of this tourist, then selecting the most similar attraction themes, and at last returning the attractions with positive sentiments on selected themes.

Existing works on personalized tourism recommendation mostly ignore the sentiment analysis on both tourist preferences and attraction themes, and also lack of processing of multi-modal data [1,8]. [9,10] directly work on tourist topics instead of mining of tourist preferences. Without analyzing the sentiments of a tourist over the topics, there is a high chance that some of the recommended attractions are not what the tourist is looking forward. [11,15] recommend top searched attractions on the selected themes. This only reflects the popularity of the attractions but fails to reveal the sentiments from the tourists who actually visited them. [20] analyzes the sentiments yet neglects to separate tourists and attractions into two domains and identify tourist topics with attraction themes. Moreover, [9,10,20] only focus on text modality, and leave image modality out of consideration.

This paper emphasizes on mining all above mentioned three aspects, and proposes a Sentiment-aware Multi-modal Topic Model (SMTM), which is capable of discovering topics/themes of tourists/attractions conditioned on multi-modal tourism data and analyzing their sentiments for better recommendation results. Specifically, we divide tourists and attractions into two domains. As shown in Fig. 1, the left side is topic mining and sentiment analysis on tourist domain, while the right side is those on attraction domain. The inputs of two sides are the multi-modal corpus from tourist and attraction domains respectively. The outputs of two sides include topics/themes of each tourist/attraction, and their corresponding sentiments. The middle of Fig. 1 shows the applications of SMTM, including personalized attraction recommendation and potential tourist recommendation.

In summary, the contributions of this work are as follows:

– We analyze the three essential aspects on attraction recommendation, and propose Sentiment-aware Multi-modal Topic Model (SMTM) whose advantages are four folds: (1) we separate tourists and attractions into two domains for better recovering tourist topics and attraction themes; (2) we investigate tourist sentiments on topics to retain those with actual preference; (3) the recommended attraction is guaranteed with positive sentiment on the related attraction themes; (4) the multi-modal data are utilized to enhance the recommendation accuracy.

– We present two applications on SMTM, including personalized attraction recommendation and potential tourist recommendation. We also construct a large scale tourism recommendation dataset including 14,648 tourists and 8,724 attractions with multi-modal data. The experiments on this dataset show the effectiveness of our proposed approach, and the high accuracies on two applications.
– The proposed model has great potential for other recommendation applications, including movie, goods recommendations, and so on.

The rest of this paper is organized as follows. In Sect. 2, we briefly review the related work. Section 3 presents the details of Sentiment-aware Multi-modal Topic Model. Section 4 introduces the above mentioned two applications. Section 5 reports and analyzes experimental results. Finally, we conclude the work in Sect. 6.

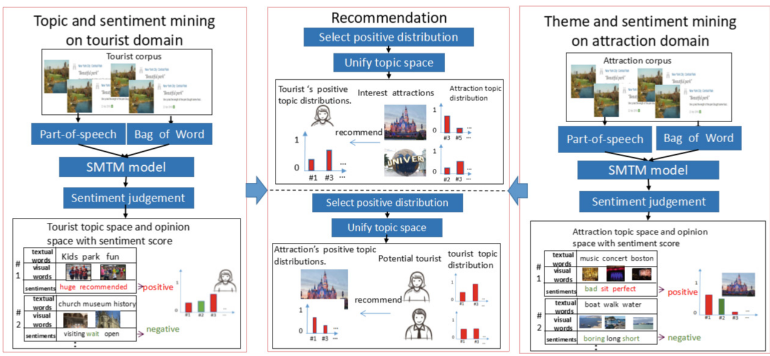


Fig. 1. Framework of sentiment-aware multi-modal recommendation

2 Related Work

Our work is related to two main research areas, that is, probabilistic topic models and personalized travel recommendation.

2.1 Probabilistic Topic Models

The Probabilistic Topic Model (PTM) is proposed to explore a set of topics from a document set, where a topic is a distribution over a fixed vocabulary and a document is a distribution over topics. The simplest probabilistic topic model [2] is Latent Dirichlet Allocation (LDA) [17]. In order to extend the LDA model to learn the joint correlations between data of different modalities, such as the texts and images, some variants of topic models are developed, such as multimodal-LDA and correspondence LDA [3]. They use a set of shared latent variables to explicitly model images and annotated text to capture correlations between the data of two modalities.

In order to mine subjective emotion, some works focus on studying topic models on sentiment/opinion mining [5–7,18,24,26–28]. Topic Sentiment Mixture model is proposed by Mei *et al.* [18] to reveal the latent topic facets in a Weblog collection, and their associated sentiments. In [27], Alam *et al.* propose a domain-independent topic-sentiment model called Joint Multi-gain Topic Sentiment to extract quality semantic aspects automatically, thereby eliminating the requirement for manual probing. These works are the pioneer studies on topic sentiment analysis. However they just consider limited states on sentiment spaces like "negative, positive, neutral" in [18]. To improve this, Titov [21] propose Multi-Aspect Sentiment model which can aggregate sentiment texts for the sentiment summary of each rating aspect. [23] propose Contrastive Opinion Modeling to present the opinions of the individual perspectives on the topic and furthermore to quantify their difference. Latter, some works research on sentiment topic model with multi-modal data. In [26], Huang *et al.* propose Multi-modal Joint Sentiment Topic Model for weakly supervised sentiment analysis on texts and emoticons in microblogging. Fang *et al.* [19] propose Multi-modal Aspect-opinion Model to consider both user-generated photos and textual documents and simultaneously capture correlations between textual and visual modalities. Extend to [23], Qian *et al.* [25] propose a multi-modal multi-view topic opinion mining model for social event analysis multiple collection sources. Different from above approaches which are based on single topic space, our work consider two domains, that is tourist and attraction, and two topic spaces associated.

2.2 Personalized Travel Recommendation

In recent years, high demand of travel recommendation solutions have led to a lot of researches. Some works [12–14,16] consider that the reviews contain diverse information which can mitigate sparse problems, and some of them also use reviews to extract sentiments for recommendation. But they ignore the theme of attractions. Some other works focus on mining the themes of attractions to facilitate trip planning [4,9,15]. By considering the tourist topics and attraction themes, Leal *et al.* [9] propose Parallel Topic Modelling to extract information and utilize semantic similarity to identify relevant recommendation. However, majority of those methods lack of analysis of sentiment on themes of attractions. [20] proposes a user interest modeling method called LSARS to represent the user's interest. [11] designs a personalized similarity (PAS) model which utilizes the heterogeneous travel information for recommendations. This method takes sentiments of the attraction themes into consideration and uses a multi-modal data, but fails to reveal the sentiments from the tourists who actually visited them.

3 Sentiment-Aware Multi-modal Topic Model

This section introduces the proposed Sentiment-aware Multi-modal Topic Model (SMTM), which is composed of theme and sentiment mining on tourist domain, topic and sentiment mining on attraction domain, and correlation analysis between tourist and attraction topic spaces, as shown in Fig. 2.

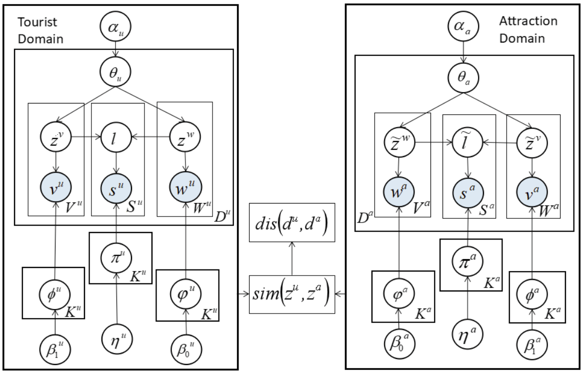


Fig. 2. Representation of sentiment-aware multi-modal topic model.

3.1 Problem Definition

Supposing that we are given a collection of tourist documents $\mathcal{U} = \{d_1^u, ..., d_{D^u}^u\}$ and a collection of attraction documents $\mathcal{A} = \{d_1^a, ..., d_{D^a}^a\}$, where $d_i^u = \{U_i^W, U_i^V, U_i^S\}$, $d_i^a = \{A_i^W, A_i^V, A_i^S\}$ are composed of three components: textual component U_i^W, A_i^W, visual component U_i^V, A_i^V, and sentiment component U_i^S, A_i^S. Table 1 lists the key notations. As discussed in Sect. 1, tourist preferences, attraction themes, and sentiments on themes of attraction are the three essential aspects for tourist attraction recommendation. We separate tourist and attraction into two domains, seek their associated theme spaces, analyze the sentiment of each topic for every tourist and attraction, and at last, find the correlation between tourist topic space and attraction theme spaces for recommendation. Thus, the problem of SMTM can be defined as follows:

Definition 1 *(Sentiment-aware Multi-modal Topic Model). Given two collections of tourists and attractions from travel websites, that is,* $\mathcal{U}$ *and* $\mathcal{A}$*, the goal of SMTM is to learn: (1) tourists topic space* φ^u*,* ϕ^u *and attraction theme space* φ^a*,* ϕ^a*; (2) corresponding sentiment space* π^u *and* π^a*; (3) the distributions of tourist domain document-topic* θ^u *and attraction domain document-theme* θ^a*; (4) the correlation between tourist and attraction topic spaces is measured by all the similarities between tourist topics and attraction themes, that is,* $sim(z^u, z^a)$*, for* $z^u \in \{1, 2, \cdots, K^u\}$ *and* $z^a \in \{1, 2, \cdots, K^a\}$*.*

3.2 Topic and Sentiment Mining on Tourist Domain

Topic and sentiment mining on tourist domain is to determine the preferences of each tourist. Specifically, each tourist document contains texts and images. By using our model, the textual and visual words are generated following document-topic distributions θ^u, while the sentiment words are generated from the sentiment distribution conditioned on the corresponding topics. Accordingly, the generative process of a tourist document d^u in SMTM can be described as follows.

Table 1. The key notations of the proposed sentiment-aware multi-modal topic model

Notations	Description
$\mathcal{U}, \mathcal{A}$	Tourist document set, attraction document set
D^u, D^a	Number of tourist documents and attraction documents
K^u, K^a	Number of tourist topics and attraction topics
U^W, U^V, U^S	Textual word vocabulary, visual word vocabulary and sentiment word vocabulary in tourist document set
A^W, A^V, A^S	Textual word vocabulary, visual word vocabulary and sentiment word vocabulary in attraction document set
W^u, V^u, S^u, W^a, V^a, S^a	Number of word in U^W, U^V, U^S, A^W, A^V, A^S
φ^u, ϕ^u, π^u	The multinomial distributions over textual words, visual words and sentiment words for tourist topics
φ^a, ϕ^a, π^a	The multinomial distributions over textual words, visual words and sentiment words for attraction topics
θ^u, θ^a	The multinomial distributions over topics for tourists and attractions
z^w, z^v, l	Topic assignment for textual word, visual word, and sentiment word on tourist domain
$\tilde{z}^w$, $\tilde{z}^v$, $\tilde{l}$	Topic assignment for textual word, visual word, and sentiment word on attraction domain
$\alpha^u, \alpha^a, \beta_0^u, \beta_0^a$ $\beta_1^u, \beta_1^a, \eta^u, \eta^a$	Dirichlet priors to multinomial distribution θ^u, θ^a, φ^u, φ^a, ϕ^u, ϕ^a, π^u, π^a
$\mathbf{w}_{-i}, \mathbf{z}_{-i}, \mathbf{v}_{-i}, \mathbf{s}_{-i}, \mathbf{l}_{-i}$	Vector values of $\mathbf{w}$, $\mathbf{z}$, $\mathbf{v}$, $\mathbf{s}$, $\mathbf{l}$ on all the other dimensions except i

1. For each tourist topic $z^u \in \{1, 2, \cdots, K^u\}$ including textual topic z^w and visual topic z^v, draw a multinomial distribution over topic words, $\varphi^u \sim Dir(\beta_0^u)$ and $\phi^u \sim Dir(\beta_1^u)$.
2. For each tourist topic $z^u \in \{1, 2, \cdots, K^u\}$, draw a multionmial sentiment word distribution $\pi^u \sim Dir(\eta^u)$.
3. For each document d^u:
   (a) draw a multinonmial distribution $\theta_d^u \sim Dir(\alpha^u)$ for document.
   (b) for each textual word w in document d^u: draw a topic $z_d^w \sim Multi\,(\theta_d^u)$, a textual word $w \sim Multi\left(\varphi^u_{z_d^w}\right)$.
   (c) for each visual word v in document d^u: draw a topic $z_d^v \sim Multi\,(\theta_d^u)$, a visual word $v \sim Multi\left(\phi^u_{z_d^v}\right)$.

(d) for each sentiment word s in document d^u: draw a topic assignment $l \sim Uniform\left(z_1^u, z_2^u, ..., z_{K^u}^u\right)$, a sentiment word $s \sim Multi\left(\pi_l^u\right)$.

We assume that the priors $\left(\alpha^u, \beta_0^u, \beta_1^u, \eta^u\right)$ follow symmetric Dirichlet in modeling learning process, where the symmetric Dirichlet are conjugate priors for multinomial distribution.

After modeling the tourist domain of SMTM, we use Gibbs sampling [25] for model inference. There are three sets of variables involved, that is, textual topic assignment $\mathbf{z}^w$, visual topic assignment $\mathbf{z}^v$ and sentiment distribution $\mathbf{l}$. In a Gibbs sampler, one iteratively samples new assignments of latent variables by drawing from the distributions conditioned on the previous state of the model. For tourist domain of SMTM model, the update rules for latent variables are as follows.

The rule for the latent variables $\mathbf{z}^w$, $\mathbf{z}^v$ and $\mathbf{l}$:

$$p(z_i^w = k^u|\mathbf{w}, \mathbf{z}_{-i}^w) \propto \frac{n_{kd,-i}^u + \alpha^u}{\sum_{k=1}^{K^u} n_{kd,-i}^u + K^u\alpha^u} \times \frac{n_{wk,-i}^u + \beta_0^u}{\sum_{w=1}^{W^u} n_{wk,-i}^u + W^u\beta_0^u} \tag{1}$$

$$p(z_i^v = k^u|\mathbf{v}, \mathbf{z}_{-i}^v) \propto \frac{n_{kd,-i}^u + \alpha^u}{\sum_{k=1}^{K^u} n_{kd,-i}^u + K^u\alpha^u} \times \frac{n_{vk,-i}^u + \beta_1^u}{\sum_{v=1}^{V^u} n_{vk,-i}^u + V^u\beta_1^u} \tag{2}$$

$$p(l_i = m^u|\mathbf{s}, \mathbf{l}_{-i}) \propto \frac{n_{sm,-i}^u + \eta^u}{\sum_{s=1}^{S^u} n_{sm,-i}^u + S^u\eta^u} \times \frac{n_{md}^u}{N_{kd}^u} \tag{3}$$

where the symbol $-i$ means a counting variable that excludes the i-th word index in the corpus. $n_{kd,-i}^u$ denotes the times of words for topic k^u being generated from document d^u except the current assignment. $n_{wk,-i}^u$ denotes the times of word w being generated from topic k^u except the current assignment. $n_{vk,-i}^u$, $n_{sm,-i}^u$,n_{md}^u, is similar. N_{kd}^u means the times of all topic words in document d^u.

After sampling, the tourist domain parameters can be estimated as follows:

$$\begin{aligned} \theta_{kd}^u &= \frac{n_{kd}^u + \alpha^u}{\sum_{k=1}^{K^u} n_{kd}^u + K^u\alpha^u}, \varphi_{wk}^u = \frac{n_{wk}^u + \beta_0^u}{\sum_{w=1}^{W^u} + W^u\beta_0^u}, \\ \phi_{vk}^u &= \frac{n_{vk}^u + \beta_1^u}{\sum_{v=1}^{V^u} + V^u\beta_1^u}, \pi_{sm}^u = \frac{n_{sm}^u + \eta^u}{\sum_{s=1}^{S^u} n_{sm}^u + S^u\eta^u} \end{aligned} \tag{4}$$

In this paper, the SentiWordNet, a popular linguistics based sentiment model, is used to set a sentimental value (between -1 and 1) to every sentiment word. The value of is closer to 1, it is more likely to be positive, otherwise to be negative. Then the sentiment score of each tourist topic is,

$$Q\left(\mathbf{z}^w, \mathbf{l}\right) = \frac{1}{2}\left[\sum_{w=1}^{Nwk} p\left(w|z^w = k\right) \cdot Q_w + \sum_{s=1}^{Nsk} p\left(s|l = k\right) \cdot Q_s\right] \tag{5}$$

In this equation, $Q\left(s\right)$ and $Q\left(w\right)$ are the individual sentiment scores of a word. $Q\left(\mathbf{z}_w, \mathbf{l}\right)$ represents the overall sentiment tendency of the topic.

3.3 Theme and Sentiment Mining on Attraction Domain

Theme mining on attraction domain is to determine attraction themes, while sentiment mining is to determine sentiments on themes of attraction. Similar to those on tourist domain, we use the same process. So we just express the key formulas below. The update rules are as follows.

$$p(\tilde{z}_i^w = k^a|\mathbf{w}, \tilde{\mathbf{z}}_{-i}^w) \propto \frac{n_{kd,-i}^a + \alpha^a}{\sum_{k=1}^{K^a} n_{kd,-i}^a + K^a\alpha^a} \times \frac{n_{wk,-i}^a + \beta_0^a}{\sum_{w=1}^{W^a} n_{wk,-i}^a + W^a\beta_0^a} \tag{6}$$

$$p(\tilde{z}_i^v = k^a|\mathbf{v}, \tilde{\mathbf{z}}_{-i}^v) \propto \frac{n_{kd,-i}^a + \alpha^a}{\sum_{k=1}^{K^a} n_{kd,-i}^a + K^a\alpha^a} \times \frac{n_{vk,-i}^a + \beta_1^a}{\sum_{v=1}^{V^a} n_{vk,-i}^a + V^a\beta_1^a} \tag{7}$$

$$p(\tilde{l}_i = m^a|\mathbf{s}, \tilde{\mathbf{l}}_{-i}) \propto \frac{n_{sm,-i}^a + \eta^a}{\sum_{s=1}^{S^a} n_{sm,-i}^a + S^a\eta^a} \times \frac{n_{md}^a}{N_{kd}^a} \tag{8}$$

After sampling, the attraction domain parameters can be estimated as follows:

$$\begin{aligned} \theta_{kd}^a &= \frac{n_{kd}^a + \alpha^a}{\sum_{k=1}^{K^a} n_{kd}^a + K^a\alpha^a}, \varphi_{wk}^a = \frac{n_{wk}^a + \beta_0^a}{\sum_{w=1}^{W^a} + W^a\beta_0^a}, \\ \phi_{vk}^a &= \frac{n_{vk}^a + \beta_1^a}{\sum_{v=1}^{V^a} + V^a\beta_1^a}, \pi_{sm}^a = \frac{n_{sm}^a + \eta^a}{\sum_{s=1}^{S^a} n_{sm}^a + S^a\eta^a} \end{aligned} \tag{9}$$

3.4 Correlation Analysis Between Tourist and Topic Space and Attraction Theme Space

To correlate tourist topic space and attraction theme space, we calculate all the similarities between tourist topics and attraction themes. Inspired by [22], we use symmetric Kullback-Leibler (KL) to measure the similarity, that is,

$$sim(z^u, z^a) = \sum_i p\left(i|z^u\right) log \frac{p\left(i|z^u\right)}{p\left(i|z^a\right)} + \sum_i p\left(i|z^a\right) log \frac{p\left(i|z^a\right)}{p\left(i|z^u\right)} \tag{10}$$

where i indexes the word which occurs in both domains.

4 Applications

In this section, we introduce how to leverage the learned SMTM to enable two interesting applications: personalized attraction recommendation and potential tourist recommendation.

4.1 Personalized Attraction Recommendation

For a tourist d_i^u and a query attraction d_i^a in $\mathcal{A} = \{d_1^a, d_2^a..., d_n^a\}$, the topic and theme distribution $\theta_{d_i}^u$,$\theta_{d_i}^a$ and the topic and theme space $\mathbf{z}^u$, $\mathbf{z}^a$ can be obtained by model. Taking into account the sentimental factors, we use the Eq. (5) in Sect. 3.2 to learn sentiment score $Q(z)$ of a topic z. Then use it to compute sentiment binary variable q^z of topic z for further comparison recommendation using the following equation:

$$q^z = \begin{cases} 1, \text{if } Q(z) \geq \sigma \\ 0, \text{if } Q(z) < \sigma \end{cases} \tag{11}$$

where σ is threshold parameter. We calculate the distance between d_i^u and each document of $\mathcal{A}$ to rank the recommended attractions by following equation:

$$\begin{aligned} dis\left(d_i^u, d_i^a\right) &= \sum_z \sqrt{\left(q^z \cup q^{z'}\right) \times \left[p\left(z|d_i^u\right) - p\left(z'|d_i^a\right)\right]^2} \\ &= \sum_z \sqrt{\left(q^z \cup q^{z'}\right) \times \left(\theta_{d_i,z}^u - \theta_{d_i,z'}^a\right)^2} \end{aligned} \tag{12}$$

Where z is the topic in $\mathbf{z}^u$. $z^{'}$ is a corresponding theme to z in $\mathbf{z}^a$, which is calculated by semantic similarity according to Sect. 3.4.

4.2 Potential Tourist Recommendation

Potential tourist recommendation is similar to the interest attraction recommendation. Specifically, given an attraction d_i^a and a tourists d_i^u in set $\mathcal{U} = \{d_1^u, d_2^u..., d_n^u\}$. The theme and topic distribution $\theta_{d_i}^a$,$\theta_{d_i}^u$ and the topic and theme space $\mathbf{z}^a$, $\mathbf{z}^u$ can be learned by model. q^r and $q^{r^{'}}$ are obtained using the method mentioned in Sect. 4.1. Then we calculate the distance between d_i^a and each tourist from $\mathcal{U}$ to rank the recommended tourists by follows:

$$dis\left(d_i^a, d_i^u\right) = \sum_r \sqrt{\left(q^r \cup q^{r'}\right) \times \left(\theta_{d_i,r}^a - \theta_{d_i,r'}^u\right)^2} \tag{13}$$

Where r is the theme in $\mathbf{z}^a$. $r^{'}$ is a corresponding topic to r in $\mathbf{z}^u$, which is also calculated by semantic similarity according to Sect. 3.4. The *topk* of the potential tourists is recommended to tourist attraction d_i^a .

5 Experiment

5.1 Experimental Settings

The evaluation dataset was constructed from TripAdvisor, an online travel website. We collected multi-modal data from tourist and attraction domains respectively. For tourist domain, we collected 14,648 tourists, including their comments, descriptions and images. For attraction domains, we collected 8724 attractions with at least 30 comments and 1 image. In total, we have 459,160 textual comments or descriptions, and 43,944 images in tourist domain, while 392,580 textual comments and 26,172 images in attraction domain.

For each image from both domains, we represent image visual content by SIFT-Bow feature with 968 visual words. With the assumption similar to that used in [23,25], we extract all the nouns in the documents as the textual words and the adjectives, verbs, adverbs as the sentiment words. To classify tokens into nouns, adjectives, verbs, and adverbs, we use the Part-of-Speech tagging function provided by Stanford NLP toolkits[2]. Here we set Dirichlet hyper parameters of $\alpha^u = \alpha^a = 50/\mathrm{K}$ and $\beta_0^u = \beta_1^u = \beta_0^a = \beta_1^a = 0.02$, $\eta^u = \eta^a = 0.01$ for all the experiments.

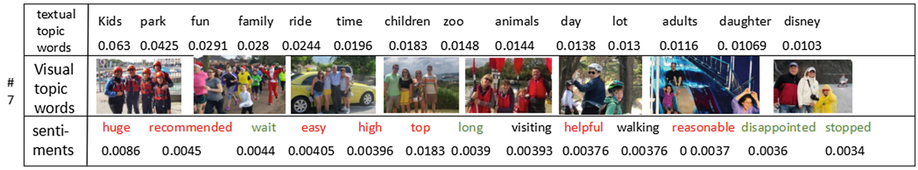


Fig. 3. Sample of topic words with corresponding sentiment words.

5.2 Evaluation of Sentiment-Aware Multi-modal Topic Model

Qualitative Analysis. We demonstrate the effectiveness of SMTM by examining the extracted topics with their sentiments and visualize them by providing the top ranked words. Both tourist and attraction domains are represented by a set of topic and theme distributions with textual words, visual words and corresponding sentiment words. Figure 3 presents a sample of topic words composed of textual words and visual words, and sentiment words, among which positive ones are in red while negative ones are in green, on tourist domain.

Figure 3 shows the partial results of the tourist domain. Here, topic #7 is about the topic "family travel", where textual words are closely related to the image patches. Sentimental words that describe the topic reflect the tourist's preferences, such as"recommended" and "disappointed".

Quantitative Evaluation. To evaluate our model, the perplexity is used as the metric. The perplexity score can be used to measure the generalization ability of a model, the lower the score is, the better capacity the topic model has. The perplexity for a set of test documents D_t is calculated as follows:

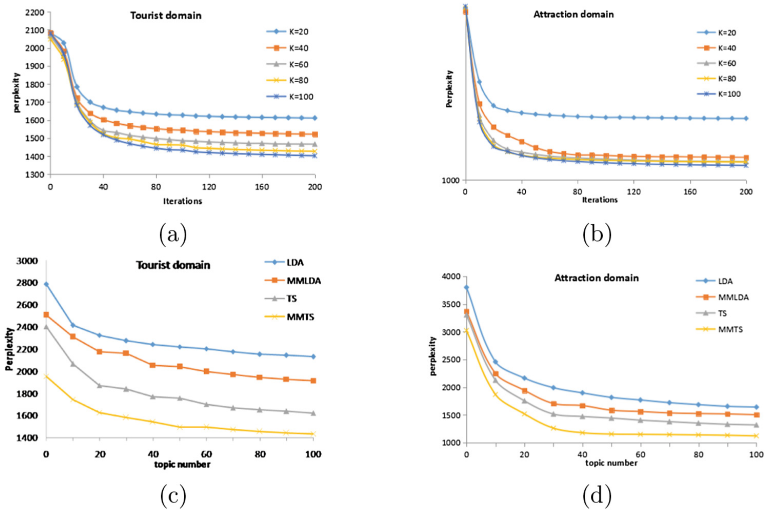
$$perplexity(D_t) = \exp - \frac{\sum_{d \in D_t} logp\left(\mathbf{w}_d, \mathbf{v}_d, \mathbf{o}_d\right)}{\sum_{d \in D_t} \left(N_{w,d} + N_{v,d} + N_{o,d}\right)} \tag{14}$$

where $p\left(\mathbf{w}_d, \mathbf{v}_d, \mathbf{o}_d\right) = p\left(\mathbf{w}_d\right) + p\left(\mathbf{v}_d\right) + p\left(\mathbf{o}_d | \mathbf{w}_d, \mathbf{v}_d\right)$.

[2] [online]. Available http://nlp.stanford.edu/software/index.shtml.

In our experiment, the data sets are divide into two parts separately: 80% are set as the training set and the rest are set as the test set. Figure 4(a) and (b) shows the perplexity of SMTM model in different topic numbers. It can be seen that, as the number of iterations increases, the degree of perplexity decreases, and it tends to stabilize when iterating about 100 times.

The baselines include: LDA, which treats all words as topic words. multi-modal LDA(mm-LDA), which model extends LDA by considering two modality of textual and visual. Topic-Sentiment(TS), which mines topic and sentiment on single domain. Figure 4(c) and (d) show the perplexity scores of all models in both data sets. From the result we can see LDA get a highest score which means the worst ability. This is because LDA only models textual words and sentiment words but does not distinguish them. The mm-LDA and TS model get a better performance than LDA, because of the additional dependencies of visual or sentiment information. The proposed SMTM model achieves the best results than other topic models on both domains.

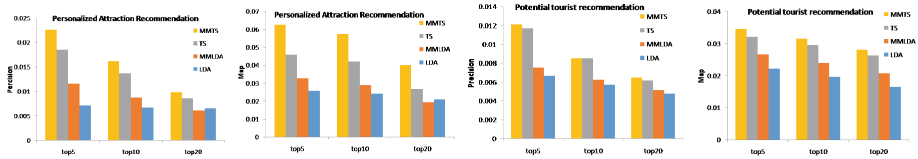


(a) (b) (c) (d)

Fig. 4. Perplexity of different topic numbers and different models

5.3 Evaluation of Cross-Domain Multi-modal Recommendation

To evaluate the validity of the two recommendations, two test sets are created. 1,261 tourists who have visited at least 15 attractions are selected from all tourists. And a total of 2411 tourist destinations, which have been visited by at lease 15 tourists, are selected from all the attractions. After the model is completed, the formula obtained in Sect. 4 was used to make two recommendations. Referring to traditional search metrics, $Precision@k$ and $MAP@k$ are used to measure two recommendations.



(a) (b) (c) (d)

Fig. 5. Precision and MAP of two recommendations

We report the *Precision*@*k* and *MAP*@*k* for two settings when K is 5, 10 and 20. Figure 5(a) and (b) show the performance comparison for personalized travel recommendation. It can be observed that LDA performance is the worst because it lacks the ability to mine the potential relationship between multi-modal topics and sentiments. The mm-LDA and TS perform better than LDA because they capture the consistency between different modalities. This indicates that visual or sentimental information is useful for improving recommendation performance. SMTM performs better than all baselines, which proves that an effective combination of textual data, visual data, and sentiments can improve data mining capabilities and help to achieve recommendations. Similar results can be observed in Fig. 5(c) and (d), which reports the performance of potential tourist recommendation. With taking both tourists' interests and attraction' comments into account, SMTM achieves the best results comparing to baselines.

6 Conclusions

In this paper, we proposed Sentiment-aware Multi-modal Topic Model to address the recommendation problem by considering tourist preference, attraction theme, and sentiment of attraction theme. SMTM is capable of mining multi-modal tourist topics and attraction themes both with corresponding sentiments. In future work, we plan to model the shared topic space to correlate two domains instead of using similarities.

Acknowledgement. This work is supported by the National Key Research & Development Plan of China (No. 2017YFB1002800), by the National Natural Science Foundation of China under Grant 61872424, 61572503, 61720106006, 61432019, and by NUPTSF (No. NY218001), also supported by the Key Research Program of Frontier Sciences, CAS, Grant NO. QYZDJ-SSW-JSC039, and the K.C.Wong Education Foundation.

References

1. Adomavicius, G., et al.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**(6), 734–749 (2005)

2. Blei, D., Carin, L., Dunson, D.: Probabilistic topic models. IEEE Signal Process. Mag. **27**(6), 55–65 (2010)
3. Blei, D.M., Jordan, M.I.: Modeling annotated data. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 127–134 (2013)
4. Huang, C., Wang, Q., Yang, D., et al.: Topic mining of tourist attractions based on a seasonal context aware LDA model. Intell. Data Anal. **22**(2), 383–405 (2018)
5. Bao, B.K., Xu, C., Min, W., Hossain, M.S.: Cross-platform emerging topic detection and elaboration from multimedia streams. TOMCCAP **11**(4), 54 (2015)
6. Bao, B.-K., Liu, G., Changsheng, X., Yan, S.: Inductive robust principal component analysis. IEEE Trans. Image Process. **21**(8), 3794–3800 (2012)
7. Bao, B.-K., Zhu, G., Shen, J., Yan, S.: Robust image analysis with sparse representation on quantized visual features. IEEE Trans. Image Process. **22**(3), 860–871 (2013)
8. Borras, J., Moreno, A., Valls, A.: Intelligent tourism recommender systems: a survey. Expert Syst. Appl. **41**(16), 7370–7389 (2014)
9. Leal, F., González–Vélez, H., Malheiro, B., Burguillo, J.C.: Semantic profiling and destination recommendation based on crowd-sourced tourist reviews. Distributed Computing and Artificial Intelligence, 14th International Conference. AISC, vol. 620, pp. 140–147. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-62410-5_17
10. Yang, D., Zhang, D., Yu, Z., et al.: A sentiment-enhanced personalized location recommendation system. In: ACM Conference on Hypertext and Social Media, pp. 119-128. ACM (2013)
11. Shen, J., Deng, C., Gao, X.: Attraction recommendation: towards personalized tourism via collective intelligence. Neurocomputing **173**, 789–798 (2016)
12. Kurashima, T., Iwata, T., Irie, G., Fujimura, K.: Travel route recommendation using geotags in photo sharing sites. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, Toronto, Canada, pp. 579–588. ACM, October 2010
13. Wu, Y., Ester, M.: FLAME: a probabilistic model combining aspect based opinion mining and collaborative filtering. In: Eighth ACM International Conference on Web Search and Data Mining, pp. 199–208. ACM (2015)
14. Arbelaitz, O., Gurrutxaga, I., Lojo, A., Muguerza, J., Perez, J.M., Perona, I.: Web usage and content mining to extract knowledge for modelling the users of the Bidasoa Turismo website and to adapt it. Expert Syst. Appl. **40**(18), 7478–7491 (2013)
15. Hao, Q., et al.: Equip tourists with knowledge mined from travelogues. In: Proceedings of the 19th International Conference on World Wide Web, pp. 401–410. ACM (2010)
16. Jiang, K., Wang, P., Yu, N.: ContextRank: personalized tourism recommendation by exploiting context information of geotagged web photos. In: 2011 Sixth International Conference on Image and Graphics, Hefei, Anhui, China, pp. 931–937. IEEE, August 2011 (2011)
17. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J Mach. Learn. Res. **3**, 993–1022 (2003)
18. Mei, Q., Ling, X., Wondra, M., et al.: Topic sentiment mixture: modeling facets and opinions in weblogs. In: Proceedings of the 16th International Conference on World Wide Web, pp. 171–180 (2007)

19. Fang, Q., Xu, C., Sang, J., et al.: Word-of-mouth understanding: entity-centric multimodal aspect-opinion mining in social media. IEEE Trans. Multimedia **17**(12), 2281–2296 (2015)
20. Xiong, H., Xiong, H., Xiong, H., et al.: A location-sentiment-aware recommender system for both home-town and out-of-town users, pp. 1135–1143 (2017)
21. Titov, I., McDonald, R.: A joint model of text and aspect ratings for sentiment summarization. In: ACL-08: HLT, pp. 308–316. Association for Computational Linguistics (2008)
22. Olszewski, D.: Fraud detection in telecommunications using Kullback-Leibler divergence and latent Dirichlet allocation. In: Dobnikar, A., Lotrič, U., Šter, B. (eds.) ICANNGA 2011. LNCS, vol. 6594, pp. 71–80. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-20267-4_8
23. Fang, Y., Si, L., Somasundaram, N. Yu, Z.: Mining contrastive opinions on political texts using cross-perspective topic model. In: Proceedings of the fifth ACM international conference on Web search and data mining, pp. 63–72. ACM (2012)
24. Lin, C., He, Y., Everson, R., et al.: Weakly supervised joint sentiment-topic detection from text. IEEE T. Knowl. Data En. **24**(6), 1134–1145 (2012)
25. Qian, S., Zhang, T., Xu, C., et al.: Multi-modal event topic model for social event analysis. IEEE Trans. Multimedia **18**(2), 233–246 (2016)
26. Huang, F., Zhang, S., Zhang, J., et al.: Multimodal learning for topic sentiment analysis in microblogging. Neurocomputing **253**(C), 144–153 (2017)
27. Alam, M.H., Ryu, W.J., Lee, S.K.: Joint multi-grain topic sentiment: modeling semantic aspects for online reviews. Inf. Sci. **339**, 206–223 (2016)
28. Min, W., Bao, B.K., Mei, S., et al.: You are what you eat: exploring rich recipe information for cross-region food analysis. IEEE Trans. Multimed. 1 (2017)