A Large-scale Analysis of Regional Tendency of Twitter Photos Using Only Image Features

Tetsuya Nagano, Takumi Ege, Wataru Shimoda and Keiji Yanai
Department of Informatics, The University of Electro-Communications, Tokyo
1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585 JAPAN
{nagano-t,ege-t,shimoda-k,yanai}@mm.inf.uec.ac.jp

Abstract

It is common activities for many people to post photos on open SNSs such as Twitter and Instagram where anyone can see all the posted photos. The difference on idea of the privacy issue at SNSs depends on culture and history. In fact, the people in Eastern Asian such Japan and Korea do not like to post human face photos to SNSs, while Western and South-Eastern-Asian people posts many human face photos. To make this fact clear and explore it deeper, we analyzed geotagged photos collected from the Twitter streams by classifying them into five kinds of rough categories, "people", "building", "document", "scene" and "food". As results, we found there are quite large differences on regional tendency of posted Twitter photos regarding the photo genre distributions.

1 Introduction

Nowadays, posting photos to SNSs such as Facebook, Instagram and Twitter is a part of usual activities in human life. Such photos are not only photos taken at special situations such as trip photos and party photos but also usual photos taken in everyday life like foods and documents. The existing work on SNS photo analysis depends on texts attached to photos in general, because the number of SNS photos is so large like from millions to billions. The analysis using textual information attached to photos was effective for photo sharing sites such as Flickr. Since many of the Flickr users want to have many people see their photos, they tend to attach keywords or tags which expresses the content of photos directly to make it easier to search for their photos. On the other hand, attached texts to photos in SNSs such as Twitter and Instagram do not tend to represent the contents of photos directly in general. This is because the objective of attached messages are not for search but for explaining additional information which cannot be understood by just seeing photos. Therefore, when analyzing the content of SNS photos, promising way to use only image features without no textual information.

Before the era of deep learning, it is not easy to recognize million-scale of photos, because it requires very large-scale computer facilities. However, now CNN running on GPUs enables us to recognize a large number of photos very effectively. In fact, in this work, by using a single server having one GPU, we achieved feature extraction from one million images in only 17 hours.

Thus, in this paper, we analyze a million-scale of Twitter photos by using CNN as a feature extractor. Especially, we try to find differences of the regional tendency of Twitter photos by analyzing geotagged Twitter photos.

2 Related Works

As typical works on Twitter photo, event detection [1, 2] has been studied by using both text analysis and image analysis. In [2], Kaneko et al. proposed a method to detect event photos for geotagged Twitter photos. First, they extract event keywords by keyword burst detection, and classify if photos are related to events or not by using clustering methods. In the experiments, the precision of event photos were 65.5% such as fireworks, concerts, sport games and Christmas party. However, in these works, there was a problem that Twitter photos with no texts or unrelated texts are discarded. On the other hand, in this paper, we analyze a million-scale of Twitter photos without using any textual information, and aim to detect the differences of regional tendency of posted photos to Twitter.

3 Overview of the Proposed Method

In this work, we analyze the regional tendency of rough categories of Twitter photos such as food, people and scenes using geotagged Twitter photos. To do that, (1) we gather a million-scale of geotagged photos from the Twitter stream, (2) extract image features from all of them with CNN running on a GPU, and (3) carry out clustering of them. After clustering of images, (4) we classify only large clusters which have more than one hundred images into one of five typical rough categories or out of them. (5) Finally, we compare the ratios of five categories between eight regions over the world. Note that our objective is to analyze regional tendency of Twitter photos but not to propose a new methodology for geotagged Twitter photo analysis. Therefore, we adopted hand operation for classifying clusters into one of the five rough genres as a noise cluster accurately, although it can be automatically by CNNs.

4 Detail of the Method

4.1 Gathering Photos

We are continuously gathering geotagged photo tweets from the Twitter stream. In the experiments, we used two million geotagged Twitter photos which we had gathered for half years in 2016. Note that we use not Twitter messages but only photos and geotags for large-scale analysis.

4.2 Image Classification

4.2.1 Image Feature Extraction

We use the modified AlexNet (CaffeNet) pre-trained with one million images of one thousand categories using Caffe [3] with a GPU PC. With Caffe and a GPU, we are able to extract CNN features of 1000 images per one minutes. That is, one million images can be processed for about 17 hours. The speed is very important factor for efficient analysis of large-scale images.

The number of the dimension of the extracted features from the FC6 layer of CaffeNet is 4096, which is a relatively high dimension. Therefore, we applied Principal Component Analysis (PCA) to compress the raw features into more compact ones. We compressed 4096-d features into 128-d features in the same as [4]. We examine the differences between the case of 4096-d and the case of 128-d regarding the results of clustering in the next subsection.

4.2.2 Clustering

In this work, to analyze Twitter images in the unsupervised way, which means analysis without textual label information, we use a common clustering methods, K-means clustering. Because CNN features reflect semantic meaning of images, clustering of images with CNN features enables grouping of the images which are semantically similar to each other. K-means requires to decide the number of clusters, k in advance. Thus, we used several numbers for k in the pre-liminary experiments, and we decided k as 100.

To examine the difference on clustering results between the cases with and without PCA-based feature compression, we make a small pre-liminary experiments with 1000 Twitter photos. Figure 1 shows example results of clustering

Figure 1. Clustering results with 4096-d raw CNN features.

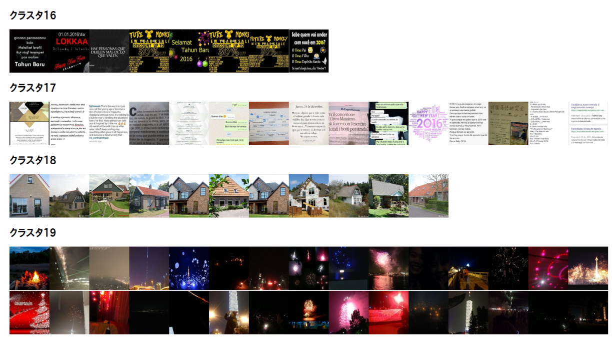

Figure 2. Clustering results with 128-d PCA-compressed CNN features.

with 4096-d raw CNN features, while Figure 2 shows results of clustering with 128-d PCA-compressed CNN features. In both cases, semantically meaningful and coherent clusters were obtained. We confirmed the significant difference was not observed between both results, and we decided to adopt PCA-based feature compression for speeding up of a large-scale analysis of Twitter images. This result is consistent with [4] which showed PCA-compressed CNN features were effective for image search.

5 Experiments

5.1 Dataset

We have collected geotagged Twitter photos, which means the Tweets having geotags and images, from the Twitter stream from January to June in 2016. As results, we built the dataset consisting of 2,161,000 geotagged Twitter images. In addition to geotag, we collected date and time information on all the photos for analyzing seasonal tendency. Note that all the time information were converted into local times depending on the locations the attached geotags indicate.

5.2 Feature Extraction and Clustering

We extracted CNN features from all the images in the dataset, and applied PCA to compress them into 128-d features. After that, we clustered all the two million images using K-means. It is computationally infeasible to carry out K-means for all the two million images. Thus, we sampled one tenth of them, calculated the center of the clusters, and assigned the rest of images into the nearest center of the clusters. We performed K-means clustering with 100, 200, and 300 as the number of clusters, K, and we examined the differences among the clustering results. For our objective which is classifying each cluster into rough categories or genres, the significant differences were not observed. Then, we decided to fix 100 to K for the rest of experiments.

5.3 How to Analyze

For regional analysis, we used the nine regional divisions: East Asia, North America, South America, Europe, Africa, Middle East, South Asia and South-East Asia, Oceania. Note that China was excluded from the target regions, since Twitter is prohibited to use in the Chinese region.

After clustering, we obtain clusters of the images which were semantically similar to each other as shown in Figure 2. To analyze the tendency of posted photos, we classify the obtained photo clusters into one of the pre-selected photo genres. As photo genres, we use “people”, “building”, “document”, “scene” and “food”. These genres are decided based on the observation of clustering results of each of the nine regions. After genre classification of clusters, we compare the genre distributions of the images in each region regarding all the region to clarify the difference of regional tendency of Tweet photos. To remove noise images, we excluded small clusters having less than one hundred images for analysis.

By using five photo genres, we analyze the regional tendency of posted photos. We classify clusters by hand into one of the five genre. Although we can build a classifier to do that automatically, at this time we regard accuracy as most important rather than fully automatic processing. Thus, we check one by one to exclude noisy clusters which contain multiple genre images or no images corresponding to one of the five genres. In addition, sometimes there are clusters within which almost all the image are identical, which means a kind of spam image clusters. We also excluded them by hand as well.

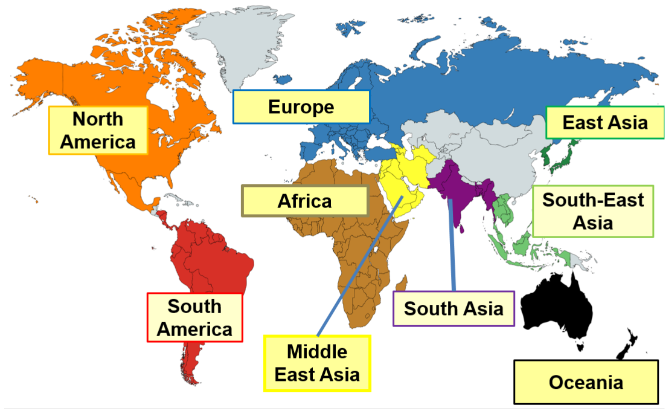

Figure 3. Nine regions over the world.

Table 1. The number of each of the five typical photo genres for the nine regions.

region	Total	People	Building	Document	Scene	Foods
East Asia	65,530	459	20,172	8,973	8,217	27,709
North America	124,141	40,231	38,336	12,015	11,314	22,245
South America	76,481	51,471	3,613	16,416	2,381	2,600
Europe	183,250	63,942	37,651	34,511	22,018	25,128
Africa	24,101	17,409	284	6,408	0	0
Middle East Asia	44,184	11,681	10,216	12,334	4,123	5,830
South Asia	13,263	5,424	277	7,214	295	53
South-East Asia	145,088	73,265	15,862	13,120	4,323	38,518
Oceania	4,695	1,049	144	2,214	953	335

5.4 Analysis of Regional Tendency of Photo Genres

Table 1 and Figure 4 shows the ratios of five typical photo genres of geotagged Twitter Photos. From this result, we found no people photos were posted in East Asia, and instead many building and food photos, the total ratio of which were more than 70%, were posted (Figure 5). In North America, the ratio of people and building (Figure 6) were high, which was more than 60%. Regarding South America, people photos are by far the most popular genre (67%) (Figure 7). In Europe, the number of posted photos was the most and the genres were well balanced. In Africa, almost no building, scene and food photos were posted and people photos occupied 70%. In Middle East, although the number of posts were fewer than Europe, all the five genres were balanced as well (Figure 8). In South Asia, more than half of the photos were document photos. This tendency was not observed in other regions. In South-East Asia, in the same way as South America and Africa, people photos are the most and in addition food photos was the second most. Regarding the absolute number of food photos, South-East Asia was the best. Regarding Oceania, the number of the photos was the least among the nine regions. Note that in Oceania, many map photos were posted, and we classified map photos as document photos. That is why the ration of document photos was the best in Oceania.

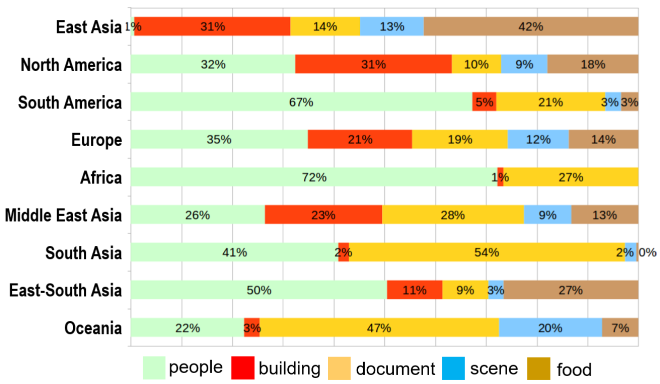

Figure 4. The ratios of five photo genres in the nine regions over the world.

Figure 5. "Food" in East Asia.

Figure 6. "Building" in North America.

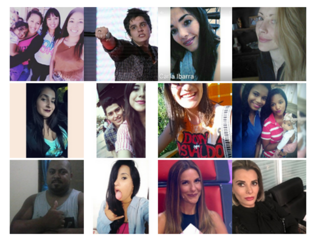

Figure 7. "People" in South America.

Figure 8. "Document" in Middle East.

6 Discussion

In this experiments, we found the five typical genres which were appropriate for tendency analysis. However, because we discarded small clusters having less than one hundred images in this work, no food and scene photos were left in the photo set of Africa.

From the typical genre distributions, we found the regional tendencies that the ratio of food photos was relatively high in East Asia and East-South Asia, while the ratio of people photos are exceptionally high in South America, South Asia and East-South Asia. In Europe and Middle-East, the typical five genres were well balanced. In addition, almost no people photos are posted in East Asia, and in South Asia half of the posted photos are document photos. In this work, we limited to use only geotagged photos. In general, posting geotagged photos to SNSs means making the current location of the user open to the public. Therefore, we expect that the users in East Asia tend to refrain from posting geotagged people photos to Twitter stronger than normal non-geotagged photos. These are the finding of the analysis in this paper. From these observation, we can estimate that the users in East Asia enjoy posting food photos and the uses in South America, South Asia and East-South Asia like to post people photos without caring privacy issue.

7 Conclusions

In this work, we have gathered geotagged Twitter photos from the Twitter stream for half a year, extracted CNN features from all the gathered images using a GPU, and carried out clustering for them. After that, we classified each cluster into one of the five typical genres regarding each of the nine regions over the world, and analyzed the differences of the tendency of the photo genres of posted photos. As results, we have found the prominent differences on regional tendency of photo genres. For example, the users in East Asia refrain from posting people face photos to Twitter, while the users in East-South Asia and South America like to post face photos.

On the other hand, because we limited only five typical photo genres into which photos were categorized, we probably missed to find out some photos having regional characteristics. In addition, by making typical genres more fine-grained, we might be able to discover more interesting tendency.tendencies. For example, "people photo" can be divided into "selfy" and "group photo", and we can get to know the answer for which region posting of selfy is the most popular.

References

[1] M. Schinas, S. Papadopoulos, P. A. Mitkas, and Y. Kompatsiaris. Visual event summarization on social media using topic modelling and graph-based ranking algorithms. In *Proc. of International Conference on Multimedia Retrieval*, pages 203–210, 2015.

[2] T. Kaneko and K. Yanai. Event photo mining from twitter using keyword bursts and image clustering. *Neurocomputing*, 172:143–158, 2016.

[3] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678. ACM, 2014.

[4] A. Babenko, A. Slesarev, A. Chigorin, and A. V. Lempitsky. Neural codes for image retrieval. In *Proc. of European Conference on Computer Vision*, pages 584–599, 2014.