Multimodal Spatio-Temporal Theme Modeling for Landmark Analysis

Weiqing Min
Institute of Automation, Chinese Academy of Sciences

Bing-Kun Bao and Changsheng Xu
Institute of Automation, Chinese Academy of Sciences, and China-Singapore Institute of Digital Media, Singapore

By distinguishing between temporal, local, and general landmark themes, the proposed model will help tourists easily identify desirable landmarks and the optimal time to visit them.

Travel has become increasingly popular in people's everyday lives, and tourism is flourishing across the globe. Landmarks are of great interest to tourists because their unique physical and historical characteristics often educate visitors about the location. Landmarks also reveal different charms at different moments. Thus, time-dependent characteristics are useful for tourists in making decisions about when to visit various landmarks.

Take the landmark Kiyomizu-dera in Japan as an example. In addition to the unique local characteristic ("temple"), two temporal characteristics are "cherry blossoms in spring" and "red maple leaves in autumn" (see Figure 1). If we can provide tourists with both local and temporal characteristics, they can easily decide what to visit and when, according to their own preferences. Therefore, it is highly desired to automatically mine and summarize landmarks' local and temporal characteristics to facilitate users' landmark browsing and trip planning.

Because they offer numerous landmark photos with comprehensive information (including associated tags and other context metadata), photo-sharing websites such as Flickr and Picasa are a good resource for landmark mining and summarization. The considerable existing work in this area includes landmark summary,[1] landmark recognition and classification,[2,3] landmark retrieval,[4] and landmark relevant tour recommendation.[5–7] However, almost all such efforts only consider local characteristics, while few have mined the landmarks' temporal characteristics.

Here, we discuss mining and summarizing landmarks' general themes as well as the local and temporal themes. General themes occur extensively in various landmarks, and include accommodations and other standard features. The local theme implies a specific theme that exists only at a certain landmark, such as a unique physical characteristic. The temporal theme corresponds to the location-time-representative pattern, which relates only to a certain landmark during a certain period—such as fleet week at the Golden Gate Bridge or red maple leaves in Kiyomizu-dera. Local themes are useful in landmark analysis for their discriminative and representative attributes. However, the ability to discover landmark diversity at different moments makes temporal themes equally important in landmark studies. Time-dependent diversity shows complete viewing angles over time and complements local themes in landmark understanding. Furthermore, it provides more comprehensive and structured information for landmark history browsing and tourist decision making.

Although general, local, and temporal themes exist in nearly all landmarks, discovering them is challenging as a result of the intrinsic issues caused by user-generated data. First, the associated tags and descriptions are inevitably noisy, and even the landmark photos themselves are often corrupted with occlusions such as tourists or other architecture. Thus, it is not easy to discover three kinds of themes from large, noisy user-generated data. Second, besides texture contexts associated with landmark photos, visual contexts can also benefit theme mining. Take

1070-986X/14/$31.00 © 2014 IEEE

Published by the IEEE Computer Society

Figure 1. The Kiyomizu-dera temple in Japan. The scenery from different seasons may appeal to different types of tourists and thus affect trip planning.

Kiyomizu-dera in autumn as an example: texture keywords corresponding to the temporal theme may be "red," "maple," and "leaves," while the visual part could be patches of the red maple. Thus, it is necessary to consider multimodal contexts together in theme mining and summarization. The third issue is how to reveal time-relevant themes in dynamic and diverse landmark samples. For temporal theme discovery, the proposed method should be able to reserve the shared characteristics in a certain landmark corresponding to local themes and to detect the landmark's discriminative patterns over time.

To tackle these challenges, we propose a probabilistic topic model called Multimodal Spatio-Temporal Theme Modeling (mmSTTM). The model considers both textual and visual contexts to learn general, local, and temporal themes, which span a low-dimensional theme space. The model also assigns all textual and visual keywords to each theme, along with a probability for each; a keyword with high weight assignment is meaningful for the theme, while low-weighted keywords are considered noise. (See the related work sidebar for earlier work in this field.)

As Figure 2 shows, based on the proposed mmSTTM, we developed a framework with three modules: data preparation, theme modeling, and theme analysis. In the first module, we download the landmark photos from Flickr and preprocess them. Next, in theme modeling, we use mmSTTM to discover the three theme types. Finally, we analyze the discovered themes, including landmark-specific time distribution and time-specific location distribution.

Data Preparation

To build the landmark dataset for mining and analyzing landmark themes, we first select a list of landmarks and then collect photos for each one. Because landmarks are places that tourists typically go for sightseeing, we chose the most-visited countries from a Wikipedia webpage (http://en.wikipedia.org/wiki/Tourism). For each country, we then visited tourism websites, such as Yahoo Travel Guide, and manually select famous landmarks. This gave us our initial landmark list. To ensure our list contained the most famous landmarks, we assumed that if numerous landmark photos were uploaded to photo-sharing websites such as Flickr, the landmark was probably famous. We therefore further filtered and selected landmarks using the Flickr API (www.flickr.com/services/api/explore/flickr.photos.search) and then chose the landmarks for which the number of returned photos is higher than a particular threshold (such as 200,000). This gave us our final list of landmarks.

Based on the selected landmark list, we collected the corresponding photo set for each landmark, focusing on downloading landmark photos from Flickr. For each landmark, we crawled the photos and associated metadata (tags, title, description, time-taken information, and geotags). If we had simply selected the

Related Work in Landmark Mining and Geographic Topic Modeling

Our work is closely related to both landmark mining and geographical topic modeling.

Landmark Mining

Lyndon S. Kennedy and Mor Namaan used both context and content information—including tags, geotags, and visual features—to summarize representative views of landmarks from Flickr.[1] In contrast, Rongrong Ji and his colleagues focused on mining landmarks from blogs using graph modeling.[2,3] Yan-Tao Zheng and his colleagues were dedicated to landmark recognition in large-scale landmark collections,[4] while Yannis Avrithis and his colleagues proposed a photo clustering scheme—that is, vector quantization—to depict different views of one location for further photo retrieval.[5] For image retrieval using a San Francisco landmark dataset, Shaoting Zhang and his colleagues proposed a graph-based,[6] query-specific fusion method.[6]

Several research projects focus on mining landmarks for developing tourism-relevant applications. One project developed a travel recommendation system, W2Go, by automatically recognizing and ranking landmarks from Flickr.[8] Another effort mainly used the geotag information of photos to generate travel routes for trip planning.[9] Almost all of this work focuses only on the landmarks' local characteristics and ignores their temporal characteristics.

Geographic Topic Modeling

In recent years, researchers have successfully applied probabilistic topic models such as probabilistic latent semantic analysis (PLSA)[10] and latent Dirichlet allocation (LDA)[11] to discover latent themes. To discover themes from geographical regions, topic model variations are developed by incorporating location information.[12,13] Zhijun Yin and his colleagues presented a joint model that combines location and text for discovering geographical themes and comparing themes across different geographical locations.[13] Based on earlier work,[14] Yanwei Pang and his colleagues assumed that geographical location involves both general and local themes, where general terms are grouped into general themes, while terms related to special locations belong to local themes.[12]

Our work relates more to this latter idea, but it differs in two ways. First, we introduce the temporal themes into the topic model and mine three kinds of themes simultaneously. Second, we introduce a complete framework for not only mining themes, but also deeply analyzing their spatio-temporal properties. Earlier work[2] considered both location and time information to discover theme patterns. However, in contrast to that work, we mine three kinds of themes simultaneously, rather than focus only on temporal topic mining and analysis. Also, our framework incorporates both textual and visual photo information to mine themes, whereas previous research conducted theme analysis based on textual information.[2]

References

1. L. Kennedy and M. Naaman, "Generating Diverse and Representative Image Search Results for Landmarks," *Proc. 17th Int'l Conf. World Wide Web*, 2008, pp. 297–306.
2. Q. Mei et al., "A Probabilistic Approach to Spatiotemporal Theme Pattern Mining on Weblogs," *Proc. Int'l Conf. World Wide Web*, 2006, pp. 533–542.
3. B. Bao et al., "General Subspace Learning With Corrupted Training Data Via Graph Embedding," *IEEE Trans. Image Processing*, vol. 22, no. 11, 2013, pp. 4380–4393.
4. Y. Zheng et al., "Tour the World: Building a Webscale Landmark Recognition Engine," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2009, pp. 1085–1092.
5. Y. Avrithis et al., "Retrieving Landmark and Non-Landmark Images from Community Photo Collections," *Proc. Int'l Conf. Multimedia*, 2010, pp. 153–162.
6. B. Bao et al., "Efficient Region-Aware Large Graph Construction towards Scalable Multi-label Propagation," *Pattern Recognition*, vol. 44, no. 3, 2011, pp. 589–606.
7. R.M. Neal and G.E. Hinton, "A View of the EM Algorithm that Justifies Incremental, Sparse, and Other Variants," *Learning in Graphical Models*, Springer, 1998, pp. 355–368.
8. Y. Gao et al., "W2GO: A Travel Guidance System by Automatic Landmark Ranking," *Proc. Int'l Conf. Multimedia*, 2010, pp. 123–132.
9. X. Lu et al., "Photo2Trip: Generating Travel Routes from Geo-Tagged Photos for Trip Planning," *Proc. Int'l Conf. Multimedia*, 2010, pp. 143–152.
10. T. Hofmann, "Unsupervised Learning by Probabilistic Latent Semantic Analysis," *Machine Learning*, vol. 42, nos. 1-2, 2001, pp. 177–196.
11. D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," *J. Machine Learning Research*, vol. 3, 2003, pp. 993–1022.
12. Y. Pang et al., "Summarizing Tourist Destinations by Mining User-Generated Travelogues and Photos," *Computer Vision and Image Understanding*, vol. 115, no. 3, 2011, pp. 352–363.
13. Z. Yin et al., "Geographical Topic Discovery and Comparison," *Proc. Int'l Conf. World Wide Web*, 2011, pp. 247–256.
14. C. Chemudugunta and P.S.M. Steyvers, "Modeling General and Specific Aspects of Documents with a Probabilistic Topic Model," *Advances in Neural Information Processing Systems*, vol. 19, MIT Press, 2007, p. 241.

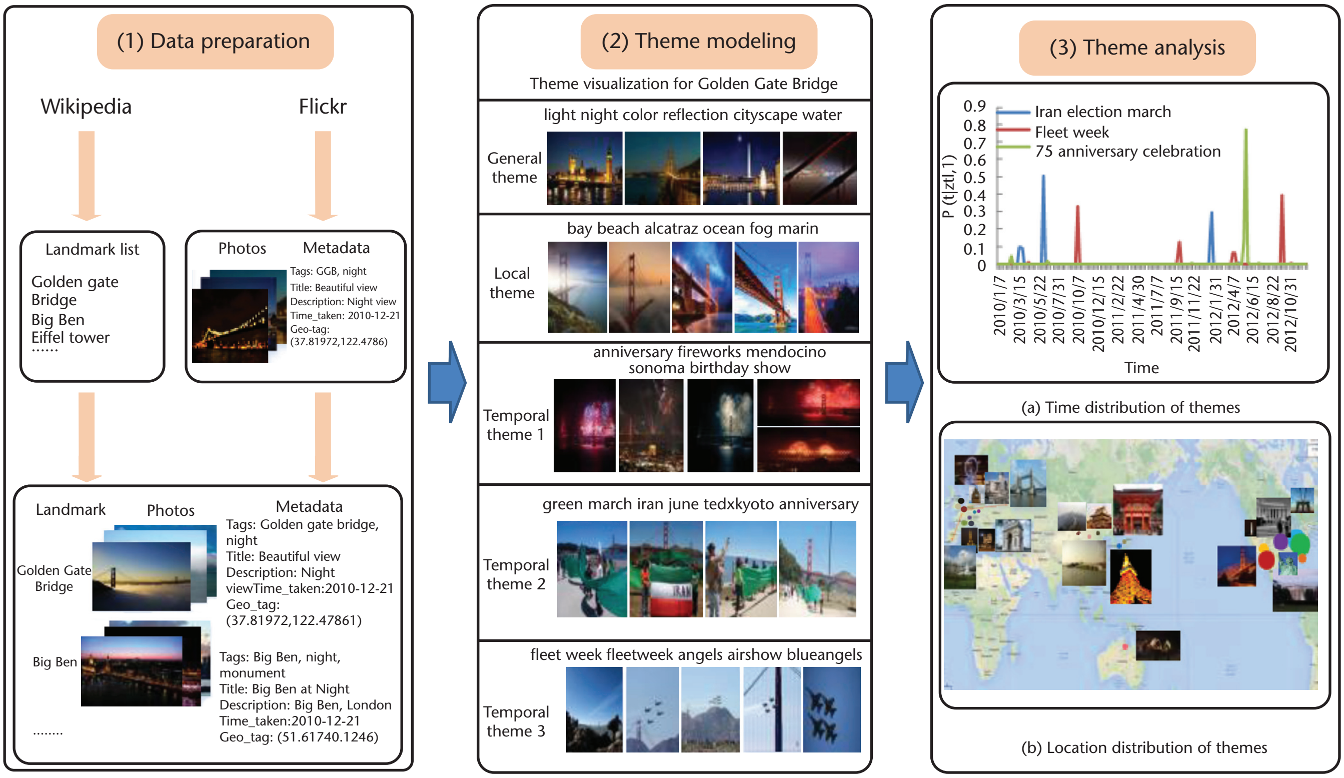


Figure 2. Overview of the multimodal spatio-temporal theme modeling framework. The mmSTTM model consists of three modules: data preparation, theme modeling, and theme analysis.

landmark name as the input, we would likely have gotten many noisy photos for two reasons:

- Some landmark names are polysemy. For example, the landmark "White House" not only represents the official residence and principal workplace of the President of the United States, but also any house that is painted white.
- Many photos are annotated with more than one landmark. For example, a photo might be annotated with both "Big Ben" and "Eiffel Tower" even though the photo itself represents the Eiffel Tower.

To solve these issues, we first combined the landmark name and corresponding city name as the query input. For instance, to crawl landmark photos of "Big Ben," we would query "Big Ben, London." The constraint created by the city name can greatly reduce the noise. Second, we used geotags of photos if available to conduct a denoising procedure on the collected dataset. Specifically, we used the geolocation (that is, the latitude and longitude) of each landmark from Wikipedia as the cluster center and then geographically grouped photos into clusters. Only photos in the correct clusters were used; we removed the rest. For photos without geotag information, we removed any that were annotated with more than one landmark.

Theme Modeling

The goal of the proposed mmSTTM model is to mine multiple kinds of themes by modeling documents' generative processes, where each Flickr photo and associated text are considered as one document. Based on the three-themes concept, we can define the problem of landmark theme mining as follows: Given the document set $D = \{d_1, d_2, \ldots, d_{|D|}\}$, where d_i is labeled with both a timestamp t_{d_i} and a landmark l_{d_i}, the goal of landmark theme mining is to learn the general theme space φ^{gl}; the local theme space φ^{loc}; and the temporal theme space φ^{tl}.

Generative Process of Documents

Given the Flickr document set D, which covers a set of words Y including the visual words V and textual words W, a set of landmarks L and a set of

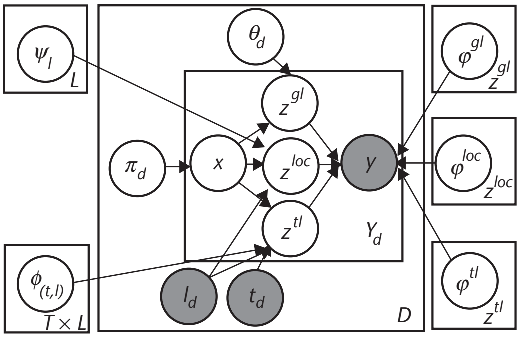


Figure 3. The graphical representation of the proposed landmark theme model.

time intervals T, each document d is associated with one landmark location $l_d \in L$ and one time interval $t_d \in T$. Based on a set of general themes Z^{gl}, local themes Z^{loc}, and temporal themes Z^{tl}, each word in document d is generated by

- a general theme $z \in Z^{gl}$, chosen according to document-specific distribution θ_d;
- a local theme $z \in Z^{loc}$, chosen according to landmark-specific distribution ψ_{l_d}; and
- a temporal theme $z \in Z^{tl}$, chosen according to landmark and time-specific distribution $\phi_{(l_d,t_d)}$.

The variable x is sampled from a document-specific multinomial distribution π_d, where π_d $\pi_d \triangleq \{p(x|d)|x \in \{gl, loc, tl\}\}$ represents the proportion distribution of the three themes. Figure 3 shows a graphical representation of the generation process. Note that ψ_1 is a $|L| \times |Z^{loc}|$ matrix and $\phi_{(l,t)}$ is an $|LT| \times |Z^{ti}|$ matrix. $|LT| = |L| \times |T|$. In our experiment, the number of landmarks $|L| = 20$, and the time scope is from 1 January 2010 to 31 December 2012. We simply divided one month into four time intervals, thus $|T| = 144$. The details of generative process for each document d in the document set are as follows. For each word $y_{d,n} \in Y$ from each document d,

1. draw $x_{d,n} \sim \text{Multi}(\pi_d)$ and
2. if $x_{d,n} = gl$, then draw a general theme $z_{d,n} \sim \text{Multi}(\theta_d)$;
3. if $x_{d,n} = loc$, then draw a local theme $z_{d,n} \sim \text{Multi}\left(\psi_{l_d}\right)$;
4. if $x_{d,n} = tl$, then draw a temporal theme $z_{d,n} \sim \text{Multi}\left(\psi_{(l_d,t_d)}\right)$; and
5. draw $y_{d,n} \sim \text{Multi}\left(\varphi_{z_{d,n}}^{x_{d,n}}\right)$.

According to the generative process of the document set D, we can write the log-likelihood of D as

$$\begin{aligned} L(D) = &\sum_{d \in D} \sum_{y \in Y} n(d,n) \\ &\times \log[p(x = gl|d) \sum_{z \in Z^{gl}} \theta_{d,z} \varphi_{z,y}^{gl} \\ &\times p(x = loc|d) \sum_{z \in Z^{loc}} \psi_{l_d,z} \varphi_{z,y}^{loc} \\ &\times p(x = tl|d) \sum_{z \in Z^{tl}} \phi_{(t_d,l_d),z} \varphi_{z,y}^{tl} \end{aligned} \tag{1}$$

Mutual-Information-Based Regularization

Similar to other work,[7] the occurrences of each local theme in documents should be highly correlated with those landmarks. We measure the correlation between the location set L and local theme set Z^{loc} with mutual information and define a location-based regularizer as

$$I_l\left(L; Z^{loc}\right) \triangleq \sum_{l \in L} \sum_{z \in Z^{loc}} p(l,z) \log \frac{p(l,z)}{p(l)p(z)} \tag{2}$$

where the probability distribution $\{p(l)\}_{l \in L}$ is set to uniform distribution if no prior knowledge of the locations exists.

Given one landmark, the occurrences of each temporal theme in a document should also be correlated with the corresponding time and location. We define the time- and location-based regularizer as

$$\begin{aligned} I_{(l,t)}\left((L,T); Z^{lt}\right) \triangleq &\sum_{(t,l) \in (T,L)} \sum_{z \in Z^{tl}} p((l,t),z) \\ &\times \log \frac{p((l,t)z)}{p((l,t))p(z)} \end{aligned} \tag{3}$$

where the probability distribution $\{p(l,t)\}_{(l,t) \in (L,T)}$ is also set to uniform distribution if no prior knowledge exists.

Parameter Estimation

Given the introduction of mutual-information-based regularization, we utilized the generalized EM algorithm (GEM)[8] for parameter estimation by solving the following regularized optimization problem:

$$\max\left(L(D) + \lambda_1 I_l\left(L; Z^{loc}\right) + \lambda_2 I_{(l,t)}\left((L,\ T);\ Z^{lt}\right)\right) \tag{4}$$

where λ_1 and λ_2 are the regularization parameters.

Theme Analysis

Once we estimate the parameters using GEM, the parameters can support further analysis, including temporal- and location-based distribution of themes.

Temporal Distribution of Themes

We can classify a landmark's temporal theme into two types:

- a periodic theme, which repeats at regular intervals, or
- an aperiodic theme, which is transient and intensively covered only in a certain time period.

We differentiate between the two temporal themes as follows: Given a specific landmark l, we compute the conditional probability of time t given the temporal theme z and the landmark l — that is, $p(t|z, l)$. Based on the estimated landmark and time-specific distribution $\phi_{(l,t)}$, we calculate $p(t|z, l)$ according to the Bayes' theorem as

$$P(t|z,l) = \frac{\phi_{(l,t)} p(t,l)}{\sum_{t' \in T} \phi_{(l,t')} p(t',l)} \tag{5}$$

where $p(t, l)$ is given by the word count in the time period t at location l divided by the total word count in the collection at location l.

Location Distribution of Themes

We can also analyze the location distribution of themes given the specific time interval. Similarly, this analysis can be inferred from the Bayes' theorem:

$$P(l|z,t) = \frac{\phi_{(l,t)} p(t,l)}{\sum_{l' \in L} \phi_{(t,l')} p(t,l')} \tag{6}$$

where $p(t, l)$ is given by the word count in the time period t at location l divided by the total word count in the collection in the time period t.

Experiment

To test our approach, we conducted an experiment using photos of 20 well-known international landmarks taken from 1 January 2010 to 31 December 2012:

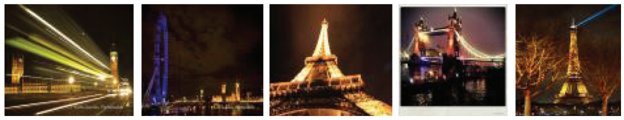
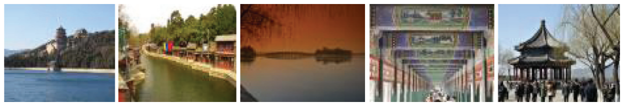
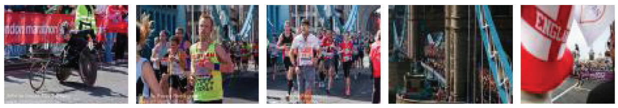
- Arc De Triomphe
- Big Ben
- Brooklyn Bridge
- Buckingham Palace
- Eiffel Tower
- Washington Monument
- Forbidden City
- Golden Gate Bridge
- Great Wall
- Kiyomizu-dera
- London Eye
- Lincoln Memorial
- Statue of Liberty
- Notre Dame
- Summer Palace
- Sydney Opera House
- Tokyo Tower
- Tower Bridge
- Trafalgar Square
- White House

We concatenated the tags, title, and description of one photo as the text information. First, because the landmark names and relevant city names are meaningless in discovering themes, we first removed them. Next, we cleaned the text by removing the stop words, HTML tags, and camera-related words such as "Cannon" and "35mm." Finally, we removed the words with a frequency lower than 50 and selected photos with more than eight words to build our final dataset. The resulting dataset has 435,810 photos and 22,703 unique words. For the visual content of photos, we choose histogram of oriented gradients (HoG) features with 1,024 dimensions.

Evaluating Theme Modeling

After preprocessing, we trained an mmSTTM on the Flickr dataset to learn the general, local, and temporal themes. We empirically set the number of general, local, and temporal themes to 50, 50, and 150, respectively. In addition, we empirically set the value of the regularization parameters λ_1 and λ_2 in Equation 4 to $1e + 6$. We set the granularity of location as one landmark; for time granularity, we simply divided one month into four parts: seven days for each of the first three parts, and the remaining days

Figure 4. Example themes discovered by mmSTTM. We label local themes with the landmark name, and we add the topic title and time for temporal themes.

Type of Topics	Top 5 words	Top 5 relevant photos
General#12	blue 0.2531 sky 0.23845 world 0.063868 nature 0.023623 cloud 0.02143	 0.312264 0.228972 0.228302 0.215036 0.203640
General#17	night 0.38077 lights 0.13088 dark 0.20122 nightshot 0.017241 nightphotography 0.01534	 0.311728 0.307387 0.297862 0.282610 0.262326
General#50	architecture 0.11974 building 0.097822 monument 0.033438 historic 0.032848 famous 0.031878	 0.146580 0.143253 0.122393 0.103129 0.099448
Local#5 Summer Palace	kunming 0.022126 lake 0.016091 longevity 0.015694 hill 0.0097054 bridge 0.0095046	 0.157004 0.152931 0.152222 0.150788 0.150391
Local#26 Big Ben	thames 0.12577 river 0.055631 southbank 0.044495 riverthames 0.031256 bank 0.024509	 0.135706 0.135705 0.131267 0.130258 0.126970
Local#27 Sydney Opera House	harbour 0.057095 vivid 0.038343 bridge 0.031438 harbourbridge 0.0206 sydneyharbour 0.0195	 0.078688 0.072973 0.071117 0.067875 0.066036
Temporal#67 10/01/2011- 10/07/2011 Brooklyn Bridge Occupy WallStreet	occupy 0.094204 occupywallstreet 0.079832 ows 0.058307 march 0.032343 wall 0.029606	 0.101686 0.101680 0.101607 0.101482 0.101471
Temporal#70 08/08/2012- 08/14/2012 Tower Bridge London Marathon	marathon 0.22029 philippound 0.040515 runner 0.025902 philip 0.023851 pound 0.23138	 0.203266 0.203159 0.202948 0.185978 0.183135
Temporal#114 04/01/2010- 04/07/2010 Washington Monument Cherry Blossom	spring 012389 cherry 0.08374 blossom 0.060646 april 0.047882 sakura 0.043024	 0.148427 0.147661 0.145635 0.142938 0.135086

for the fourth part. Our framework can be easily extended to different time granularities—such as one day—depending on the requirements.

Qualitative Case Study. Here, we illustrate three kinds of discovered themes by providing representative keywords and photos. The tag words are sorted according to $p(w|z)$, while photos are sorted by the cosine similarity between the topic Z_i distribution and the document d_j.

$$sim(Z_i, d_j) = \frac{(\mathbf{w}_{Z_i}, \mathbf{v}_{Z_i})\left(\mathbf{w}_{d_j}, \mathbf{v}_{d_j}\right)}{\left|(\mathbf{w}_{Z_i}, \mathbf{v}_{Z_i})\right|\left|\left(\mathbf{w}_{d_j}, \mathbf{v}_{d_j}\right)\right|} \quad (7)$$

where $(\mathbf{w}, \mathbf{v})$ is the concatenated word and visual word vector.

As Figure 4 shows, "temporal (or local or general) theme #j" denotes the *j*th temporal (or local or general) theme discovered by the model. Meanwhile, we label discovered general, local, and temporal themes with the help of θ_d, ψ_l, and $\phi_{(l,t)}$. To facilitate the interpretation, we label local themes with the landmark name. For temporal themes, we also add the topic title and time. As Figure 4 shows, some temporal themes are about distinctive views at special moments (such as temporal theme #114). Other temporal themes characterize activities, such as temporal theme #67 "Occupy Wall Street" in "Brooklyn Bridge," and temporal theme #70 "London Marathon" in "Tower Bridge." Such temporal themes help us more deeply understand a landmark's social function. Local themes are mainly related to the styles of certain landmarks. For example, local theme #27 characterizes the natural style "harbor." In contrast, general themes such as weather (general theme #12) and buildings (general theme #50) tend to apply to many landmarks.

Table 1. Performance comparison on the discovered themes.

Methods	Local themes	Temporal themes
Probabilistic latent semantic analysis (PLSA)	17	22
Travelogue model (TM)	20	25
mmSTTM_Text	27	39
mmSTTM	330	45

Table 2. Performance comparison on the discovered themes.

Methods	MAP@10 for words	MAP@10 for photos
PLSA	0.6188	0.5887
TM	0.6867	0.6368
mmSTTM_Text	0.7837	0.7411
mmSTTM	0.8399	0.8025

Quantitative Evaluation. We can now quantitatively evaluate mmSTTM. To do this, we use three baseline methods for comparison:

- Probabilistic latent semantic analysis (PLSA)[9] learns 250 themes without considering the visual modality.
- Travelogue model (TM)[7] learns general and local themes without considering the visual modality. We set the number of general and local themes to 50 and 200, respectively, and empirically set TM's regularization parameter (λ) values to $1e + 6$.
- mmSTTM—Text is similar to mmSTTM, but it doesn't consider the visual modality. Here, we set the number for each of the three themes to be the same as in mmSTTM (that is, 50, 50, and 150, respectively).

As in standard PLSA, the discovered themes have no explicit ranking. Because our goal is to primarily analyze the landmarks using discovered local and temporal themes, we first evaluate them on the number of semantically meaningful local and temporal themes. To further evaluate the semantic consistency for each semantically meaningful theme, we evaluate mmSTTM using MAP@K.

We asked 10 users to label semantically meaningful themes, consulting both the Flickr dataset and external resources for help. Table 1 shows the statistics on the number of discovered themes. As the table indicates, mmSTTM outperforms the three baselines. Our analysis of the experimental results reveals the following three key points:

- Because TM differentiates between the general and local themes, it filters words representing general themes and thus outperforms PLSA.[7]
- mmSTTM—Text significantly outperforms TM because it can differentiate the three kinds of themes.
- Using mmSTTM offers further improvements because the multimodal information used in mmSTTM can compensate and enhance each other.

To further evaluate the semantic consistency of meaningful themes, we asked the same 10 users to annotate the top 10 keywords for each labeled theme with a "theme-relevant tag" and a "theme-irrelevant tag" for our method and baselines. If more than six users thought a word was relevant to a theme, we labeled it with a 1; otherwise, it was labeled with a 0. We evaluated the tag ranking list that each method generated using MAP@10. We used similar methods to evaluate the photos for labeled themes. Table 2

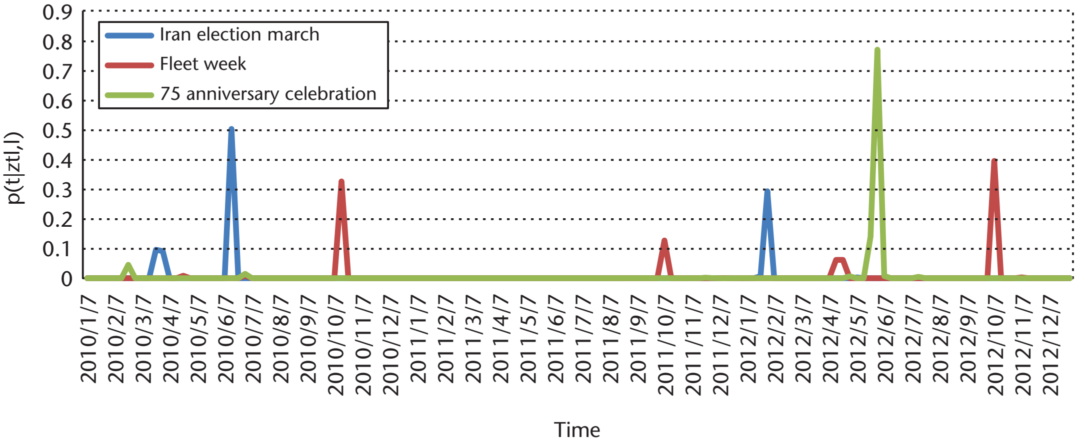


Figure 5. Time distribution of themes for Golden Gate Bridge. The "Fleet Week" theme has a clear periodic pattern, whereas other themes such as the "75 anniversary celebration" appear less frequently.

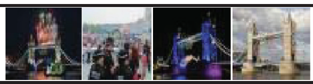
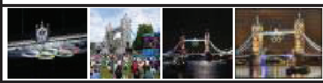
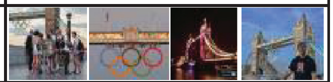
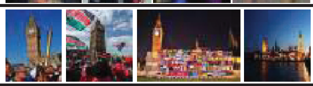
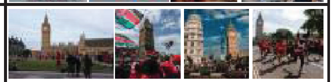
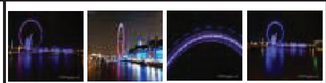
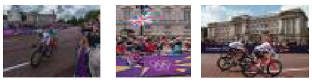
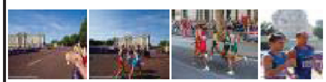
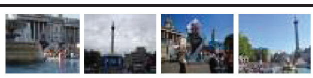
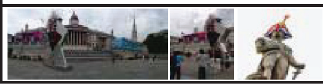
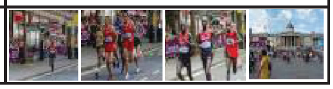
Landmarks	07/22/2012-07/31/2012		08/01/2012-08/07/2012		08/08/2012-08/14/2012	
	P(l|ztl,t)	Relevant photos	P(l|ztl,t)	Relevant photos	P(l|ztl,t)	Relevant photos
Tower Bridge	0.64		0.58		0.30	
Big Ben	0.30		0.11		0.10	
London Eye	0.05		0.01		0.04	
Buckingham Palace	0.003		0.07		0.33	
Trafalgar Square	0.0001		0.22		0.219	

Figure 6. Location distribution of the theme "2012 Olympic games in London" between 22 July and 14 August 2012.

shows the results; once again, our model outperforms other baselines.

Analysis of Mined Themes

For temporal distribution of themes, we plotted the time distribution of themes at the same landmark according to Equation 5. Without loss of generality, we randomly selected the Golden Gate Bridge landmark as an example to analyze the temporal distribution. We first manually labeled each temporal theme from this landmark, with the help of $\phi_{(l,t)}$. Figure 5 shows the time distribution of different Golden Gate Bridge themes; the "Fleet Week" theme has a clear periodic pattern. Fleet Week occurs between 6–9 October each year. In contrast, other temporal themes—especially the "75 anniversary celebration"—are aperiodic. We can easily apply this analysis to other landmarks to differentiate between the periodic and aperiodic themes.

To evaluate the location distribution of a special theme in a given time interval, we used Equation 6 to analyze the temporal theme "2012 Olympic games in London" from 22 July to 14 August 2012. Figure 6 shows the results. Because the 2012 Olympic Games were held in London, this theme is mainly distributed across London landmarks, particularly the Tower Bridge, which was illuminated with the Olympic Rings and thus was a key tourist attraction during the games. In addition, the theme is strongest at Buckingham Palace in the third week, during which it hosted the popular men's marathon race. As these results show, in addition to making recommendations for tourists, our analysis shows how important events impact a landmark's popularity.

Future Work

We will continue our future work on mmSTTM in two directions. First, we plan to change our model from offline to online to make it more practical. Second, we will apply mmSTTM to other applications, such as event-based landmark browsing and landmark popularity analysis.

MM

Acknowledgments

This work is supported in part by National Basic Research Program of China (grants 2012CB316304), National Natural Science Foundation of China (grants 61225009 and 61201374), and Beijing Natural Science Foundation (No. 4131004). Our work is also supported by the Singapore National Research Foundation under its International Research Centre@Singapore Funding Initiative and administered by the IDM Programme Office.

References

1. L. Kennedy and M. Naaman, "Generating Diverse and Representative Image Search Results for Landmarks," *Proc. 17th Int'l Conf. World Wide Web*, 2008, pp. 297–306.
2. Y. Li, D. Crandall, and D. Huttenlocher, "Landmark Classification in Large-Scale Image Collections," *Proc. IEEE Int'l Conf. Computer Vision*, 2009, pp. 1957–1964.
3. Y. Zheng et al., "Tour the World: Building a Web-scale Landmark Recognition Engine," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2009, pp. 1085–1092.
4. Y. Avrithis et al., "Retrieving Landmark and Non-Landmark Images from Community Photo Collections," *Proc. Int'l Conf. Multimedia*, 2010, pp. 153–162.
5. Y. Gao et al., "W2GO: A Travel Guidance System by Automatic Landmark Ranking," *Proc. Int'l Conf. Multimedia*, 2010, pp. 123–132.
6. X. Lu et al., "Photo2Trip: Generating Travel Routes from Geo-Tagged Photos for Trip Planning," *Proc. Int'l Conf. Multimedia*, 2010, pp. 143–152.
7. Y. Pang et al., "Summarizing Tourist Destinations by Mining User-Generated Travelogues and Photos," *Computer Vision and Image Understanding*, vol. 115, no. 3, 2011, pp. 352–363.
8. R.M. Neal and G.E. Hinton, "A View of the EM Algorithm that Justifies Incremental, Sparse, and other Variants," *Learning in Graphical Models*, Springer, 1998, pp. 355–368.
9. T. Hofmann, "Unsupervised Learning by Probabilistic Latent Semantic Analysis," *Machine Learning*, vol. 42, nos. 1-2, 2001, pp. 177–196.
10. R. Ji et al., "Mining City Landmarks from Blogs by Graph Modeling," *Proc. 17th ACM Int'l Conf. Multimedia*, 2009, pp. 105–114.
11. S. Zhang et al., "Query Specific Fusion for Image Retrieval," *Computer Vision–ECCV*, Springer, 2012, pp. 660–673.

Weiqing Min is a research associate and doctoral candidate at Institute of Automation, Chinese Academy of Sciences. His current research interests include landmark-based multimedia search and mining, computer vision, and pattern recognition. Min has an ME in communications and information systems from Wuhan University, China. Contact him at wqmin@nlpr.ia.ac.cn.

Bing-Kun Bao is an assistant researcher at Institute of Automation, Chinese Academy of Sciences, and a researcher at China-Singapore Institute of Digital Media. Her research interests include cross-media cross-modal image search, social event detection, image classification and annotation, and sparse/low rank representation. Bao has a PhD in control theory and control application from the Department of Automation at the University of Science and Technology of China (USTC). Contact her at bingkunbao@gmail.com.

Changsheng Xu is a professor at Institute of Automation, Chinese Academy of Sciences, and an executive director at the China-Singapore Institute of Digital Media. His current research interests include multimedia content analysis, indexing, and retrieval; pattern recognition; and computer vision. Xu has a PhD in electric engineering from Tsinghua University, Beijing. Contact him at csxu@nlpr.ia.ac.cn.