2018 IEEE Fourth International Conference on Big Data Computing Service and Applications

ASSR:An Adjustable Scenic Spot Rating System Based on Travel Note Mining

Xiaowen Mei, Yushun Fan*
Tsinghua National
Laboratory for Information
Science and Technology
Department of Automation
Tsinghua University
Beijing 100084, China
mxw16@mails.tsinghua.edu.cn
fanyus@tsinghua.edu.cn

Junqi Zhang, Bing Bai, Zhenfeng Gao
Tsinghua National
Laboratory for Information
Science and Technology
Department of Automation
Tsinghua University
Beijing 100084, China
jq-zhang15/bb13/gzf13@mails.tsinghua.edu.cn

***Abstract*—Scenic spot rating is a helpful and straightforward metric to evaluate scenic spots and assist travel plan making. However, existing ratings from the tourism websites are always unadjustable and strongly correlated with the popularity of the spots while different tourists may have different tastes. To solve this problem, we propose an Adjustable Scenic Spot Rating System (ASSR) based on travel note mining. With the help of the information mined from numerous travel notes, which contains the sentiment of tourists towards these numerous spots, our system is able to evaluate the scenic spots properly. We implement sentiment analysis to detect sentiment polarity of the travel notes. Based on this, we bring up a Spot-Topic LDA (ST-LDA) to generate a sentiment score for each spot which reflects its quality. In this case, the popularity of spots becomes an adjustable factor to satisfy the demands of different people. Our model is validated on public databases and has been embedded in the Smart Tourism Services Platform (STSP).**

***Index Terms*—scenic spot rating; travel note; sentiment analysis**

I. Introduction

With the rapid development of toursim, personalized services for tourists have drawn more attention in recent years [1], [2]. Thanks to the explosive growth in tourism data, we are able to mining meaningful information from massive tourism data [3]. As the rating of scenic spots can assist tourists to evaluate the spots directly, providing a personalized spots ranking for the specific tourist can be a meaningful service. The tourists will make a suitable travel plan more efficiently with the highly recommended spots from the spot ranking. However, there are some limitations on the existing rankings. Most existing rankings are static, while different people may have different preferences on the spots. For example, newcomers often prefer popular spots while natives often like some non-popular spots better. Moreover, the existing rankings often rely on direct feedbacks from tourists, thus they are highly biased to the popularity of spots, which means that we can hardly find obscure spots on the list which are little-known but worth a visit. So it is challenging and meaningful to provide an adjustable ranking of spots.

In this paper, we use travel notes to evaluate scenic spots. On the one hand, unlike the official description of a scenic spot, which is full of praises and positive opinions, travel notes represent the personal opinions on the spots from numerous and various tourists who have been there [4]. Therefore, we will have an unbiased evaluation of the spots. On the other hand, travel notes can cover a wider range of spots than collecting the explicit feedbacks from tourists, which make it possible to find more highly praised but little-known spots.

However, there are still some challenges to extract useful information from the notes and link them to the corresponding spots. For example, a travel note usually involves multiple scenic spots which means the opinions on different spots are mixed up and cannot be separated directly. Moreover, there are too many irrelevant descriptions in the travel notes, for example, "It's a nice day! ", which are completely unrelated to any spot. And these irrelevant descriptions will obviously impede extracting useful opinions on the spots.

To settle the aforementioned issues and provide both credible and adjustable rankings, we propose a novel system termed Adjustable Scenic Spot Rating System (ASSR), which can mine the sentiment of tourists and the popularity of scenic spots from travel notes and then quantitatively rate these spots. The structure of our model is illustrated in Fig. 1.

To detect the sentiment polarity of a travel note, we divide the travel note into several paragraphs which we call "sub-notes". Then we apply sentiment analysis to the sub-notes instead of the whole note. So we can acquire the sentiment scores of different parts, which are more accurate and meaningful than only one score generated from the whole note.

To settle the problem of linking scores with spots, we introduce Spot-Topic LDA (ST-LDA), a probabilistic topic model extending the Latent Dirichlet Allocation (LDA) [5]. The basic idea of ST-LDA is that both spots and scores are related to a latent variable "topic". By estimating the spot-topic proportions and topic-score proportions, we can

* Corresponding Author

0-7695-6396-1/18/$31.00 ©2018 IEEE
DOI 10.1109/BigDataService.2018.00023

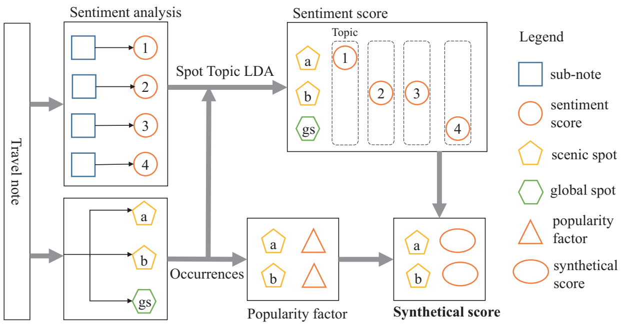

Fig. 1. The Framework of ASSR. By applying sentiment analysis and ST-LDA, we generate a distribution over sentiment scores to each spot. Meanwhile, we measure the popularity factor based on the occurrences of the spot. By integrating these two parts and adjusting the weight of popularity factor, we can generate the adjustable synthetical score for each spot.

automatically assign the score to the corresponding spot to generate a sentiment score for each spot.

To reduce the influence of irrelevant descriptions, we introduce the "global spot". The hypothesis is that if a description is irrelevant to individual spot, it will be assigned to the global spot automatically and thus it can reduce the negative impact of the irrelevant descriptions.

To meet the various demands from various tourists, such as the new-comers and the natives, we also take the popularity into consideration and design an adjustable factor based on popularity. For people with different preferences on popularity, we can change the popularity factor and provide personalized rankings for them.

The main contribution of this paper are summarized as follows:

1) We propose a rating system ASSR to quantitatively rate the spots based on numerous travel notes, and provide unbiased rankings of scenic spots.
2) The ratings are adjustable according to the preferences of tourists. When we choose a relatively low weight of popularity, we can find some obscure spots which are less well known but worth visiting in the top of the ranking list. And these spots are often ignored by most of famous travel platforms.
3) Comprehensive experiments on the web data sets show that our personalized rankings of the scenic spots are effective and reasonable. And the result of our model has been embedded in the Smart Tourism Service Platform[1].

The rest of this paper is organized as follows. Section II presents the framework of our ASSR. Section III exports the experiment results of applying our model to the web data sets. Section IV summarizes the related work and Section V concludes the whole paper.

II. MODEL FRAMEWORK

In this section, we present several notations used in our model, then describe the framework of the model, including three main components: sentiment analysis, ST-LDA and popularity adjustment.

A. Background Notations

In our model, we mainly use the travel notes and scenic spots as the dataset. The notations for them are defined as follows.

For travel notes, we use subscript "i" to denotes that this notation is related to travel note i, where $i = 1 : I$. A travel note is divided to several paragraphs and each of them is regarded as a sub-note, and we use subscript "j" to denote each sub-note, where $j = 1 : J$. So the i-th travel note F_i is defined as $F_i = \{f_{i1}, \cdots, f_{ij}, \cdots, f_{iJ_i}\}$, where f_{ij} is the j-th sub-note in i-th travel note. Moreover, g_{ij} denotes the sentiment score of sub-note f_{ij} respectively.

For scenic spots, we use subscript "k" to denote that this notation is related to spot k, where $k = 1 : K$. And $S_i = \{s_{ik_1}, s_{ik_2}, \cdots, s_{ik_i}\}$ is the set of scenic spots occurring in the travel note i.

B. Sentiment Analysis

Sentiment analysis, a branch of natural language processing (NLP), has received increasing attention and developed rapidly these years [6], [7]. Conventional sentiment analysis usually includes two parts, emotion recognition and polarity detection [8]. Here we mainly use the second part. By applying sentiment polarity detection on the travel notes, which record tourists' opinions on spots, we can transform the semantic descriptions into numerical data and find out whether a description expresses a positive or a negative opinion.

Since positive descriptions and negative descriptions may be neutralized in a whole travel note, it is hard to extract the sentiment information exactly. To solve this problem, we split up each travel note into small paragraphs with an appropriate size, which we called "sub-notes", and then apply sentiment analysis to them separately. The sub-notes should not be too long otherwise the separation will make no sense. But if sub-notes are too short, the comments on one spot will be separated into many parts, and most of them are incomplete sentences which will lead to the inaccuracy of sentiment analysis.

To get sub-notes in an appropriate size, we split up the long travel notes into short sentences based on all kinds of punctuations (both in Chinese input methods and English input methods), including ",", ".", "...", "!", "?", etc. Then we combine the short sentences into sub-notes with a length around 80 words.

After splitting up all the travel notes, we apply sentiment polarity detection to each sub-notes separately. We choose an open API named TextSentiment from Tencent NLP[2], a platform focusing on Chinese natural language processing. The interface information is demonstrated in Table I.

For an input f_{ij}, the j-th sub-note of i-th travel note, we use the output "positive" value as the sub-note's sentiment score g_{ij}, and $g_{ij} \in [0, 1]$. A high sentiment score means the sentiment polarity is positive.

[1]https://mp.weixin.qq.com/s/28eaqQVEaUSCZp3aSEzHpw

[2]http://nlp.qq.com/

TABLE I
INPUT AND OUTPUTS OF TEXTSENTIMENT API

Input	content	Chinese text waiting to be analyzed
Outputs	positive	probability of positive sentiment
	negative	probability of negative sentiment

At last, we discretize the sentiment score g_{ij} into discrete sentiment score $\widetilde{g_{ij}}$, and $\widetilde{g_{ij}} \in \{1, \cdots, V\}$. Then we use several discrete sentiment scores to describe every travel note. The discretization criteria is:

$$\widetilde{g_{ij}} = \lceil g_{ij} * V \rceil \tag{1}$$

C. Spot Topic LDA

After the sentiment analysis, every travel note is transformed to several discrete sentiment scores. Also each travel note is related to several spots. In ST-LDA, both spots and scores have a relationship with the latent topic.

What we need to mention here is that we introduce a global spot, denoted as gs, to filter the scores of background descriptions in the travel notes. So we add gs to the spots set of every travel note. For travel note F_i, we have the discrete scores set $\widetilde{G_i} = \{\widetilde{g_{i1}}, \widetilde{g_{i2}}, \cdots, \widetilde{g_{in_i}}\}$ and the spots set with global spot $\widetilde{S_i} = \{s_{i1}, s_{i2}, \cdots, s_{ik_i}, gs\}$.

In ST-LDA, to generate each score in travel note F_i with spots set $\widetilde{S_i}$, we can choose one spot uniformly at random. Then we choose a topic based on the distribution over topic specific to the spot. At last we choose a score base on the distribution over the scores to the topic chosen before.

The generative process can be described more clearly as follows. Assuming that there are T topics, V scores, K spots and a global spot gs.

1) For each spot $s \in \{s_1, s_2, \cdots, s_K, gs\}$, draw topic proportions $\theta_s \sim Dirichlet(\alpha)$ in T-dimensional.
2) For each topic $t \in \{t_1, t_2, \cdots, t_T\}$, draw score proportions $\phi_t \sim Dirichlet(\beta)$ in V-dimensional.
3) For each discrete score $\widetilde{g_{in}} \in \{\widetilde{g_{i1}}, \widetilde{g_{i2}}, \cdots, \widetilde{g_{in_i}}\}$ in each travel note F_i:
   a) Draw a scenic spot assignment $s_{in} \sim Uniform(s_{i1}, s_{i2}, \cdots, s_{ik_i}, gs)$
   b) Based on spot s_{in}, draw a topic assignment $t_{in} \sim Multinomial(\theta_{s_{in}})$.
   c) Based on topic t_{in}, draw the score $\widetilde{g_{in}} \sim Multinomial(\phi_{t_{in}})$.

For every score in a travel note, it will be assigned to the spots involved in this note or the global spot. So for the scores representing the background descriptions, which show in almost every note, will more likely to be assigned to the global spot so that other spots can get more reasonable and accurate scores.

By applying the Gibbs sampling to estimate the parameters θ_s and ϕ_t, we can get the spot-topic proportions $\Theta = \theta_{1:K}$ and topic-score proportions $\Phi = \phi_{1:T}$. Based on total probability formula, the spot-score proportions can be calculated as follows:

$$\Pi = \Theta * \Phi \tag{2}$$

where

$$\Pi \in R^{K*V} \;\; \Theta \in R^{K*T} \;\; \Phi \in R^{T*V} \tag{3}$$

The sentiment score of spot k is defined as :

$$sen_k = \sum_{v=1}^{V} \widetilde{g_v} * \pi_{kv} \tag{4}$$

where $\widetilde{g_v}$ is the v-th discrete score,

$$\pi_{kv} = p(\widetilde{g} = \widetilde{g_v} | s = s_k) \tag{5}$$

D. Popularity Adjustment

As we have got a sentiment score of each spot by ST-LDA, we adjust the score according to the popularity of each spot to satisfy the taste of different tourists. Typically, we divide the tourists into two type. One type of them prefers the popular spots like the landmark of a city and most of them are common tourists who come to a tourist city for the first time(e.g., The new-comers to Beijing always visit The Forbidden City first). The other type of tourist may have visited those famous spots before, like local people and someone who have been to the tourist city for several times. So they are more likely to prefer those spots which are little-known but still worth visiting.

To measure the popularity factor of spot k, which denoted as pop_k, we count the occurrences of spot k and denote it as occ_k. Because of the wide gap between the occurrences of popular spots and unpopular spots (e.g., The Forbidden City occurs 1539 times and some of other spots only occur one time), we use logarithmic functions to compress the occurrences and the popularity score pop_k is calculated as follows:

$$pop_k = \log(occ_k + \tau), \;\; where \;\; \tau > 0 \tag{6}$$

By using a logarithmic function with parameter τ, we can adjust the weight of popularity. When τ is big, the popularity factors of popular spots and unpopular spots are very close which means we hardly consider about popularity. When τ is small, the popular spot will gain a much higher score than the unpopular spots which means we consider a lot about the popularity. In short, lower the τ is, higher the weight of popularity is, and vice versa.

At last, we multiply the sentiment score of a spot by the popularity factor and normalize all the scores in the range 0 through 1. So the normalized synthetical score of spots k is given by:

$$NScore_k = \frac{Score_k - \min_{i \in 1,2,\cdots,K} Score_i}{\max_{i \in 1,2,\cdots,K} Score_i - \min_{i \in 1,2,\cdots,K} Score_i} \tag{7}$$

where

$$Score_k = sen_k * pop_k \tag{8}$$

TABLE II
PROPERTIES OF DATASET

Number of travel notes	6819
Number of sub-notes	34631
Number of scenic spots(Occur in the travel notes)	712
Total occurrences of all spots	25100
Average number of spots in a travel note	3.681

III. EXPERIMENTS

In this section, we introduce the data set used for experiments and compare our results with existing rating systems. We evaluate our results in both quantitative and qualitative ways.

A. Dataset

Baidu[3] and Dianping[4] are two famous and comprehensive websites in China which contains massive and multi-dimension tourism data. We crawled the travel notes and spots list from Baidu for experiments. Besides, we crawled two rankings of spots from Baidu and Dianping to evaluate the ranking results of ASSR. In this paper, we choose Beijing (the capital city of China, also a very popular tourist city) as the example and all the tourism data are associated with Beijing. The properties of our dataset are summarized in Table II.

B. Dataset Preprocessing

1) **Cleaning the text.** Delete strange symbols and messy codes in the text brought by the Chinese character coding problems to make sure the text can be processed.
2) **Filtering the useless travel notes.** Some travel notes mention Beijing but actually record a journey in other city. So we remove this kind of travel notes which is useless for us. And this makes the number of travel notes reduce from 7049 to 6819.
3) **Simplifying the name of spots.** Tourists usually mention the spots in short name while some of the spots in our spots list are using full name. By simplifying some spots' name, the occurrences of all spots increase from 24301 to 25100.

C. Evaluation and Analysis

Our model is to rate the spots based on the information mining from travel notes and generate synthetical scores for all spots according to people's different demands. By adjusting the parameter associated with the popularity of spots, we are able to rate the spots in different ways.

Here, we concentrate on two typical kinds of tourists. For the common tourists who prefer the famous and popular spots like landmarks of a city, we increase the weight of popularity factor and generate a ranking named Popularity-Preferred-Ranking(PPR) by sorting the spots' synthetical scores. To satisfy the people who only concentrate on the quality of the spot and ignore whether the spot is popular or not, we reduce the weight of popularity factor and then generate a Quality-Preferred-Ranking(QPR). To evaluate the results of our model, we conduct quantitative analysis on PPR and qualitative analysis on QPR.

*1) **Quantitative Analysis on PPR**:* To evaluate the effectiveness of PPR, we quantitatively compare the PPR with Baidu ranking and Dianping ranking, two existing rankings from the web.

To measure the similarity between two rankings, we introduce two evaluation metrics, Fitness and Kendall tau distance.

Fitness is defined:

$$F = \frac{1}{Z}\sum_{j}\frac{\omega_j}{\alpha(|p_j - q_j|) + (1-\alpha)p_j} \tag{9}$$

where Z is the normalization factor, p_j denotes the j-th item's position in one ranking while q_j denotes the position in another. ω_j represent the weight of this item and we set it as a same value for all spots.

Kendall tau distance can be calculated by using following equation:

$$K = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j<i}k(i,j) \tag{10}$$

where

$$k(i,j) = \begin{cases} 0 & \text{(if i,j is in same order)} \\ 1 & \text{(if i,j is not in same order)} \end{cases} \tag{11}$$

Both Fitness and Kendall tau distance are real numbers between 0 and 1. For Fitness, the higher values indicate the higher similarity between two rankings. But for Kendall tau distance, it measures one kind of distance between two rankings, so the lower the distance is, the more similar two rankings are.

To determine the proper popularity factor corresponding to PPR, we compare our result with two existing rankings under different parameters. Here, we use the intersection of these three rankings as the common spot set, which contains 172 spots. Then we adjust the parameter $\tau = 1 : 300$ to control the weight of popularity factor and get different rankings on the common spot set. The similarity between our result and the two existing rankings is shown in the Figure 2.

As illustrated, when τ is small, which means a high weight for popularity factor, the value of Fitness becomes larger while the value of Kendal tau distance becomes smaller. This indicates that when the weight of popularity factor is high, we can generate a spot ranking relatively similar to the existing rankings. Typically, we set $\tau = 10$ and generate the ranking as PPR based on our model.

Table III demonstrates the similarity of three rankings between each other. Three Fitness values are very close to each other and relatively high in a range of 0 to 1. The three Kendall tau distance values are also very close and relatively low. The values of two indicators prove that the similarity among these three rankings is almost in the same and relatively high level. Considering that Baidu ranking and Dianping ranking are two existing rankings provided by two large comprehensive

[3]https://www.baidu.com/
[4]http://www.dianping.com/

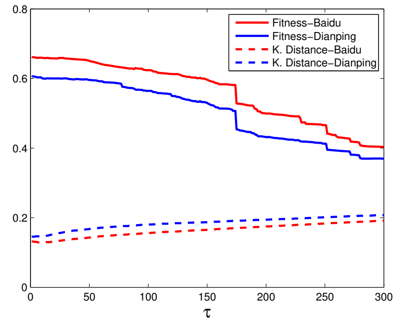

Fig. 2. Fitness and Kendall tau distance results for the similarity comparison between our rankings and existing rankings, Baidu Ranking and Dianping Ranking. τ indicates the weight of popularity factor, higher τ means lower weight of popularity factor.

TABLE III
SIMILARITY AMONG PPR, BAIDU AND DIANPING RANKINGS

Indicator	Fitness	Kendall tau distance
PPR vs Dianping	0.6004	0.1465
PPR vs Baidu	0.6588	0.1278
Dianping vs Baidu	0.6579	0.1197

websites, the high similarity between PPR and the existing rankings can prove that the PPR is reasonable and effective for the common tourists.

*2) **Qualitative Analysis on QPR**:* To generate QPR which cares more about the quality information of spots, we should decrease the weight of popularity by increasing the parameter τ. As illustrated in Figure 2, with the increase of τ, our spot ranking becomes more and more different from the existing rankings and the differences exactly show that we find some other information about the spots which is sometimes covered by the popularity. Since our QPR has different target from existing rankings, we can't evaluate QPR based on its similarity to existing rankings. So we discuss the differences in detail and provide a qualitative analysis.

Here, we set $\tau = 150$ to generate the ranking as QPR and list some typical popular spots and obscure spots in the top of QPR. Moreover, we compare the ranks of these spots in QPR and the two existing rankings. The comparison results among three rankings are illustrated in Table IV.

Firstly, we can find some very famous and popular spots rank highly in QPR, like "The Forbidden City", "Temple of Heaven", "Beihai Park", while these spots also have a similar and high rank in both Baidu ranking and Dianping ranking. So for these spots with both high popularity and high quality, QPR can give them a high rank which is similar to the existing rankings.

What's more important, some less well-known spots also have a high rank in QPR, which are exactly the obscure spots we want. Typically we find "Fortuna's park", ""QianChi Pearl Fall", "YuDu Mountain", "OuFeiBao Chateau", "KaiYueLai Hot Spring" and "Qianling Mountain" in top 20 of the QPR.

TABLE IV
SPOTS RANKING COMPARISON AMONG QPR, BAIDU, DIANPING

Spot	QPR	Baidu	Dianping
The Forbidden City	1	1	1
Temple of Heaven	7	9	6
Beihai Park	13	10	15
Fortuna's park	5	935	644
QianChi Pearl Fall	6	-	-
YuDu Mountain	11	-	41
OuFeiBao Chateau	15	-	-
KaiYueLai Hot Spring	19	-	-
Qianling Mountain	20	398	37

"-" means the spot is not included in this ranking.
First three spots are typical popular spots and the rest of spots are less well known.

Although two of them, "Yudu Mountain" and "Qianling Mountain", have an acceptable rank in Dianping ranking, many of other obscure spots are not included in the two existing rankings or have a very low rank, which means most of these unpopular spots are ignored by existing rankings.

To identify whether these obscure spots in QPR worth visiting, we search the related information of these spots on the Internet as much as possible. Except for the fine introduction and beautiful photograph given by the official website, we also discover a few comments on these unpopular spots from different kinds of tourism websites. For example, one tourist comment on "YuDu Mountain" like "I come to YuDu Mountain every year, and I think it's the most beautiful place in Beijing suburb...". To summarize, most of these comments are positive and show that the obscure spots we find have a relatively fine review by most of the tourist who have been there. According to the information we find, we believe that the obscure spots we find are actually unpopular and worth visiting.

IV. RELATED WORK

A. Obscure Scenic Spots Discovering

In the field of research, most existing work about tourism data mining focus on two types of data, i.e., photograph with geo-tag or other labels [9], and semantic description. For example, [10] using a large-scale geo-tagged web photo to suggest tourist destinations with the input of a photo or a keyword describing the place.

Most especially, [11] had done a very similar work with us which develop a system named ANABA to discover the obscure scenic spot. And this system also used geo-tagged images as the dataset and the system mainly calculate the obscurity score based on the social appreciation and the content of images.

In our rating system ASSR, we choose travel notes as dataset. Travel notes contain a lot of opinions and comments towards the scenic spots, and are playing a more and more important role in tourists decision-making [12]. By mining information from numerous travel notes, we can discover many obscure spots.

Actually, we are using definitely different dataset and method to solve the same problem and we can try to combine them together and find better results of this problem.

B. Author Topic Model

LDA is widely used in text mining, especially to find the distribution of words over the latent topic. There are lots of models extends LDA and focusing on different problems [13], [14]. Author-Topic Model is also a generative model that extends LDA [15]. The difference between Author Topic model and LDA is that Author Topic Model conclude the authorship of a document. While LDA can only find the relationship between words and topics, the Author Topic model can discover the relationship among the author, topic, and words.

The algorithm ST-LDA in our model is quite similar to Author Topic Model while we regard the spots as the authors of travel notes and use the scores to describe the document. And the main improvement of ST-LDA is that we introduce a global spot to filter the scores of background descriptions to make our result more accuracy.

V. Conclusion

The ratings of spots are useful information for tourists to evaluate spots and make travel plans. However, the existing rankings provided by various tourism platform are static and highly related to the popularity of the spot. To solve this problem, we proposed our ASSR to rate the scenic spot automatically based on numerous travel notes. The result of our model can be adjusted by the popularity of the spot to satisfy the demands of different tourists. Typically, we provide two rankings named PPR and QPR for tourists with different demands. PPR is designed for the common tourists who prefer the famous and popular spots. QPR is designed for some particular tourists who only concentrate on the quality of the spot and even prefer the spots which are less well known but worth visiting.

The experiments on the web data sets have provided the result of PPR and QPR among Beijing Spots. We verified the effectiveness of PPR by calculating the similarity between PPR and real world rankings of the spots. As for QPR, we have conducted qualitative analyses on the obscure spots we have found. Although it's hard to strictly prove the correctness of these spots now because they are rarely known by common people, we will monitor the popularity of these obscure spots continuously. If these spots become more popular in the future, it can be a strong evidence to prove the effectiveness of our model. Also, we have published our QPR in the Smart Tourism Services Platform (STSP) and we will collect the feedback from the users who visit those obscure spots for further proof.

Besides finding ways to further prove the effectiveness of our model, we also consider introducing more adjustable factors into our model like environment, price, traffic, type of the spots, etc. By synthesizing different factors together in a proper way, we can acquire more personalized rankings of spots based on different preference and demands of tourists.

Acknowledgment

This research has been partially supported by the National Nature Science Foundation of China (No.61673230) and High-Tech Ship Research Project of China (17GC26102.01).

References

[1] L. Ardissono, A. Goy, G. Petrone, M. Segnan, and P. Torasso, "Intrigue: Personalized recommendation of tourist attractions for desktop and hand held devices," 2003, pp. 687–714.

[2] L. Sebastia, I. Garcia, E. Onaindia, and C. Guzman, "e-tourism: A tourist recommendation and planning application," in *IEEE International Conference on TOOLS with Artificial Intelligence*, 2008, pp. 89–96.

[3] I. Olmeda and P. J. S. PhD, "Data mining techniques and applications for tourism internet marketing," *Journal of Travel & Tourism Marketing*, vol. 11, no. 2-3, pp. 1–20, 2002.

[4] M. Banyai and T. D. Glover, "Evaluating research methods on travel blogs," *Journal of Travel Research*, vol. 51, no. 3, pp. 267–277, 2012.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[6] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov, "Semeval-2016 task 4: Sentiment analysis in twitter," in *International Workshop on Semantic Evaluation*, 2016, pp. 1–18.

[7] F. Bravo-Marquez, M. Mendoza, and B. Poblete, "Meta-level sentiment models for big social data analysis," *Knowledge-Based Systems*, vol. 69, pp. 86–99, 2014.

[8] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016.

[9] J. Luo, D. Joshi, J. Yu, and A. Gallagher, "Geotagging in multimedia and computer vision–a survey," *Multimedia Tools & Applications*, vol. 51, no. 1, pp. 187–211, 2011.

[10] L. Cao, J. Luo, A. Gallagher, X. Jin, J. Han, and T. S. Huang, "Aworldwide tourism recommendation system based on geotaggedweb photos," vol. 23, no. 3, pp. 2274–2277, 2010.

[11] C. Zhuang, Q. Ma, X. Liang, and M. Yoshikawa, "Anaba: An obscure sightseeing spots discovering system," in *IEEE International Conference on Multimedia and Expo*, 2014, pp. 1–6.

[12] L. L. Guo and Z. Y. Qin, "Research on tourism destination image perception of rizhao based on network travel notes," *Journal of Qufu Normal University*, 2016.

[13] B. Bai, Y. Fan, W. Tan, and J. Zhang, "Sr-lda: Mining effective representations for generating service ecosystem knowledge maps," in *IEEE International Conference on Services Computing*, 2017, pp. 124–131.

[14] Z. Gao, Y. Fan, C. Wu, W. Tan, J. Zhang, Y. Ni, B. Bai, and S. Chen, "Seco-lda: Mining service co-occurrence topics for recommendation," in *IEEE International Conference on Web Services*, 2016, pp. 25–32.

[15] R. Z. Uc, M. Rosenzvi, and T. Griffiths, "The author-topic model for authors and documents," pp. 487–494, 2012.