

Chapter 2

Mapping Geotagged Tweets to Tourist Spots Considering Activity Region of Spot

Kenta Oku and Fumio Hattori

Abstract We are developing a recommender system for tourist spots. The challenge is mainly to characterize tourist spots whose features change dynamically with trends, events, season, and time of day. Our method uses a one-class support vector machine (OC-SVM) to detect the regions of substantial activity near target spots on the basis of tweets and photographs that have been explicitly geotagged. A tweet is regarded as explicitly geotagged if the text includes the name of a target spot. A photograph is regarded as explicitly geotagged if the title includes the name of a target spot. To characterize the tourist spots, we focus on geotagged tweets, which are rapidly increasing on the Web. The method takes unknown geotagged tweets originating in activity regions and maps these to target spots. In addition, the method extracts features of the tourist spots on the basis of the mapped tweets. Finally, we demonstrate the effectiveness of our method through qualitative analyses using real datasets on the Kyoto area.

Keywords Geotagged user generated content · Geotagged tweet · Tourist spot analysis

2.1 Introduction

There is a rising demand for reinvigoration of the tourist industry through information technology. Because of the enormously wide variety of tourist spots all over the world, there is a significant need to apply particular search and recommendation technologies [1, 2] to the field of tourism in order to provide relevant spots with visitors.

K. Oku (✉) · F. Hattori
College of Information Science and Engineering, Ritsumeikan University,
1-1-1 Nojihigashi, Kusatsu-city, Shiga 525-8577, Japan
e-mail: oku@fc.ritsumei.ac.jp

F. Hattori
e-mail: fhattori@is.ritsumei.ac.jp

In order to search and recommend tourist spots effectively, it is first necessary to characterize the tourist spots. Although there are currently many tourism information websites, most of the information that appears at such websites was collected at certain points in time and thus is static.

The difficulty is that the features of tourist spots change dynamically with trends, events, season, and time of day. For instance, a spot that is famous for its red leaves is more attractive in the autumn, while a spot that is famous for its bright lights is more attractive by night. In addition, a spot where a festival is held is more attractive during the period of the festival. However, to maintain information on such features of tourist spots at fixed intervals is very costly, because most existing tourism information websites are managed manually.

In order to avoid such problems, we focus on geotagged user-generated content (G-UGC), which has rapidly been increasing on the Web of late. There are certain service providers, such as Foursquare,¹ Twitter,² and Panoramio,³ which enable users to post a variety of G-UGC. In the case of Twitter, particularly, a great number of messages (called tweets) are posted daily because of its simplicity. Furthermore, geotagged tweets are exponentially increasing with the spread of GPS-equipped mobile devices. In addition to message text, a geotagged tweet includes not only user information and the submission time but also the device location (i.e., latitude and longitude). We focus on the principle that the features of tourist spots, which may change dynamically, can be extracted by mapping geotagged tweets to tourist spots.

A problem to be solved is how to map geotagged tweets to tourist spots. However, based solely on the text and locations of tweets, it is difficult to infer which tweets are related to which spots. For instance, if a tweet includes the name of tourist spot like the Kiyomizu-dera Temple, then mapping the tweet to the spot is easy, but most tweets do not explicitly include the name of a tourist spot. One approach is mapping a tweet to a spot if the tweet originates within some arbitrary radius from the location of the spot. Still, to define the region around a tourist spot appropriately is not easy, because the extent of such a region is usually not clear.

In order to solve the above mentioned problem, we propose a method that infers the regions of substantial activity surrounding target spots on the basis of tweets and photographs that are explicitly geotagged and originated near the spots. A tweet is regarded as explicitly geotagged if the message text includes the name of a target spot. A photograph is regarded as explicitly geotagged if the title includes the name of a target spot. Then, we propose a method for taking geotagged tweets originating in the activity regions and mapping these to the target spots. The activity region of a target spot is defined as the region that people actually visit to enjoy the spot, not as the region indicated by its address or location. We use a one-class support vector machine (OC-SVM) [3, 4] to infer the activity regions of target spots.

¹<https://foursquare.com/>.

²<https://twitter.com/>.

³<http://www.panoramio.com/>.

The remainder of this paper is organized as follows: Sect. 2.2 discusses related work. Section 2.3 explains the geotagged user-generated content used in this study. Section 2.4 presents our proposed method for mapping geotagged tweets to tourist spots, and Sect. 2.5 presents that for extracting the features of the tourist spots on the basis of the mapped tweets. Section 2.6 shows results from qualitative analyses using real datasets to demonstrate the effectiveness of the above methods. Section 2.7 concludes this paper and discusses future directions.

2.2 Related Work

Several studies propose POI (points of interest) recommendations, which provide locations suitable for users' preferences.

Crandall et al. [5] proposed a system that displays photographs of landmarks on a map. Their system extracts landmarks where many people take photographs. The mean-shift algorithm, which is a clustering method, is applied to geotagged photographs that are posted on Flickr. Zheng et al. [6–8] proposed a system that extracts POIs. The POIs are places where many people stay, and these are extracted by clustering the GPS trajectory data collected. They utilize Tree-Based Hierarchical Graph (TBHG) to cluster the GPS trajectory data. Leung et al. [9] propose a collaborative location recommendation framework, which incorporates user activity in addition to user and location relations. They also extract POIs by TBHG based on GPS trajectories.

Ye et al. [10] exploit geographical and social influence to recommend POIs based on location-based social networks (LBSNs). Gao et al. [11] exploit temporal effects for POI recommendations. These studies extract POIs based on check-in histories from foursquare.

Lee et al. [12] proposed a method that extracts geographical events based on geotagged tweets. The method depends on geographical regularities deduced from the usual patterns of geotagged tweets. It focuses on temporal variations within the target regions as important clues for extracting the geographical events. Lee [13] and Wakamiya et al. [14, 15] proposed another method that extracts characteristics of urban areas by monitoring crowds through geotagged tweets.

As stated above, there have been many studies attempting to extract characteristic regions based on geotagged user-generated content. While these studies mainly attempted to extract POIs by clustering methods for regions where users generate content densely, we attempt to extract features of tourist spots by mapping geotagged tweets to the spots.

2.3 Geotagged User-Generated Content

In this study, we obtain data on tourist spots from Foursquare, which is one of the location-based social networking services (LBSNs). Then, we obtain geotagged tweets from Twitter, which is one of the microblog services, and we map these tweets

Table 2.1 Names of categories targeted as tourist spots

Hiking Trail	Religious Center	Playground	Zoo
Mountain	Shrine	Park	Museum
Lake	Temple	Sculpture Garden	Art Gallery
River	Campground	Monument or Landmark	Art Museum
Beach	Dog Run	Bridge	Arts and Entertainment
History Museum	Farm	Harbor or Marina	Science Museum
Cemetery	Garden	Boat or Ferry	Aquarium
Historic Site	Garden Center	Pier	Scenic Lookout

to the tourist spots. In order to map the tweets to the tourist spots, we utilize geotagged photographs from Panoramio, which is one of the photograph sharing websites.

In the subsections below, we describe the three types of geotagged user-generated content (G-UGC) and how this is collected.

2.3.1 Tourist Spots

Foursquare is one of the LBSNs and began in March 2009. Users can participate in the service by using mobile devices such as smartphones to check in at places of interest called venues. Currently, over 30 million people are participating in the service, and over 3 billion check-ins have been posted.

We collected venue data by using the application program interface (API)⁴ released by Foursquare. The venues represented various categories, such as universities, train stations, and bookstores. In this study, tourist spots were defined as venues in the categories listed in Table 2.1.

We inserted the collected tourist spot data into the following *spot* table:

spot(id, name, address, latitude, longitude, category_name, url)

2.3.2 Geotagged Tweets

Twitter is one of the microblog services and began in July 2006. Users can post messages called tweets. Due to the spread of mobile devices with GPS receivers, geotagged tweets have been increasing recently. The number of tweets has been growing yearly and reached about 35 million tweets per day in 2010.

We collected geotagged tweets by using the streaming API⁵ released by Twitter. We inserted the collected tweets into the following *tweet* table:

⁴<https://developer.foursquare.com>.

⁵<https://dev.twitter.com/docs/streaming-apis>.

Table 2.2 Differences between types of user-generated content

User-generated content	Amount of data	Clarity of target	Target region	Text data	Temporal data
Tourist spot	Small	Clear	No	No	No
Geotagged tweet	Enormous	Unclear	Yes	Yes	Yes
Geotagged photograph	Small	Clear	Yes	No	No

tweet(id, user_id, user_name, text, year, week_of_year, hour, latitude, longitude)

2.3.3 Geotagged Photographs

Panoramio is one of the photograph sharing sites and began in October 2005. Users can upload photographs for display on a map.

We collected geotagged photographs by using the API⁶ released by Panoramio. We inserted the collected photographs into the following *photograph* table:

photo(photo_id, photo_title, photo_url, longitude, latitude, owner_id, owner_name)

Table 2.2 summarizes the features of the abovementioned three types of G-UGC.

Twitter enables its users to post tweets of up to 140 characters. Because this limitation encourages users to post tweets frequently, an enormous number of tweets are posted each day. Compared to the tweets on Twitter, the photographs on Panoramio are carefully screened. Panoramio users specially upload favorite photographs from among many taken. Most geotagged photographs also include the names of tourist spots in the titles. This helps to clarify the associated target spots. Furthermore, geotagged photographs are higher in quality and fewer in number than are geotagged tweets.

Geotagged tweets include an enormous amount of data, such as text and temporal information. Hence, these are useful sources from which to extract the features of tourist spots. However, the target spots of geotagged tweets are unclear, while those of geotagged photographs are clear. It is thus a challenge to link tweets with tourist spots and geotagged photographs. Section 2.4 describes how to map geotagged tweets to tourist spots.

⁶<http://www.panoramio.com/api/data/api.html>.

In this paper, we respectively denote the collected tourist spot set, geotagged tweet set, and geotagged photograph set as follows:

$$S = \{s_1, s_2, \dots\}, \quad (2.1)$$

$$T = \{t_1, t_2, \dots\}, \quad (2.2)$$

$$P = \{p_1, p_2, \dots\}. \quad (2.3)$$

We represent each attribute in the form *object.attribute* (e.g., $t_1.text$).

2.4 Mapping Geotagged Tweets to Tourist Spots

By mapping geotagged tweets to tourist spots, features of the tourist spots can be extracted from the mapped tweets. However, based solely on the text and locations of tweets, it is difficult to infer which tweets are related to which spots. For instance, if a tweet includes the name of tourist spot like the Kiyomizu-dera Temple, then mapping the tweet to the spot is easy, but most tweets do not explicitly include the name of a tourist spot. One approach is mapping a tweet to a spot if the tweet originates within some arbitrary radius from the location of the spot. Still, to define the region around a tourist spot appropriately is not easy, because the extent of such a region is usually not clear.

In order to solve the above mentioned problem, we propose a method that infers the regions of substantial activity surrounding target spots on the basis of tweets and photographs that are explicitly geotagged and originated near the target spots. A tweet is regarded as explicitly geotagged if the message text includes the name of a target spot. A photograph is regarded as explicitly geotagged if the title includes the name of a target spot. Then, we propose a method for taking geotagged tweets originating in the activity regions and mapping these to the target spots.

The activity region of a target spot is defined as the region that people actually visit to enjoy the spot, not as the region indicated by its address or location. For example, those who visit Kiyomizu-dera Temple visit not only its main hall but also nearby buildings and approaches. These attractions should also be included in the activity region of Kiyomizu-dera. The address of Kiyomizu-dera, which is 1-294 Kiyomizu, Higashiyama-ku, Kyoto, Kyoto Prefecture, Japan, is not even roughly equivalent to its activity region. Figure 2.1 shows the activity region of Kiyomizu-dera (dark gray region) and the region indicated by its address (light gray region).

We use a one-class support vector machine (OC-SVM) [3, 4] to infer the activity regions of tourist spots. The OC-SVM can extract a high-density region based on a given training dataset.

As training datasets we utilize explicitly geotagged photographs and explicitly geotagged tweets that originated in the surrounding region of the target spot. Then, the OC-SVM learns the activity region of the spot based on the datasets. Finally, our

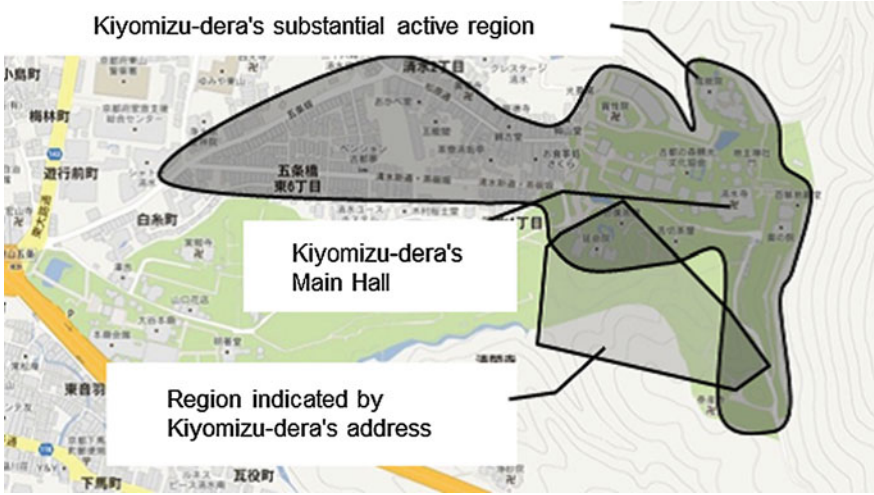


Fig. 2.1 Activity region of Kiyomizu dera Temple

proposed method maps unknown geotagged tweets to the target spot based on the learned region.

According to the example in Fig. 2.2, the steps for mapping geotagged tweets to a tourist spot are as follows:

- (1) Select the target spot $s_i \in S$.
- (2) Obtain the photograph set $P_i^* \subseteq P$ whose titles $p_i.photo_title$ include the target spot name $s_i.name$ and whose locations are within a radius r from the target spot.
- (3) Learn the region $R_{P_i^*}$ based on the photograph set P_i^* as the training set by using the OC-SVM.
- (4) Obtain the tweet set $T_i^* \subseteq T$ whose texts $t_i.text$ include the target spot name $s_i.name$ and whose locations are within a radius r from the target spot.
- (5) Learn the region $R_{T_i^*}$ based on the tweet set T_i^* as the training set by using the OC-SVM.
- (6) Obtain the region R_i by combining the region $R_{P_i^*}$ and the region $R_{T_i^*}$. We define the region R_i as the activity region of the target spot s_i .
- (7) Map the geotagged tweets T_i originating in the region R_i to the target spot s_i .

2.5 Extracting Features of Tourist Spots

Our proposed method extracts features of the tourist spot s_i based on the tweet set T_i mapped to the spot s_i . The method extracts the following two types of features of tourist spots: 1) temporal features, and 2) phrasal features.

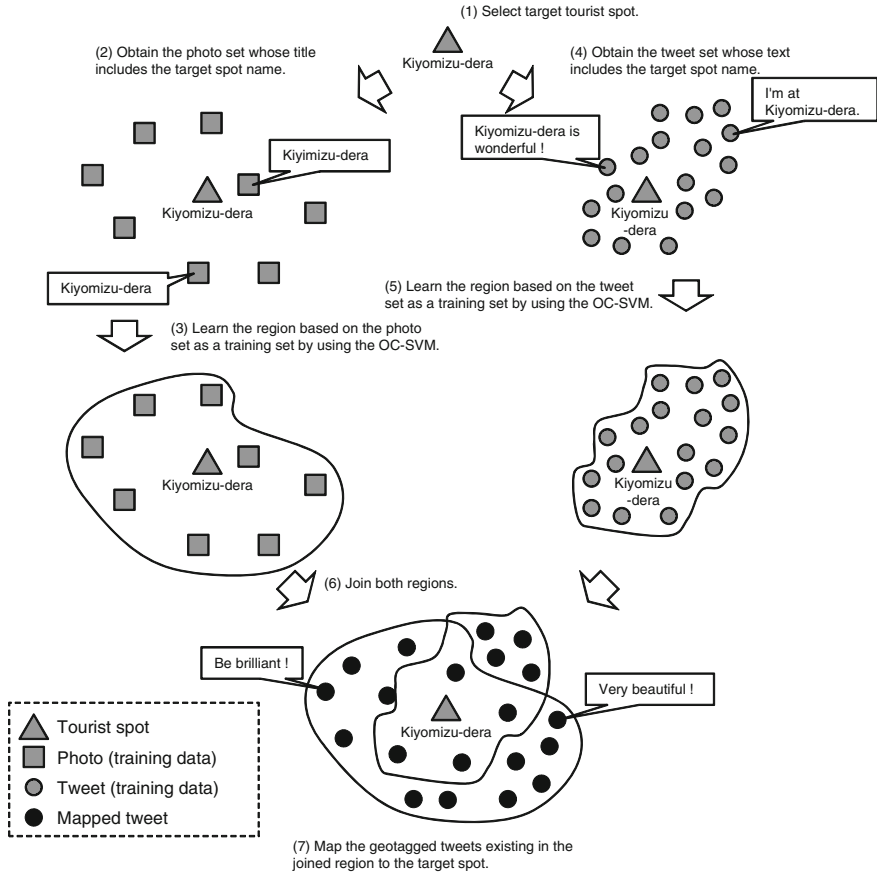


Fig. 2.2 Steps of method for mapping geotagged tweets to tourist spot

We explain each method in a subsection below.

2.5.1 Extracting Temporal Features of Tourist Spots

Temporal features of tourist spots can be extracted by analyzing the distributions of submission times ($t_i.year$, $t_i.week_of_year$, and $t_i.hour$) in the tweet set T_i . For instance, more tweets than usual are posted from Kiyomizu-dera in autumn, when the leaves change color, because Kiyomizu-dera is famous for its autumn leaves. In addition, even more tweets are posted at night, because Kiyomizu-dera is lit at night during this season.

First, using the example of Kiyomizu-dera in 2011, we describe how to extract temporal features related to *week_of_year* for each year.

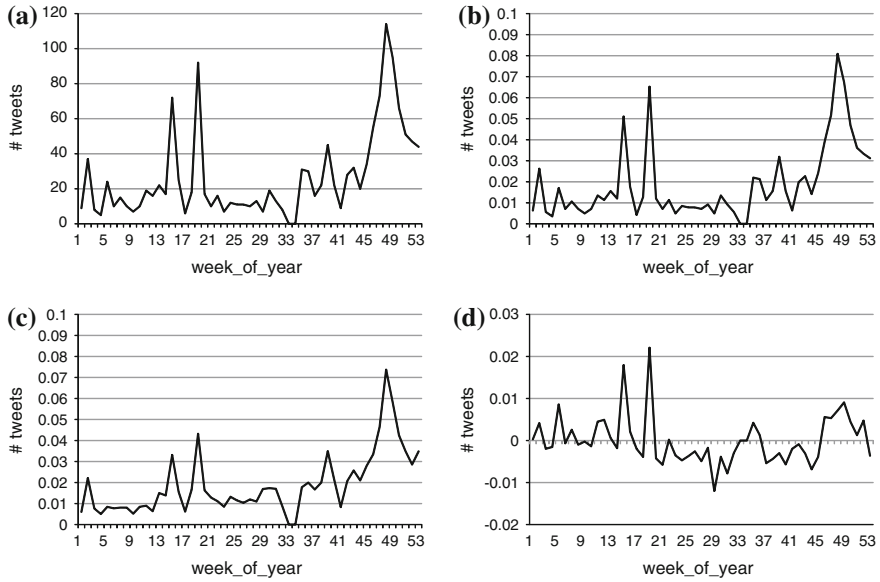


Fig. 2.3 Temporal features of Kiyomizu-dera related to *week_of_year* (in 2011). **a** Number of tweets by week (in 2011) for Kiyomizu-dera Temple. **b** Normalized number of tweets by week (in 2011) for Kiyomizu-dera Temple. **c** Mean number of tweets by week (in 2011) for all spots. **d** Difference in number of tweets by week (in 2011) for Kiyomizu-dera Temple

- (1) Count the number of tweets posted in each week ($week_of_year = 1, 2, \dots, 53$). We define the feature vector $\mathbf{W}_i = (w_{i1}, w_{i2}, \dots, w_{i53})$, where w_{ij} denotes the number of tweets in the j th week. Figure 2.3a shows the feature vector \mathbf{W}_i in a line graph (the horizontal axis denotes the week and the vertical axis denotes the number of tweets).
- (2) Normalize the number of tweets in each week by entering the total number of tweets $|T_i|$ in the tweet set T_i as 1. The well-known spots like Kiyomizu-dera tend to have many tweets mapped while other spots have fewer tweets. The normalization should be done to eliminate the differences between spots in the number of mapped tweets. We define the normalized feature vector $\mathbf{W}_i^* = (w_{i1}^*, w_{i2}^*, \dots, w_{i53}^*)$, where w_{ij}^* denotes the normalized number of tweets in the j th week. Figure 2.3b shows the normalized feature vector \mathbf{W}_i^* in a line graph.
- (3) Perform the above steps (1) and (2) for each tourist spot. Then, obtain the average of the normalized number of tweets over all spots. We define the average feature vector $\mathbf{W}_i^{\text{all}} = (w_{i1}^{\text{all}}, w_{i2}^{\text{all}}, \dots, w_{i53}^{\text{all}})$, where w_{ij}^{all} denotes the average of the normalized number of tweets in the j th week. Figure 2.3c shows the average of the normalized feature vector $\mathbf{W}_i^{\text{all}}$ in a line graph.
- (4) Obtain the difference feature vector $\mathbf{W}_i^{\text{diff}} = \mathbf{W}_i^* - \mathbf{W}_i^{\text{all}}$, which represents the difference between the normalized feature vector \mathbf{W}_i^* and the average feature vector $\mathbf{W}_i^{\text{all}}$. The number of tweets tends to be biased according to season.

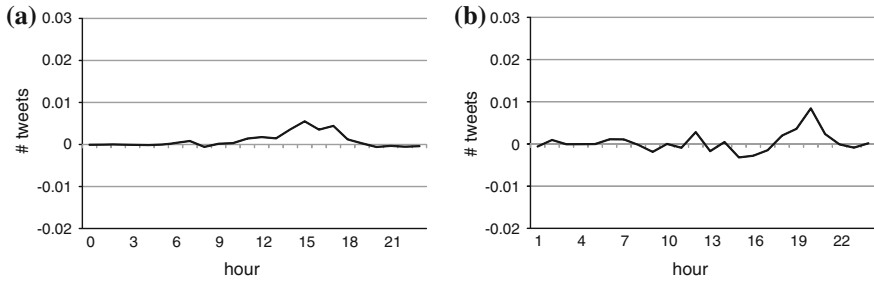


Fig. 2.4 Temporal features of Kiyomizu-dera related to *hour* (in 2011). **a** Difference in number of tweets by hour for Kiyomizudera Temple (in the 19th week of 2011). **b** Difference in number of tweets by hour for Kiyomizudera Temple (in the 49th week of 2011)

Therefore, the difference feature vector should be obtained to eliminate the bias. Figure 2.3d shows the difference feature vector $\mathbf{W}_i^{\text{diff}}$ in a line graph. We can see that Kiyomizu-dera is more attractive during the 19th week (called the Golden Week holiday in Japan) and the 49th week (for the autumn leaves).

In the same way, the temporal features related to *hour* can be extracted for each week of each year. By extracting for each week, the method can extract features depending on season, such as daytime popularity during the summer and nighttime popularity during the winter. Figures 2.4a, b show the temporal features related to *hour* as line graphs for the 19th week of 2011 and the 49th week of 2011, respectively. We can observe in the figures that at 15 and 20, respectively, the numbers of tweets in the 19th week and the 49th week are higher than usual. This is because many people visit to see light-up events held at Kiyomizu-dera.

In steps similar to those above, we define the feature vector \mathbf{H}_i , the normalized feature vector \mathbf{H}_i^* , the average feature vector $\mathbf{H}_i^{\text{all}}$, and the difference feature vector $\mathbf{H}_i^{\text{diff}}$, respectively.

2.5.2 Extracting Phrasal Features of Tourist Spots

Our method extracts phrasal features of tourist spots from the text ($t_i.\text{text}$) in the tweet set T_i . We use ChaSen⁷ as a morphological parser to extract Japanese phrases. Our method extracts parts of speech, such as nouns, adjectives, and unknown words. In the case of a noun phrase that can be formed with a particle (pronominal), the words are combined into one phrase. This process enables the method to extract compound phrases, such as “Kiyomizu-dera Main Hall” and “the stage of Kiyomizu.”

⁷<http://chasen.naist.jp/hiki/ChaSen>.

Table 2.3 Top 10 phrasal features of Kiyomizu-dera

Ranking	Feature phrase (value in parentheses denotes <i>tf-idf</i>)	
	16th week	49th week
1	Kiyomizu-dera (52.8365)	Kiyomizu-dera (161.3635)
2	Kiyomizu-dera (41.5633)	Higashiyama-ku, Kiyomizu (99.0543)
3	Higashiyama-ku, Kiyomizu (29.3494)	Zenko-ji Temple (22.9248)
4	the stage of Kiyomizu (13.9454)	the stage of Kiyomizu (20.9181)
5	Kyoto-city (5.9212)	Kyoto-city (19.9841)
6	cherry blossoms (5.5048)	Kiyomizu-dera Main Hall (15.0961)
7	active cherry blossoms (3.8918)	Jishu Shrine (11.3221)
8	the stage (3.8918)	the precincts of Kiyomizu-dera (7.5481)
9	Yae (3.8918)	Light up (6.1777)
10	visiting the temple at night (3.8918)	ganbare (4.5850)

The phrasal features are extracted for each week of each year. We denote as follows the phrases extracted from the tourist spot s_i for the j th week:

$$L_{ij} = \{l_{ij1}, l_{ij2}, \dots\}, \quad (2.4)$$

where l_{ijk} is the weight calculated by our method for the k th phrase extracted. To assign the weight of each phrase, we use (*tf-idf*), which is widely used in the field of document retrieval. The *tf* stands for term frequency. The *tf* of the phrase l_{ijk} corresponds to the number of instances of the phrase l_{ijk} from the tourist spot s_i during the j th week. The *idf* stands for inverse document frequency. The *idf* of the phrase l_{ijk} can be calculated as follows:

$$idf = \log \frac{|S|}{n_{ijk}}, \quad (2.5)$$

where $|S|$ denotes the total number of spots and n_{ijk} denotes the number of spots whose mapped tweet texts include the phrase l_{ijk} . Hence, the *tf-idf* of the phrase l_{ijk} can be calculated as follows:

$$tf-idf = tf \times idf. \quad (2.6)$$

Finally, the phrases in the phrase set L_{ij} are sorted by *tf-idf* in descending order. Based on the sorted phrases, our method can provide phrasal features for each spot in each week. For example, Table 2.3 lists the top 10 phrasal features of Kiyomizu-dera in the 16th and 49th weeks. We can see from the table that the phrase “cherry blossoms” was extracted in the spring while the phrase “light up” was extracted in the autumn.

Table 2.4 User-generated content used in the analyses

User-generated content	Items	Term of collection
Tourist spots	1,006	13 Jun 2012
Geotagged tweets	389,579	1 Jan 2011 to 31 Dec 2011
Geotagged photographs	12,480	30 Jun 2012

2.6 Qualitative Analyses

We conducted qualitative analyses in order to evaluate the effectiveness of the proposed methods explained in Sects. 2.4 and 2.5. This section describes the datasets used in these analyses and then describes the results of the analyses.

2.6.1 Datasets

We used three types of G-UGC as the datasets: tourist spots, geotagged tweets, and geotagged photographs, as described in Sect. 2.3.

In the analyses, we considered the Kyoto area, where a number of major Japanese tourist spots are located. We defined the target region as the rectangle with southwestern corner (34.87069°N, 135.566713°E) and northeastern corner (35.12967°N, 135.935152°E). We collected G-UGC originating within this rectangle. For each type of G-UGC, Table 2.4 lists the number of items collected and the term of collection.

2.6.2 Analysis of Mapping Method

We analyzed the effectiveness of the mapping method in the cases of a) Kiyomizudera, b) Kinkaku-ji, and c) Tetsugaku-no-michi. Figure 2.5 shows the geotagged tweet sets mapped to each tourist spot.

In the case of Kiyomizudera, we can see in the figure that the mapped tweets also include tweets not originating in the region associated with its official address. In addition, the tweets mapped to Kiyomizudera originate not only in the main hall but also in the nearby buildings and on the approaches.

In the case of Kinkaku-ji, the mapped tweets also include those posted while traveling from the nearest bus stop (Kinkaku-ji). Indeed, the tweets include many references such as “now going to Kinaku-ji” and “to Kinkaku-ji on foot.”

In the case of Tetsugaku-no-michi, the mapped tweets originated along the stretch of street from north to south. Our proposed method can extract such tourist routes that can be represented as linear.

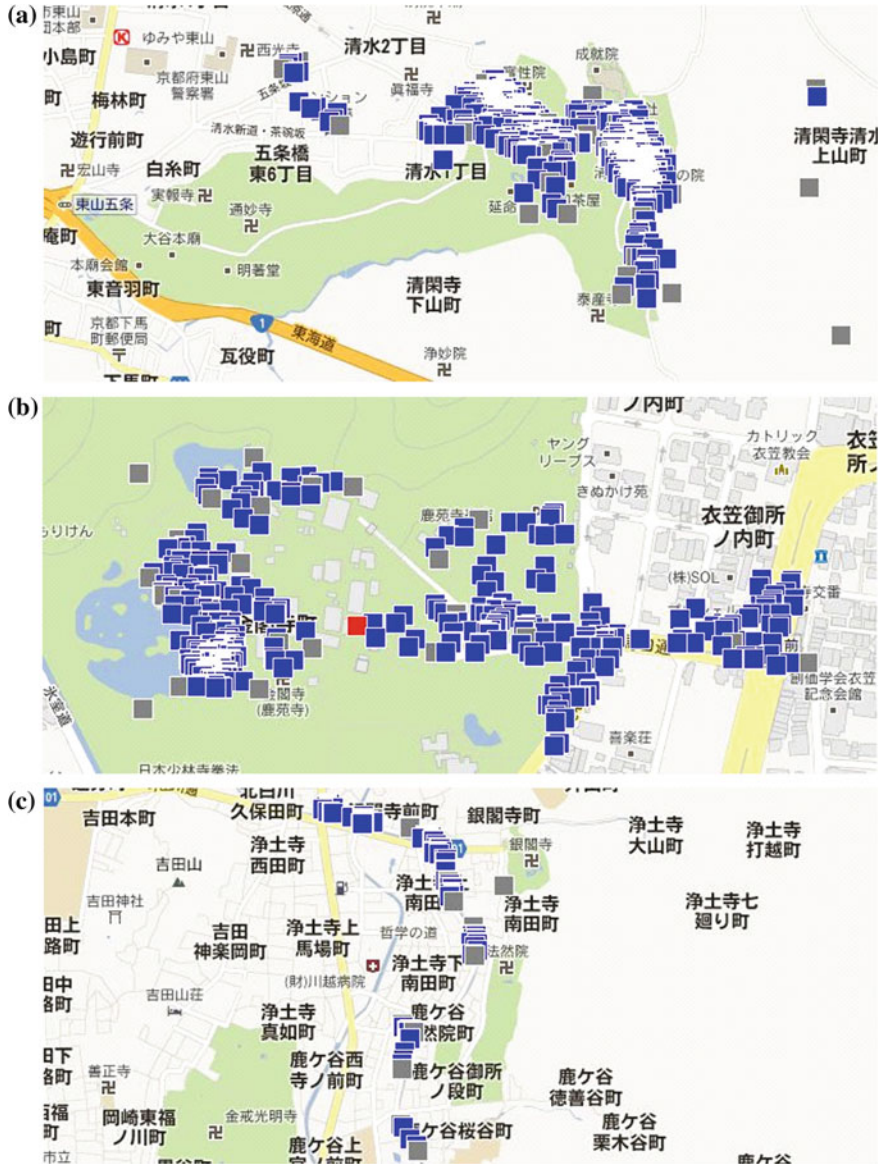


Fig. 2.5 Geotagged tweet sets mapped to each tourist spot. a Kiyomizu-dera. b Kinkaku-ji. c Tetsugaku-no-michi

Although we have merely provided some examples in this analysis, the results demonstrated that our mapping method can adequately map geotagged tweets to tourist spots by considering the regions of substantial activity.

2.6.3 Analysis of Feature Extraction Method

Figure 2.6 shows the temporal features related to *week_of_year* for each tourist spot.

In the case of Kinkaku-ji, more tweets were posted in the winter season than in other seasons. Focusing on the phrasal features extracted for this season, which extends from the 53rd week to the 3rd week, the phrase “snowscape of Kinkaku” was found.

In the case of Tetsugaku-no-michi, more tweets were posted in April (from the 15th week to the 16th week) and June (from the 20th week to the 27th week). Focusing on the phrasal features extracted for these weeks, the phrases “cherry blossoms” and “full bloom” were associated with April, while the word “hydrangea” was associated with June.

In the case of the Yasaka Shrine, more tweets were posted in July (from the 29th week to the 31st week). Focusing on the phrasal features extracted for these weeks, some phrases associated with the Gion Festival were found, such as “viewing of Gion Festival” and “Shinkosai” (Mikoshi togyo). The Gion Festival is held during those weeks in July.

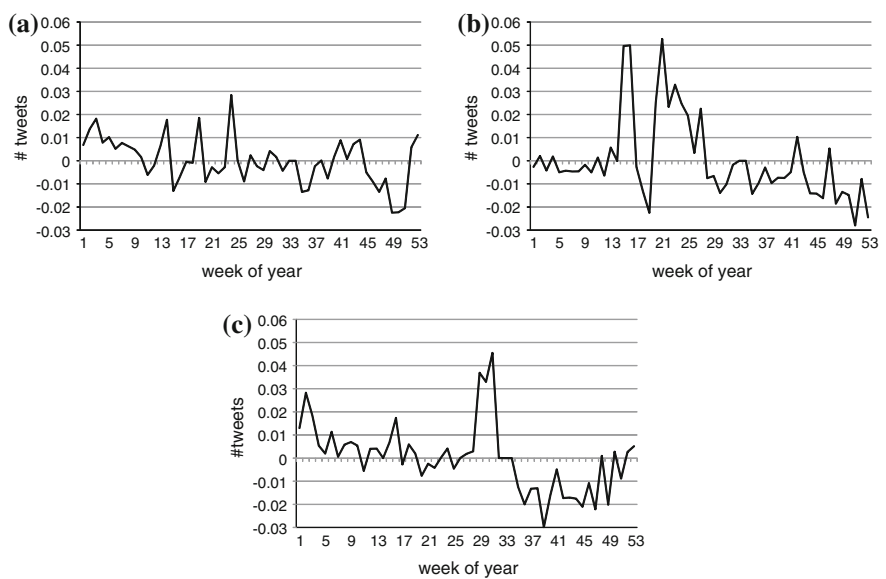


Fig. 2.6 Temporal features related to *week_of_year* for each tourist spot. **a** Kinkaku-ji. **b** Tetsugaku-no-michi. **c** Yasaka Shrine

The results demonstrated that our extraction method can extract temporal features and phrasal features of tourist spots. In future, we can develop a tourist spot recommender system based on such features.

2.7 Conclusion

In this paper, we proposed a method for mapping geotagged tweets to tourist spots on the basis of the substantial-activity regions of the spots as learned using an OC-SVM. We also proposed a method for extracting temporal features and phrasal features based on the mapped tweets. We showed the effectiveness of our methods through qualitative analyses using real datasets on the Kyoto area.

In future work, we would like to conduct further quantitative and qualitative experiments. We also would like to compare our method with other location clustering methods. Furthermore, we will specifically study how to extract features of spots for developing spot recommender system, and develop a tourist spot recommender system based on the extracted features.

Acknowledgments This work was supported by a Grant-in-Aid for Young Scientists (B) (23700132) from the Japan Society for the Promotion of Science (JSPS).

References

1. Resnick, P., Iacovou, N., Suchak, M., Bergstrom P., Riedl, J.: GroupLens: an open architecture for collaborative filtering of netnews. In: *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, pp. 175–186 (1994)
2. Adomavicius, G., Tuzhilin, A.: 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)
3. Cortes, C., Vapnik, V.: Support-Vector Networks. *Mach. Learn.* **20**(3), 273–297 (1995)
4. Manevitz, L.M., Yousef, M.: One-class SVMs for document classification. *J. Mach. Learn. Res.* **2**, 139–154 (2002)
5. Crandall, D.J., Backstrom, L., Huttenlocher, D., Kleinberg, J.: Mapping the world's photos. In: *18th International Conference on World Wide Web*, pp. 761–770 (2009)
6. Zheng, Y., Zhang, L., Xie, X., Ma, W.-Y.: Mining interesting locations and travel sequences from GPS trajectories. In: *18th International Conference on World Wide Web*, pp. 791–800 (2009)
7. Zheng, Y., Zhang, L., Ma, Z., Xie, X., Ma, W.-Y.: Recommending friends and locations based on individual location history. *ACM Trans. Web (TWEB)* **5**(1), 1–44 (2011)
8. Zheng, V.W., Zheng, Y., Xie, X., Yang, Q.: Collaborative location and activity recommendations with GPS history data. In: *19th International Conference on World Wide Web*, pp. 1029–1038 (2010)
9. Leung, K. W., Lee, D. L., Lee, W.: CLR: a collaborative location recommendation framework based on co-clustering categories and subject descriptors. In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information*, pp. 305–314 (2011)

10. Ye, M., Yin, P., Lee, W.C., Lee, D.L.: Exploiting geographical influence for collaborative point-of-interest recommendation. In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information*, pp. 325–334 (2011)
11. Gao, H., Tang, J., Hu, X., Liu, H.: Exploring temporal effects for location recommendation on location-based social networks. In: *Proceedings of the Seventh ACM Conference on Recommender Systems*, pp. 93–100 (2013)
12. Lee, R., Sumiya, K.: Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection. In: *2nd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, pp. 1–10 (2010)
13. Lee, R., Wakamiya, S., Sumiya, K.: Discovery of unusual regional social activities using geo-tagged microblogs. *World Wide Web* **14**(4), 321–349 (2011)
14. Wakamiya, S., Lee, R., Sumiya, K.: Crowd-based urban characterization: extracting crowd behavioral patterns in urban areas from twitter. In: *3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, pp. 101–109 (2011)
15. Wakamiya, S., Lee, R., Sumiya, K.: Crowd-sourced urban life monitoring: urban area characterization based crowd behavioral patterns from twitter. In: *6th International Conference on Ubiquitous Information Management and Communication*, pp. 261–269 (2012)

Tourism Informatics

Towards Novel Knowledge Based Approaches

Matsuo, T.; Hashimoto, K.; Iwamoto, H. (Eds.)

2015, XI, 164 p. 82 illus., 1 illus. in color., Hardcover

ISBN: 978-3-662-47226-2