

Recommending Prime Spots of a Destination and Time to Visit from Geo-tagged Social Data

Vishal Sharma
Department of Computer Science
Utah State University
Logan, Utah 84322
vishal.sharma@aggiemail.usu.edu

Kyumin Lee
Department of Computer Science
Utah State University
Logan, Utah 84322
kyumin.lee@usu.edu

Jinwook Chung
Department of Computer Science
Utah State University
Logan, Utah 84322
jinwook.chung@aggiemail.usu.edu

Abstract—Planning a trip can be a tedious task. One has to search for what places to visit at a destination (i.e. area) and what time to visit the destination. Sometimes this can be a time-consuming task because there are too much information available, and it is hard for one to choose which information to trust. In this paper we present a recommendation system clustering geo-tagged social data in a destination from each information source – Flickr and Foursquare – and combining the results from these diverse information sources to recommend places to visit. Our experimental results show that our recommendation system automatically suggests prime spots in Yellowstone national park with 0.83 precision and 0.927 NDCG, and in Yosemite national park with 0.8 precision and 0.912 NDCG. In addition, visualizing temporal information of social data helps travelers to decide when to visit a destination.

I. INTRODUCTION

When a tourist wants to start a new trip, it is natural to make plans prior to the trip. But planning a travel is sometimes a complex task which includes many factors such as a travel destination, when to visit the destination, which specific places in the destination to visit, travel duration, etc. A user may search for the related travel guides or ask questions in web-community, but it is still time-consuming to get a good travel plan. Since provided information varies from individual to individual and it becomes hard to choose which one to follow.

To assist travelers, Clements et al. [1] and Shi et al. [2] developed personalized favorite locations or landmark recommendation system based on collaborative filtering approach. Majid et al. [3] developed a context and preference aware travel guide system based on a traveler’s profile containing the traveler’s previous visit history.

But the previous research required collecting a traveler’s previous travel history, did not use multiple information sources to suggest places to visit, or did not consider when is a favorable period to visit an area. In this paper, we are interested to study the following research questions: What if we aggregate and cluster geo-tagged information (e.g., photos with GPS location information, and check-ins)? Can we find popular places in a specific area? Can we build a system recommending popular places automatically? Do we get different recommendation results from different information sources? What if we combine results from multiple information sources, and can we get better recommendation

results instead of using a single information source? Can we help a traveler to decide when to visit the area?

To answer these questions, we make the following contributions in this paper:

- First, we develop a recommendation system based on clustering geo-tagged social data in each destination for recommending prime spots of the destination.
- Second, we compare recommended results from Flickr and Foursquare to see how different information source drives different recommendation results.
- Third, we propose a result combination approach which takes results from Flickr and Foursquare and combines them to produce better recommendation results.
- Finally, we visualize temporal information of social data in each destination and help travelers to determine when to visit the destination.

II. DATASET

To conduct this preliminary research in this paper, we first choose two destinations – Yellowstone and Yosemite national parks. We then collect geo-tagged (latitude and longitude) social data from Flickr and Foursquare. Flickr is one of the most popular sites for sharing photos, some of which contain geo-tagged information, and more than 6 billion photos have been posted to the site. More than 3,100 photos per minute has been uploaded to Flickr. To collect Flickr data, we used a software Bulkr¹. Specifically, we have collected 4,061 photos with meta information of Yellowstone national park, and 1,298 out of them has contained geo tagged information. Similarly, we have collected 4,578 photos with meta information of Yosemite national park, and 1,753 out of them has contained geo-tagged information. Finally, Flickr dataset has contained 1,298 geo-tagged photos and 1,753 geo-tagged photos of Yellowstone and Yosemite national parks, respectively. 86% of the geo-tagged photos were posted between 2011 and 2013.

Foursquare is a leading location sharing service where users post their location information (called check-in). It has more than 4.5 billion check-ins, and 2,000 check-ins per minute are posted. Since Foursquare limits API calls for collecting

¹<http://clipyourphotos.com/bulkr>

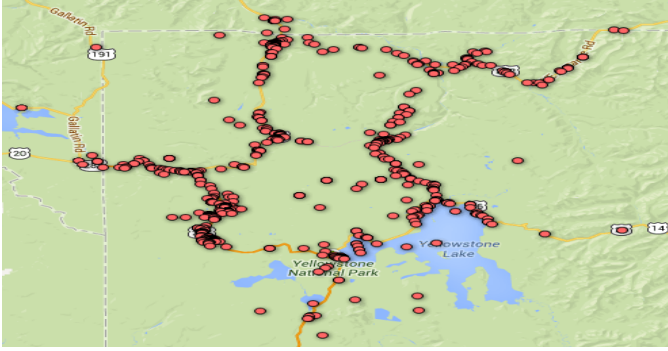


Fig. 1. Geographical location of photos in Yellowstone.

check-ins, we directly retrieved popular 30 places of each of Yellowstone and Yosemite national parks from Flickr² instead of collecting the raw data (i.e., check-ins). Foursquare returns popular 30 places based on check-ins in a specific destination. Figure 1 shows geographical location of the photos of Yellowstone national park in our Flickr dataset, each dot represents a photo, and we have collected enough number of photos for conducting further study.

In the following sections, we mainly use Flickr dataset for recommending prime spots of a destination and time to visit the destination. We then compare prime spots recommended from Flickr dataset with Foursquare data to see whether different information source may drive different recommendation results. Finally we combine the recommendation results from Flickr and Foursquare to see whether we can produce better recommendation result.

III. APPROACHES

The aims of this research work are to (i) recommend prime spots (e.g., Old Faithful, Yellowstone Lake) in a destination (e.g., Yellowstone national park) and (ii) help users to decide when is good time to visit the destination.

Clustering Approach. For the first objective recommending prime spots, we first cluster geo-tagged photos of Flickr by using K -means clustering algorithm [4]. Specifically, the input of K -means clustering algorithm consists of geographical locations of photos in a destination, and a number K which determines how many clusters we want to find. The clustering algorithm initially assigns K geo-tags as centroids, each of which is an initial center of a cluster. In each iteration, the clustering algorithm reassigns a membership of each geo-tag and recomputes a centroid of each cluster. After running the iteration several times, centroids will become stable (not be changed) and then the clustering algorithm stops the iteration (i.e., *residual sum of squares* will not be reduced any more). At that time, we stop the clustering and now we have K clusters as the output of the clustering process. Each of the final centroids of the clusters is a prime spot in the destination. The reason why we chose K -means clustering algorithm compared with other clustering algorithms (e.g., EM clustering algorithm and

Mean Shift clustering algorithm) is that we can control the number of clusters as the output of the clustering process. A use case of this approach is a user wants to find K prime spots in a destination and types the K in our system which recommends K prime spots. Since we get top 30 places in a destination or an area by Foursquare API, we take top K out of 30 places and compare the result with our ground truth described in Section IV-A.

Combination Approach. To test whether combining recommended results from two information sources – Flickr and Foursquare – would produce better recommendation result, we propose a result combination approach. We first obtain K places (centroids) from clustering Flickr data, and sort them by descending order based on $\frac{\# \text{ of geo-tags of the place}}{\text{total \# of geo-tags in the destination}}$. Then, we compute a Flickr score β of each place in a destination by

$$\frac{n - \text{current rank of the place} + 1}{\sum_{n=1}^n n} \quad (1)$$

, where n is a number of recommended places in the destination. n is K in Flickr data. For example, there are 10 places (again, centroids) in Yellowstone, and Old Faithful is ranked 3rd. Its score β is $\frac{8}{55} = 0.145$. A Foursquare score γ of each place in the destination is measured by the same Equation 1. For example, Old Faithful’s rank is the 1st out of 10 places in Yellowstone from Foursquare. Its score γ is $\frac{10}{55} = 0.181$. Finally, given Flickr score β and Foursquare score γ of a place in a destination, we compute the final score of the place as follows:

$$\text{FinalScore}(\beta, \gamma) = \alpha * \beta + (1 - \alpha) * \gamma \quad (2)$$

, where α is 0.5 to give the equal weight to both information sources. The Old Faithful’s final score in the example is $0.5 * 0.145 + 0.5 * 0.181$. After running this process for each place in a destination, we rank all places by their scores in descending order, and recommend the final top places to travelers.

Visualizing Good Time to Visit. For the second objective helping users to decide when is good time to visit the destination, we use temporal information of photos collected from Flickr. In this context, definition of “good time” to visit depends on a user’s preference. The user may want to visit the place in the hottest time (i.e., the most popular time), in less crowded time or in less crowded time during the popular season. In the meantime, we want to help users to know when is favorable period or when is the worst period or season. To achieve the goals, we first extract temporal information of photos and interpret them as previous tourists’ visiting history. We count how many photos (i.e., previous tourists) have been taken in this destination in each day in a range of 1st day and 365th day during a year. Finally, we visualize this frequency of each day, so that a user can determine when to visit the destination, depending on her preference. The visualized graph also reveals that when is favorable period to visit.

²<https://developer.foursquare.com/docs/venues/explore>

TABLE I
PRIME SPOTS IN YELLOWSTONE NATIONAL PARK SUGGESTED BY FOUR POPULAR TOURISM WEBSITES.

Ranking	<i>Yellowstone.net</i>	<i>Trip Advisor</i>	<i>US News Travel</i>	<i>Enjoyyourpark.com</i>
1	Old Faithful	Grand Canyon	Old Faithful	Mammoth Hot Springs
2	Grand Canyon	Lamar Valley	Grand Canyon	Rosevelt Valley
3	Hayden Valley	Lower Yellowstone Falls	Mammoth Hot Springs	Mount Washburn Area
4	Mammoth Hot Springs	Grand Prismatic	Yellowstone Lake	Canyon Area
5	Yellowstone Lake	Midway Geyser	Grand Prismatic	Hayden Valley
6	Norris Geyser	Artist Point	Hayden Valley	Yellowstone Lake
7	Lamar Valley	Upper Geyser	Mount Washburn	West thumb
8	Tower Fall	Hayden Valley	Lower Geyser	Old Faithful
9	Lower Geyser	Old Faithful	Grizzly and Wolf Discovery Center	Madison Area
10	West Thumb	Yellowstone Lake		Norris Area

TABLE II
PRIME SPOTS IN YOSEMITE NATIONAL PARK SUGGESTED BY THREE POPULAR TOURISM WEBSITES.

Ranking	<i>Trip Advisor</i>	<i>Yosemite.com</i>	<i>Travel Channel</i>
1	Glacier Point	Yosemite Falls	Yosemite Falls
2	Half Dome	Half Dome	Half Dome
3	Sentinel Dome	Glacier Point	Toulumne Mea...
4	Tunnel View	Mariposa Grove	El Capitan
5	Yosemite Valley	Ansel Adams Gallery	Yosemite Valley...
6	Traft Point	Tunnel View	Lembert Dome
7	Tioga Pass	Yosemite Village	Tunnel View
8	El Capitan	Mirror Lake	Cathedral Peak
9	Panorama Trail	Tuolumne Meadows	Bridalveil Fall
10	Vernal Fall	Hetch Hetchy	Glacier Point

IV. EXPERIMENTS

In this section, we develop our proposed approaches described in the previous section, run experiments and evaluate the approaches.

A. Automatically Recommending Prime Spots of a Destination

Ground Truth and Evaluation Metric. Since there is no benchmark data describing what is the best prime spots in Yellowstone and Yosemite national parks, we analyzed popular tourism websites such as Trip Advisor, US News Travel, Yellowstone.net and Enjoyyourpark.com for Yellowstone national park, and such as Trip Advisor, Yosemite.com and Travel Channel for Yosemite national park. Then we collected top places suggested by these websites as shown in Tables I and II. We consider commonly recommended places across the websites as the ground truth. In other words, we aim to produce the same result as the ground truth by using our proposed approaches. The most popular places as the ground truth in Yellowstone national park are Old Faithful, Grand Canyon, Mammoth Hot Spring, Hayden Valley, Yellowstone Lake, and Grand Prismatic in order of ranking. Likewise, the most popular places in Yosemite national park are Yosemite Falls, Half Dome, Glacier Point, Tunnel View, El Capitan in order of ranking as shown in Table III.

To evaluate our proposed approaches, we use $\text{precision}@k$ and $\text{NDCG}@k$ [5] as evaluation metrics, where k is a total number of recommended places. $\text{Precision}@k$ is measured by $\frac{|\text{correctly recommended places}|}{k}$. Normalized Discounted Cumu-

TABLE III
PRIME SPOTS IN YELLOWSTONE AND YOSEMITE NATIONAL PARKS SUGGESTED BY POPULAR TOURISM WEBSITES.

Ranking	Yellowstone	Yosemite
1	Old Faithful	Yosemite Falls
2	Grand Canyon	Half Dome
3	Mammoth Hot Spring	Glacier Point
4	Hayden Valley	Tunnel View
5	Yellowstone Lake	El Capitan
6	Grand Prismatic	

lative Gain (NDCG) $@k$ is measured by:

$$\text{NDCG}@k = Z_k \sum_{j=1}^k \frac{2^{R(j)} - 1}{\log(1 + j)} \quad (3)$$

where $R(j)$ is a relevance score of a place, and Z_k is a normalization factor.

To calculate $\text{precision}@k$, if a recommended place is included in the ground truth, we consider the place is correctly recommended. Otherwise, it is not correctly recommended. To calculate $\text{NDCG}@k$, we assign 2 as $R(j)$ if a recommended place is in the ground truth, 1 if a recommended place is in top 10 places (excluding places in the ground truth) in at least one of tourism sites, and 0 if a recommended place is not in top 10 results in any of the tourism sites.

Comparison. To find prime spots of a destination, we first extract geo-tags of photos taken in a destination from Flickr, and run K -means clustering algorithm. We choose K as 6 for Yellowstone national park and 5 for Yosemite national park to compare the results of this clustering with the same number of popular places suggested from the popular tourism websites in

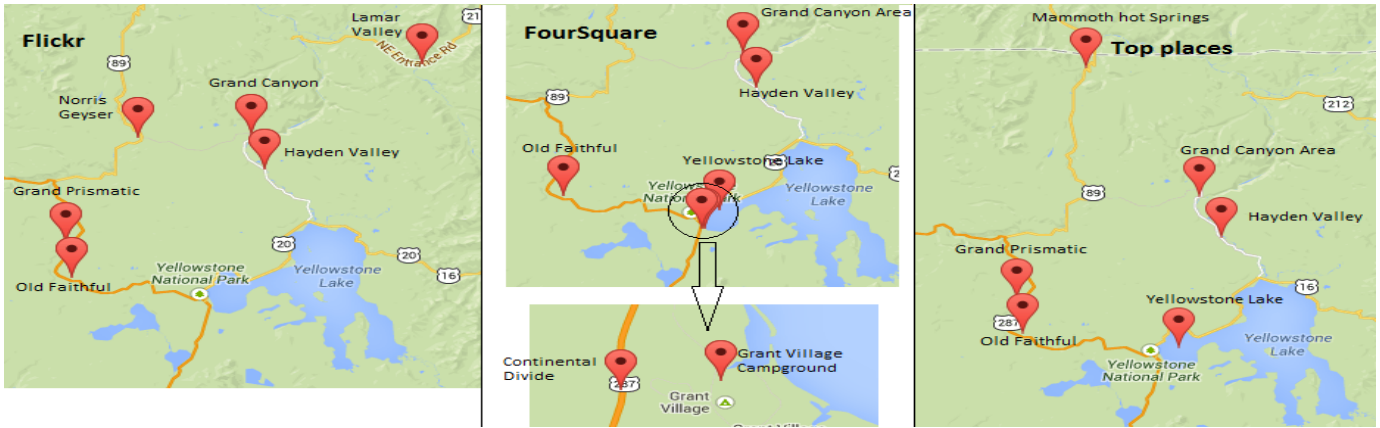


Fig. 2. Top 6 places in Yellowstone from Flickr, Foursquare and popular tourism websites.

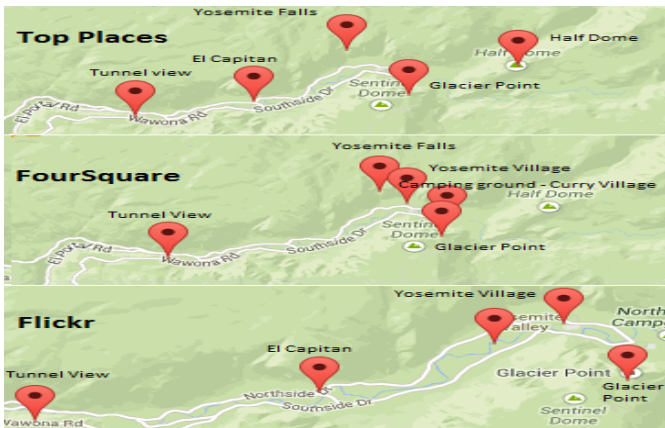


Fig. 3. Top 5 places in Yosemite from Flickr, Foursquare and popular tourism websites.

Table III. The K -means clustering algorithm for Yellowstone national park returns 6 places which are Old Faithful, Hayden Valley, Grand Prismatic, Grand Canyon, Lamar Valley, Norris Geyser as shown in Figure 2. Comparing this result with 6 popular places suggested by the popular tourism sites, we found that the clustering approach was able to correctly identify 4 (Old Faithful, Grand Canyon, Hayden Valley, Grand Prismatic) out of 6 prime spots, achieving $\text{precision}@6 = 0.67$ and $\text{NDCG}@6 = 0.849$. The rest two places (Lamar Valley, Norris Geyser) obtained from the clustering approach are still popular places in Yellowstone.net and Trip Advisor. Similarly, we run K -means clustering algorithm for Yosemite national park by using Flickr data. The clustering algorithm finds 5 places which are Yosemite Village, Tunnel View, Glacier point, El Capitan and somewhere near campground (camp 4) displayed in Figure 3. We compared this result with 5 popular places suggested by the popular tourism sites, and found that the clustering approach was able to correctly identify 3 (Glacier Point, Tunnel View and El Capitan) out 5 prime spots, achieving $\text{precision}@5 = 0.6$ and $\text{NDCG}@5 = 0.642$. Yosemite Village recommended by our clustering approach is still one of

top places in Yosemite.com. Overall, our clustering approach does work well to recommend prime spots in both Yellowstone and Yosemite national parks.

Next, we are interested to study whether different information source, especially Foursquare, gives us different results. As we mentioned in the previous section, we retrieved top 30 places (i.e., the most popular 30 places checked in by Foursquare users) in Yellowstone and Yosemite national parks from Foursquare. We compared top 6 places in Yellowstone and top 5 places in Yosemite from Foursquare with the popular 6 places and 5 places suggested by the tourism websites, respectively. Foursquare check-in data revealed 4 (Old Faithful, Grand Canyon, Yellowstone Lake and Hayden Valley) out of 6 prime spots correctly for Yellowstone national park, achieving $\text{precision}@6 = 0.67$ and $\text{NDCG}@6 = 0.761$. Foursquare check-in data also revealed 3 (Glacier Point, Yosemite Falls, Tunnel View) out of 5 prime spots correctly for Yosemite national park, achieving $\text{precision}@5 = 0.6$ and $\text{NDCG}@5 = 0.642$.

Overall, even though both recommendation results obtained from Flickr and Foursquare contained 4 out of 6 prime spots and 3 out of 5 prime spots of the ground truth for Yellowstone and Yosemite national parks respectively (achieving the same precision), the clustering approach using Flickr data was able to recommend other popular places including Lamar Valley and Norris Geyser for Yellowstone national park and Yosemite Village for Yosemite national park. Specifically, Flickr data based approach achieved higher NDCG than Foursquare data based approach for Yellowstone, but they achieved the same NDCG for Yosemite.

Combination. We combine recommended results from Flickr and Foursquare to see whether we can produce better recommendation results. Sometimes one information source (say, Flickr) may give us biased results because it may be used by certain users. If we combine results from multiple information sources, we may get better and unbiased results. To prove this hypothesis, we run the result combination approach described in the previous section by combining results from Flickr and Foursquare. Specifically, we input 10 as a K value for clustering Flickr data and also get top 10 results from Foursquare

TABLE IV
PRIME SPOTS IN YELLOWSTONE AND YOSEMITE NATIONAL PARKS
RECOMMENDED BY OUR COMBINATION APPROACH.

Ranking	Yellowstone	Yosemite
1	Old Faithful	Yosemite Falls
2	Grand Canyon	Glacier Point
3	Yellowstone Lake	Tunnel View
4	Hayden Valley	El Capitan
5	Grand Prismatic	Yosemite Village
6	Lamar Valley	

data to have a reasonable number of candidate places. Then, we computed each place's β and γ scores based on Equation 1. Finally, we computed each place's combined score based on Equation 2, and selected top 6 and 5 places (getting most highest scores) in Yellowstone and Yosemite respectively from combined results.

Surprisingly, 5 (Old Faithful, Grand Canyon, Yellowstone Lake, Hayden Valley, Grand Prismatic) out of 6 prime spots in Yellowstone national park, and 4 (Yosemite Falls, Glacier point, Tunnel View, El Capitan) out of 5 prime spots in Yosemite national park are the same with the ground truth. Our combined results of both areas are shown in Table IV. The left-over place Lamar Valley recommended for Yellowstone national park and Yosemite Village for Yosemite national park by the combination approach are still popular places even though they are not commonly suggested by the popular tourism sites. Overall, the result combination approach achieved $\text{precision}@6 = 0.83$ and $\text{NDCG}@6 = 0.927$ for Yellowstone, and $\text{precision}@5 = 0.8$ and $\text{NDCG}@5 = 0.912$ for Yosemite. These experimental results showed that the combination approach automatically recommended popular places with high quality results.

In summary, these experimental results confirmed that it is possible for our framework to automatically recommend prime spots of a destination. In a user perspective, he/she can get a list of commonly (and less biased) popular prime spots from our framework without visiting and comparing various tourism sites.

B. Good Time to Visit a Destination

Another interesting research question is "When is good time to visit a destination?". Can we suggest when is a favorable period, and help users to decide when to visit? To answer this research question, as we described in the previous section, we counted number of photos taken in a destination in each day in a range of 1st day and 365th day during a year from Flickr, and plotted a figure for each destination.

Figure 4 represents the temporal information of Yellowstone national park, and we observe that the highest peaks are in between the month of May and mid-August (during the summer), and there are very less people visiting the park between October and March, especially during the winter. This observation might imply that the winter is not a favorable season to visit because most of the roads in and near the national park during the winter are blocked due to heavy snow,

and temperature goes down to 0F (the record low temperature is -66F). After analyzing peaks of the graph, we conclude that the summer is the most popular season, and people are crowded during weekends. By observing this graph, users can determine when to visit Yellowstone.

Figure 5 shows temporal information of Yosemite national park. We can observe that there are several spikes throughout the year which means conditions to visit this park throughout the year are favorable. By verifying the information from USA national forest website, temperature during winter in this park reaches maximum 50F and minimum 30F which are still a fine condition to visit the park even during the winter. After analyzing peaks of the graph we conclude that people have visited this park throughout the year and weekends or holiday season are the most crowded time. In summary, temporal information of social data revealed properties of each destination, and the visualized graph would help users to decide when to visit the place.

V. CONCLUSION AND FUTURE WORK

In this paper, we have developed K -means clustering algorithm based prime spot recommendation system for a destination. We have compared how different information source such as Flickr and Foursquare would drive different recommendation results. As a result, recommendation results from Flickr have been slightly better than results from Foursquare. Then, we have combined the results from both Flickr and Foursquare by using the proposed combination approach which has achieved better recommendation results with 0.83 precision and 0.927 NDCG for Yellowstone national park, and 0.8 precision and 0.912 NDCG for Yosemite national parks. The experimental results confirm that our proposed recommendation system has successfully recommended prime spots of Yellowstone and Yosemite national park compared with commonly popular places among popular tourism sites. In a user's perspective, she can save her time by not visiting and comparing various tourism sites, and get a list of commonly popular prime spots of a destination from our recommendation system.

In addition, visualizing temporal information of Flickr data has revealed when is a favorable time to visit Yellowstone and Yosemite national parks, helping travelers to decide when to visit.

With these positive preliminary results, we plan to apply our recommendation system to various tourist areas to conduct large-scale experiments. We are also interested in grouping areas based on their temporal information to suggest other areas which have similar temporal pattern of another area where a traveler visited and liked before.

REFERENCES

- [1] M. Clements, P. Serdyukov, A. P. de Vries, and M. J. Reinders, "Using flickr geotags to predict user travel behaviour," in *SIGIR*, 2010.
- [2] Y. Shi, P. Serdyukov, A. Hanjalic, and M. Larson, "Personalized landmark recommendation based on geotags from photo sharing sites." in *ICWSM*, 2011.

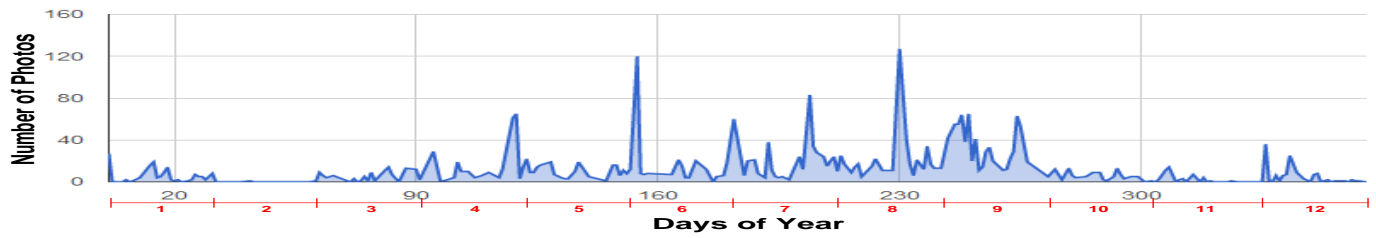


Fig. 4. Temporal distribution of photos from Flickr of Yellowstone.

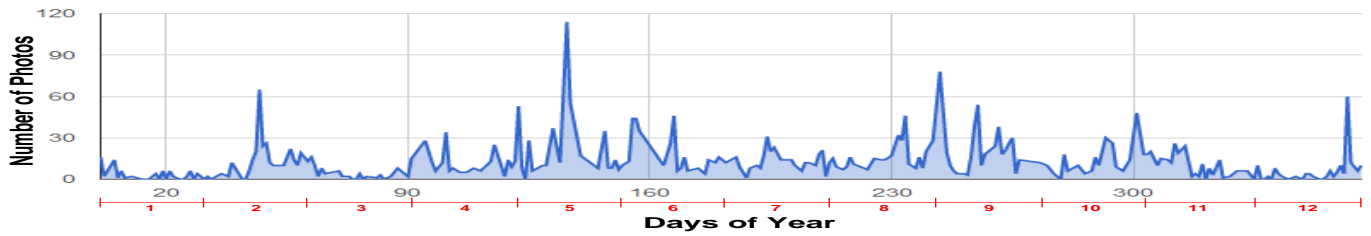


Fig. 5. Temporal distribution of photos from Flickr of Yosemite.

- [3] A. Majid, L. Chen, G. Chen, H. T. Mirza, I. Hussain, and J. Woodward, "A context-aware personalized travel recommendation system based on geotagged social media data mining," *Int. J. Geogr. Inf. Sci.*, vol. 27, no. 4, pp. 662–684, Apr. 2013.
- [4] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. University of California Press, 1967, pp. 281–297.
- [5] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.