Travel and Tourism Research Association: Advancing Tourism Research Globally

2017 ttra International Conference

Destination advertising in a smarter way: A machine learning model for DMOs' photo selection

Ning Deng

Xiang (Robert) Li
Temple University

Destination advertising in a smarter way:
A machine learning model for DMOs' photo selection

Introduction

The importance of tourism destination image (TDI) in destination marketing has been well documented. The classic conceptualization divides TDI into the projected image and received image, with the former referring to the image being portrayed to the target market, and the latter being the image reflected in tourists' (or any receivers') mind (Bramwell and Rawding 1996). Previous studies have largely focused on (potential) tourists' received image; when studying projected image, emphases tend to be given to images projected by destination marketing organizations (DMOs). Since the emergence of Web 2.0 applications, more tourism stakeholders, such as tourists and destination residents, have been involved in portraying TDI. Put differently, both the supply and demand sides are now engaging in image projection these days and no entity is solely at the receiving end of TDI anymore. As such, the line between projected and received images is getting blurred. This paper will focus on analyzing user-generated destination photos to assist DMOs in image projection.

"An image is worth a thousand words." Photos, or pictorial images, are very important carriers for TDI communication. In the past, DMOs were at the forefront of projecting destination visual images. These days, people tend to publish interesting photos, and the contents shared by relatives or friends are usually more likely to be viewed and trusted; thus, User Generated Content (UGC), e.g., destination photos taken and shared by the general public, have increasingly become an effective vehicle to form TDI for potential visitors (Lo, McKercher et al. 2011). For DMO interested in visual advertising, a good starting point is to understand what pictorial images have been portrayed and how viewers have received and reacted to those pictorial images. For this purpose, this study proposes a machine learning based model to select photo elements from the viewers' perspective and assist DMOs' photo selection.

Literature Review

Projected and received image

TDI may be divided into the supply-side projected image and the demand-side received image (Baloglu and McCleary 1999, Tasci and Gartner 2007). Traditionally, projected image usually appears in the form of tourism advertisement (e.g. tourist brochures and videos for commercial purpose). With the emergence of UGC, messages and photos on the social network also possess the attribute of promulgating the TDI unintentionally. Comparing to projected image, received image has drawn much more research attention. Received image is a mental construct or representation of a destination formed through the interaction between the projected image, personal characteristics, and external stimuli (Baloglu and McCleary 1999).

In today's technological environment, more parties, notably tourists and destination residents, are all actively involved in the process of image projection and sharing. Using photo sharing as an example, viewers themselves are constantly also publishers (Figure 1); thus, the projected online image of a destination agglomerates both DMO-generated and user-generated images. Further, the online and social media not only provide a platform for projecting images, but also publicize received images, which allows DMOs and researchers immediate access to analyze received images. The interactive nature of this process provides conceptual foundation for the method proposed in the present study.

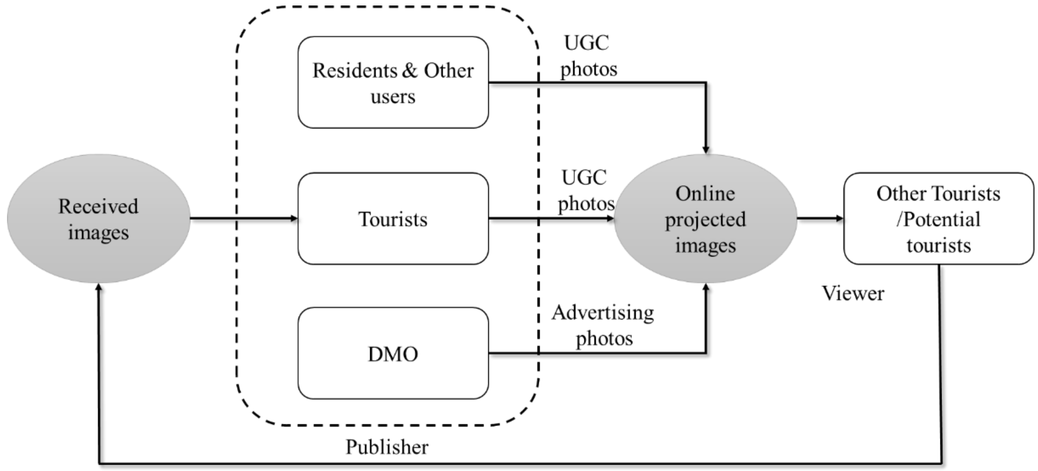

Figure 1. A conceptual model of pictorial TDI projection and receiving in the online world

Photo-induced TDI

Previous studies (Baloglu and McCleary 1999, Kim and Richardson 2003) suggest that affective image tends to be more impactful on building TDI than cognitive image. Pan, Lee et al. (2014) is one of the few studies that discusses the relationship between image dimensions and the affective qualities of destinations. By analyzing 145 photos and captions published in the New York Times's Travel Section ("Why we Travel") from 2008 to 2012, a correlation model between 9 cognitive dimensions and 8 affective categories was constructed. Similarly, we also believe that a deeper understanding of "what cognitive image can induce which emotion" is the premise of the TDI projection. However, we derived the correlation model through a large-scale dataset and used a machine learning model to summarize the relationship between photo contents and emotions.

The sentiment analysis of photos

In the field of information technology, images' sentiment analysis has also drawn great attention (Chen, Borth et al. 2014, Chen, Chen et al. 2014, Machajdik and Hanbury 2010). Unlike the conventional approaches in social science, e.g. the content and semiotic analysis, researchers in the computer science field analyze the content and sentiment with the aid of computers.

Earlier studies (Machajdik and Hanbury 2010, Wang and He 2008) attempted to map low-level visual features such as color, texture, objects, facial expressions etc., with high-level image sentiments, feelings and emotions. However, the semantic gap (Wang and He 2008) between the low-level features and high-level emotions is still an obstacle for visual content sentiment analysis. Recently, some researcher have begun to use mid-level representations, e.g. texts and metadata, to predict the visual content sentiments. In general, photos from social network contain abundant information describing themselves, namely, the metadata. Some interesting information, such as the title, tags, and description information, is of great significance for content analysis (Hollenstein and Purves 2010). Meanwhile, photos on social network usually allow viewers to leave comments, which are ideal materials for sentiment analysis.

Methodology

Flickr dataset—YMCC 100M

Photos taken by digital devices usually carry some information describing themselves, namely the metadata. In July, 2015, Yahoo released a visual content dataset for researchers named "Yahoo Flickr Creative Commons 100M" (YFCC 100M). It contains over 100 million multimedia metadata published on Flickr from 2004 to 2014, and will be used in this study.

Target city—New York City

New York City (NYC) is selected as a case study. There is no structured circle boundary in NYC and we restricted the analysis range in this study to be a rectangle, within longitude: [-73.7125, -74.0991], latitude: [40.5854, 41.8688]. The three most important boroughs of NYC, Brooklyn, Queens, Manhattan, are included in this range, which also covers most of the famous attractions in NYC.

In the data cleaning process, we filtered the metadata in YFCC 100M by the preset geographic boundary. A clean dataset of NYC containing 192,677 data items was captured after the data cleaning process. Each item contains the photo related information, including the coordinate, tag, title, description, and time that the photo was taken. We visualized the items by Carto, which is an open, powerful, and interactive ARC GIS system. As shown in Figure 2, every single yellow point represents a photo taken at that location; the deeper the color is, the more photos were taken at that region.

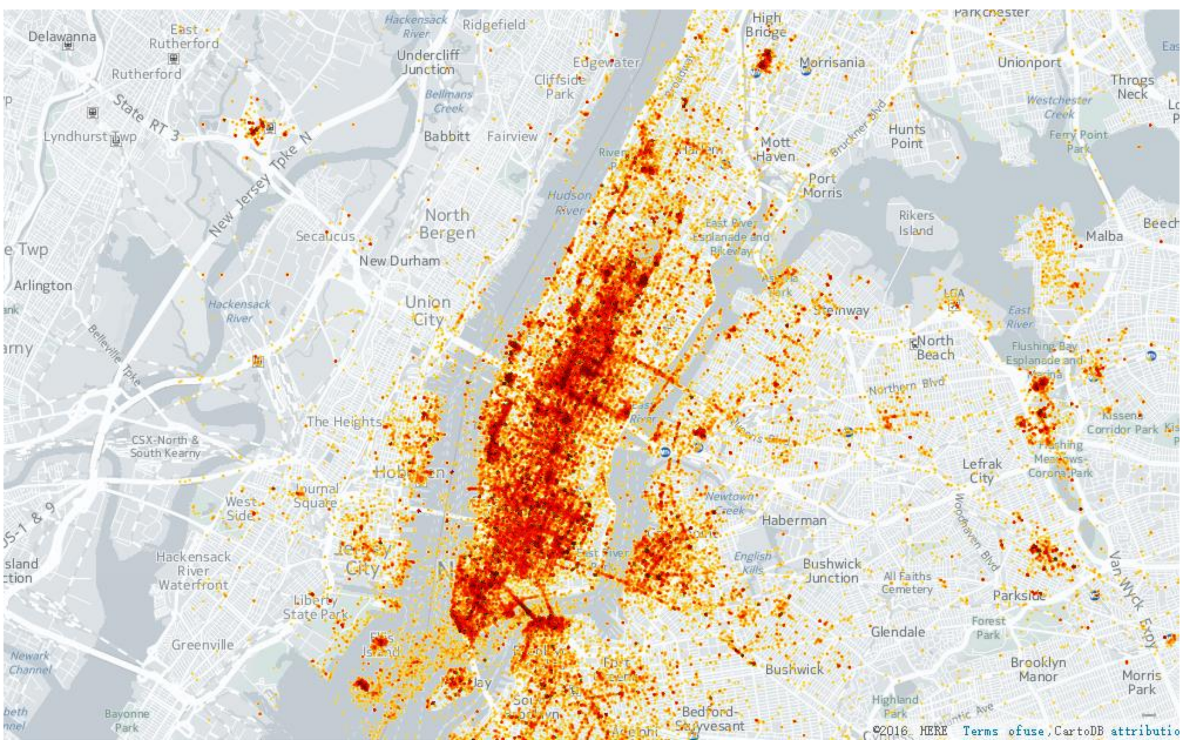

Figure 2. The Flickr photo distribution in NYC

4.4 The PCC-VAC model

PCC: In this study, we identified a set of Publisher Cognitive Concept (PCC) to represent popular items mentioned in the title, tag, and fields. A publisher is encouraged to leave information describing the photo, including the place where it was taken, people, attraction, building, and objects inside the photo. All these things can be considered as part of the publisher's own received cognitive image of the destination. Thus, PCC can be considered as a set of keywords to describe photo contents and summarize the cognitive image of destination.

VAC: A set named viewer affective concept (VAC) is defined, representing the emotions embedded in the comments. It is noted that, emotions are more likely to be expressed by adjectives, rather than nouns or verbs. Comments on photos can be considered as the attitudes and emotions a viewer expressed toward a photo. We utilize the Flickr API to achieve the comments of 192,677 photos from Flickr. Among these, a total of 15,606 photos possesses at least one comment, about 0.08 comments per photo on average.

The problem of selecting photos eliciting target emotions is, in the nutshell, to classify photos into the most probable affection categories. This can be translated into a classification problem in the machine learning field (Machajdik and Hanbury 2010). By analyzing a large-scale Flickr dataset, the model is trained by numerous "content-emotion" pairs. Then, the emotion of a new photo can be predicted according to its content. On the contrary, for a certain affect, we can also calculate each photo's probability of stimulating that emotion. All the candidate photos are ranked in a descending order.

As one of the most classical algorithms in classification problems, Naïve Bayesian Classifier (Pang, Lee et al. 2002) is widely adopted as a supervised learning method in classification problems. Briefly, naïve Bayes is a conditional probability based model, given a problem instance to be classified, represented by a vector $x=(x_1,...,x_n)$, representing some n independent variables, it assigns to this instance probabilities $p(C_k \mid x_1,...,x_n)$, for each of K possible outcomes or classes C_k. In the Naïve Bayes model, the result of the classification is labeled as $\hat{y}=C_K$ for some k as follows:

$$\hat{y}=\underset{k\in\{1,...,k\}}{\arg\max}\, p(C_k)\prod_{i=1}^{n} p(x_i \mid C_k)$$

Quantitatively, the basic rule of the Naïve Bayes Classifier is to classify the particular instance into a class that is most probable, according to its conditional probability.

As for photo recommendation, the result is actually the probability of $P(di \mid vj,\theta)$, where θ is the training photo set. For each $di\in\theta$, the larger $P(di \mid vj)$ is, the more probable that the affect v_j could be stimulated by image di. The conditional probability $P(p_k \mid v_j)$ can be determined by:

$$P(pk \mid vj;\theta)=\frac{\sum_{i=1}^{|D|} B_{ik} P(vj \mid di)}{\sum_{i=1}^{|D|} P(vj \mid di)},$$

where B_{ik} is a variable to indicate the presence/absence of P_k in the publishers' metadata of image d_i, and $|D|$ is the number of images. $P(v_j|d_i)$ is calculated by:

$$P(vj \mid di)=\frac{the\ occurrence\ counting\ of\ v_j}{the\ words\ counting\ of\ d_i\ comment}.$$

Based on the correlations $P(pk \mid vj;\theta)$, we can measure the likelihood of a photo d_i and an affect v_j by multivariate Bernoulli formulation (McCallum and Nigam 1998).

$$P(di \mid vj;\theta) = \prod_{k=1}^{|A|} (P(pk \mid di) * P(pk \mid vj;\theta) + (1 - P(pk \mid di)) * (1 - P(pk \mid vj;\theta))) \ ,$$

where |A| is the set of PCC achieved in next section. Thus, a candidate photo set $\{d_1, ..., d_i\}$ with an affect v_j could be ranked by their results of $P(di \mid vj)$, and the outcome of this model is the recommended photo sequence to illustrate the emotion v_j.

Results

PCC of NYC

To retrieve *PCC* from the *NYC* related metadata, we utilize Textblob (Loria 2014) to calculate the occurrence frequencies of nouns appeared in title, tags, and description fields. To identify popular items, a threshold of 1,000 is set to identify the PCC from all the NYC related items. Eventually, a PCC of 234 nouns is achieved, with the top 50 keywords listed in Table 1. The top 5 keywords are "brooklyn", "avenue", "heights", "yards", "barclays" respectively.

Table 1. The top 50 PCC keywords mined from 192,677 NYC related metadata items.

PCC	Count	PCC	Count
brooklyn	185614	island	11945
avenue	130132	river	10817
heights	101025	queens	10505
yards	84950	world	9870
barclays	84281	photography	9373
manhattan	81939	newyorkcity	9320
square	80880	canon	8944
street	59470	party	8538
museum	58662	stadium	8206
center	46498	hours	7662
ratner	41240	village	7569
states	35995	jersey	7431
times	31476	subway	7264
people	29510	camera	6753
music	24190	color	6752
building	23429	church	6691
newyork	19231	hudson	6627
bridge	18427	yamamoto	6246
garden	17979	station	6123
iphoneography	17861	statue	6000
format	15263	parade	5807
night	15109	gothamist	5783
photos	13913	landmark	5377
photo	13905	metropolitan	5212
state	12268	prospect	5062

VAC of NYC

To ensure that we can get the most representative affective adjectives, we focus on two factors when selecting the VAC items: the occurrence frequency and the sentiment strength. Textblob is widely adopted in the field of natural language processing, which is applied in our study to do part-of-speech tagging and to extract adjectives. Further, an embedded sentiment analysis tool, SentiWordNet, is adopted to measure the sentiment values of adjectives. The sentiment result ranges from -1 (negative sentiment) to +1 (positive sentiment). Eventually, we only kept adjectives with an occurrence frequency > 30 times, and an absolute value of sentiment strength > 0.1 in VAC. The result of VAC is shown in Table 2.

Table 2. VAC elements extracted from Flickr comments

Adjective	frequency	Adjective	frequency
great	16578	sweet	120
beautiful	2302	favorite	116
awesome	948	black	114
wonderful	513	funny	96
little	428	least	90
right	420	small	73
amazing	361	whole	60
interesting	351	super	58
other	320	green	57
pretty	278	wrong	46
excellent	245	sharp	45
fantastic	229	delicious	43
happy	200	lucky	38
first	197	large	37
light	165	special	35
perfect	157	original	34
better	156	weird	32
gorgeous	150	fabulous	31
lovely	146	crazy	31

Ranking of the candidate photos

In order to evaluate the algorithm's ability of ranking the photos in a viewer-friendly way, we randomly select five photos from Flickr that are relevant to the affective adjectives v_j: "beautiful", "delicious" and "favorite". Once a $v_j \in VAC$ can be found in a photo's comments, the photo is considered a relevant image to the affective concept v_j. The candidate photos are calculated using the PCC-VAC model and are ordered according to their relevancies to the themes. Further, we downloaded the original image files from the website using the Flickr API, and the photo contents are shown in Figure 3. As shown in the figure, photos containing the feeling of "beautiful" in New York mainly include buildings, however, the people with air balloons in the second photo is also considered as relevant. As for the feeling of "delicious", the photo contents are highly related to foods of New York, including the bar, sandwich, orange etc. Photo contents of "favorite" are much more diversified, the body, nature, activities and architecture are all considered to be relevant. The ordered photo contents provide some references to DMOs, who could select photos from their own gallery with the similar contents for marketing purpose.

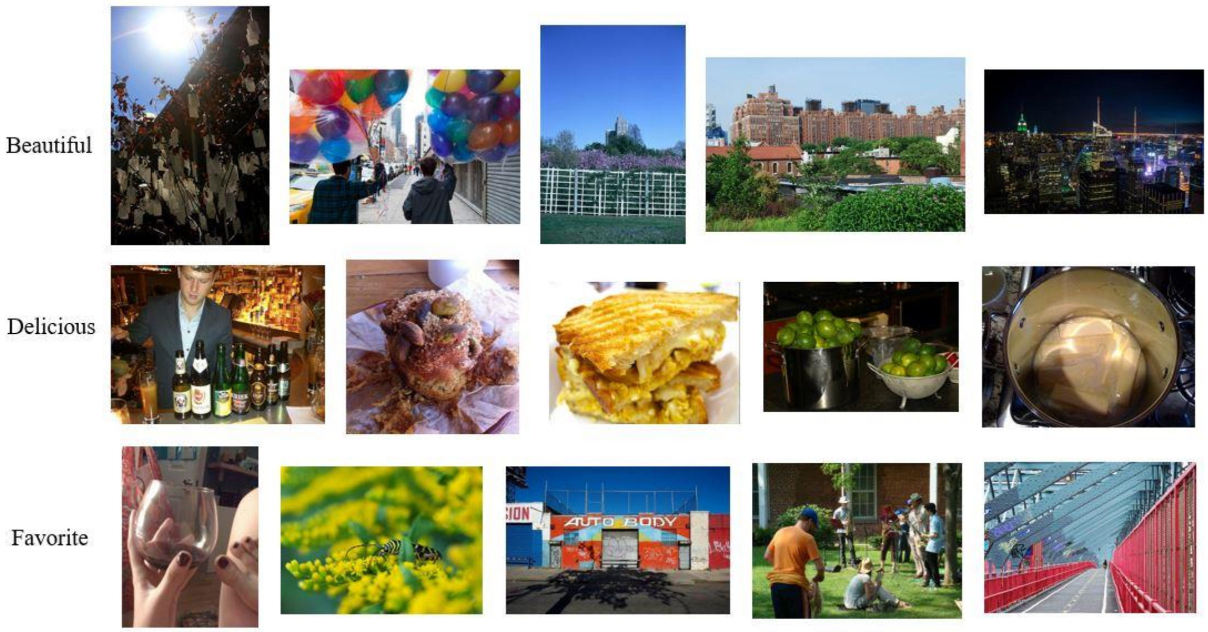

Figure 3. The ordered photos of selected affective adjectives in New York

Conclusion and Discussion

Social network has become a mainstream channel for TDI promotion, and visual contents have become the most effective vehicle for image promotion. Recently, UGC from tourists has played a more significant role in the TDI projection process. However, photo selection in marketing activities is still mainly based on DMO’s subjective judgement, largely ignoring the view of the ordinary people. This paper proposes a machine learning based model to assist the DMO with photo content selection. The proposed method utilizes a data mining technique to mine two crucial corpus, namely the PCC and VAC, from photo social network website—Flickr. This method enables the recommendation of the photos that can best stimulate an emotion. We take NYC as a case study to demonstrate the rationality and effectiveness of our approach.

Despite of its limitations, this study contributes to the literature by considering the influence from UGC in TDI projection and extending the classical “circle of representation” proposed by Urry (2002). Also, for DMOs, the proposed model could be instrumental in identifying the most suitable contents when publishing marketing photos, addressing the well-known gap between projected and received image.

References

Baloglu, S. and K. W. McCleary (1999). "A model of destination image formation." Annals of Tourism Research 26(4): 868-897.

Bramwell, B. and L. Rawding (1996). "Tourism marketing images of industrial cities." Annals of Tourism Research 23(1): 201-221.

Chen, T., et al. (2014). "Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks." arXiv preprint arXiv:1410.8586.

Chen, Y.-Y., et al. (2014). Predicting viewer affective comments based on image content in social media. Proceedings of International Conference on Multimedia Retrieval, ACM.

Hollenstein, L. and R. Purves (2010). "Exploring place through user-generated content: Using Flickr tags to describe city cores." Journal of Spatial Information Science 2010(1): 21-48.

Kim, H. and S. L. Richardson (2003). "Motion picture impacts on destination images." Annals of Tourism Research 30(1): 216-237.

Lo, I. S., et al. (2011). "Tourism and online photography." Tourism Management 32(4): 725-731.

Loria, S. (2014). "TextBlob: simplified text processing." Secondary TextBlob: Simplified Text Processing.

Machajdik, J. and A. Hanbury (2010). Affective image classification using features inspired by psychology and art theory. Proceedings of the 18th ACM international conference on Multimedia, ACM.

Machajdik, J. and A. Hanbury (2010). Affective image classification using features inspired by psychology and art theory. Proceedings of the international conference on Multimedia, ACM.

McCallum, A. and K. Nigam (1998). A comparison of event models for naive bayes text classification. AAAI-98 workshop on learning for text categorization, Citeseer.

Pan, S., et al. (2014). "Travel photos: Motivations, image dimensions, and affective qualities of places." Tourism Management 40: 59-69.

Pang, B., et al. (2002). Thumbs up?: sentiment classification using machine learning techniques. Empirical Methods in Natural Language Processing.

Tasci, A. D. and W. C. Gartner (2007). "Destination image and its functional relationships." Journal of Travel Research 45(4): 413-425.

Urry, J. (2002). The tourist gaze, Sage.

Wang, W. and Q. He (2008). A survey on emotional semantic image retrieval. ICIP.