Analyzing Regional Food Trends with Geo-tagged Twitter Food Photos

Kaimu Okamoto Keiji Yanai
Department of Informatics, The University of Electro-Communications, Tokyo
1-5-1 Chofugaoka, Chofu-shi, Tokyo, Japan 182-8585
{okamoto-ka,yanai}@mm.inf.uec.ac.jp

Abstract—Twitter is a world-wide popular SNS where many images are being posted along with tweet messages and location information. It is known that large regional differences exist regarding posted images. The differences are expected to be prominent especially for food images, since there are many kinds of regional foods over the world. However, the regional difference on food images over Twitter has not been explored so far. Therefore, in this paper, to make the difference clear, we analyzed geo-tagged food image gathered from the Twitter stream on the basis of the six regions and 17 kinds of rough food categories. For image analysis, we used only image features without any textual analysis, since Twitter messages are not always directly related to image contents. In addition, we visualized discriminative parts of local food images by applying the visualization method of CNNs.

Index Terms—food image analysis, Twitter, social multimedia analysis, Twitter photo analysis, geotagged photo analysis

I. INTRODUCTION

For these ten years, social networks have become widespread, many people have posted tweets and images on Twitter. The posted images are a wide variety of images including in people, landscapes and food, and they are familiar images of human life. Some of them have geotags which indicate the location of the origins of the Tweet messages. In addition, there are also regional differences in the posted images, because peoples' lifestyles differ from each region. Since foods are essential to human life, it can be expected that the difference will be larger. However, the regional difference on food images over Twitter has not been explored so far.

Thus, in this paper, in order to discover the regional food trends, we analyze the food images selected from large-scale geo-tagged Twitter images using CNN food-specialized features and clustering. Also, we tried to identify discriminative parts of food images by applying the CNN visualization method, Grad-CAM [1]. To analyze food images, we used only image features without any textual analysis, since Twitter messages are not always directly related to image contents.

II. RELATED WORKS

As typical works on Twitter photo, event detection [2], [3] has been studied by using both text analysis and image analysis. Because these works heavily relied on text analysis, they had a problem that Twitter photos with no texts or unrelated texts were discarded. On the other hand, Nagano et al. [4] analyze a million-scale of

978-1-7281-4673-7/19/$31.00 ©2019 IEEE

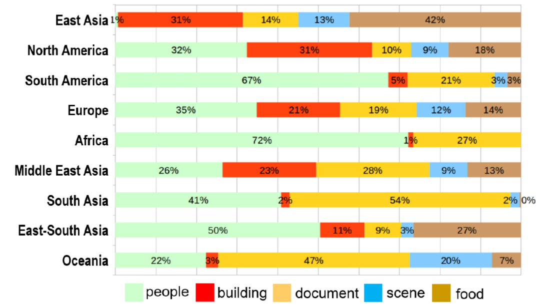

Fig. 1. The ratios of five representative categories of Twitter photos (Nagano et al. [4]).

Twitter photos without using any textual information, and found the differences of regional tendency of Twitter photos. First, CNN features were extracted using Twitter images collected for half a year. Next, they clustered those images by K-means and classified the created clusters into five image categories for nine regions over the world. After categories classification, they analyzed the regional tendency.

Fig.1 shows regional tendency of nine regions over the world with the ratios of five representative Twitter image categories. In South America and Africa, the ratio of people images was very large, while in East Asia the ratio of food photos was relatively larger. They analyzed regional tendency of Twitter photos for all kinds of images without specific targets. On the other hand, in this paper, we focus on only specific category, food, and like to perform more detailed analysis of regional tendency.

III. OVERVIEW OF THE PROPOSED METHOD

We analyze regional trends of several representative food categories using geo-tagged images gathered from the Twitter stream. The procedure consists of the following four steps as shown in Fig.2:

1) Classify food and non-food photos with a fine-tuned food/non-food classifier for Twitter photos.
2) Extract food-specialized CNN features with a CNN fine-tuned with Triplet loss [5].
3) Cluster food images
4) Analyze world food trends and visualize regional parts in food images

IV. DETAIL OF THE METHOD

A. Classification of food/non-food photos

In this work, we use the raw photos gathered from Twitter. To select only food images, we need to prepare

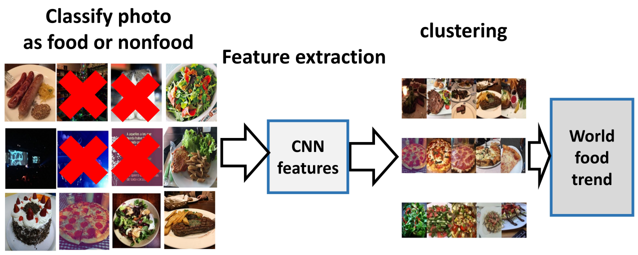

Fig. 2. The proposed procedure for analyzing regional food trends.

a food/non-food classifier. We fine-tune the ResNet [6] pre-trained with ImageNet using food image datasets and non-food image datasets. As food image datasets, we use all the images in both the UECFOOD100 [7] dataset containing 14,000 images and the Food-101 [8] dataset containing 101,000 images. As non-food image datasets, we use about 120,000 randomly extracted images from the ILSVRC2012 version of the ImageNet 1000-class dataset and about 13,000 non-food images used in the work of Kawano et al. [9]. We discriminate food images from non-food images with this classifier, and discard all the non-food images.

B. Food image feature extraction

To perform clustering food images, we use CNN features of the pre-trained VGG16 [10]. Since we focus on only food photos, we fine-tune the VGG16 so that it can extract food-specialized features which can discriminate small differences on various kinds of food images. To do that, we use the Triplet network [5] as a method of feature learning. It is known that the Triplet network can improve the image retrieval accuracy of food images [11], which is expect to allow similar kinds of food images to have closer features to each other than the CNN features pre-trained with the ImageNet dataset. The Triplet network [5] is trained by a triplet of a query image, a positive image and a negative image. Triplet loss [5] is used for the loss function to be optimized so that the Euclidean distance between the query image becomes smaller and one between the query image and the positive image becomes larger. The equation for Triplet loss L_T used in the work is:

$$L_T = max(0, g + ||f_+ - f||_2 - ||f_- - f||_2), \qquad (1)$$

where f, f_+, f_- represents CNN features of the query image, the positive image, and the negative image, respectively. The constant g is the margin between the two distances, using $g = 0.3$ for the experience. By following Shimoda et al. [11], we use a classification loss as well. The entire loss function combining the triplet loss and the class classification loss is as follows:

$$L = L_T + L_C \qquad (2)$$

For fine-tuning of the VGG16 network for food feature extraction with the triplet loss, we used the images of both UECFOOD100 and FOOD-101 by integrating them into one training image dataset. We extract the 4096-d CNN features from the FC7 of the trained network, normalize them with L2 norm, and then compress them with PCA to 128-d vectors to make large-scale clustering feasible. We follow Nagano et al. [4] regarding the PCA-based CNN feature compression.

C. Clustering and analyzing of regional tendency

In the same as Nagano et al. [4], 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 food CNN features reflect semantic meaning of food images, clustering of food images with food CNN features enables grouping of the food images which are semantically similar to each other [11].

We perform K-means clustering for the food images in each of the pre-defined regions using the food-specialized CNN features. After clustering, we obtain clusters of the images which were semantically similar to each other. To analyze the tendency of posted food photos, we classify the obtained photo clusters into one of the pre-selected representative food categories. As the representative food categories, we use 17 kinds of foods such as "meat", "noodle", "rice" and "bread". These categories are decided based on the observation of clustering results of each of the regions. After representative category classification of clusters, we compare the food distributions of the images in each region regarding all the region to clarify the difference of regional tendency of Tweet photos. In addition, the analysis of area distribution of each of the representative foods is also performed.

Although each cluster contains semantically similar food images, their appearances are diverse in general. In this work, we rank images in each cluster using a similarity-based image ranking method, VisualRank [12]. To compute VisualRank, we prepare a similarity matrix S as being dot product of image feature vectors. The equation to calculate the VisualRank is as follows:

$$r = \alpha S r + (1 - \alpha) p \qquad (0 \leq \alpha \leq 1) \qquad (3)$$

where S is the column-normalized similarity matrix of images, p is a damping vector, and r is the ranking vector each element of which represents a ranking score of each image. α plays a role to control the extent of effect of p. In this paper, we set α as 0.8. The final value of r is estimated by updating r iteratively with Eq.(3). Because S is column-normalized and the sum of elements of p is 1, the sum of elements of ranking vector r also stays 1.

D. Visualization of regional parts

Finally, we find out which parts of the food images corresponds to regional features. by using a visualization method of CNNs, Grad-CAM [1]. Grad-CAM [1] is a method for specifying which part of an image affects classification based on the class score and the gradient in the final output value of the convolutional layer. In Eq.(4), for the score c in the class y^c, a gradient ∂A^k_{ij}, which is a differentiated value of the feature map A^k_{ij} of the final convolutional layer, is averaged in all pixels i, j for each channel k and weight α^c_k is determined. Then, in Eq.(5), the mask image L^c is generated by applying the

TABLE I
Image statistics.

region	East Asia	South Asia	South-East Asia	North America	South America
all	488,609	79,179	609,671	786,093	410,086
food	69,826	999	59,459	23,867	8,211
region	Europe	Africa	Oceania	Middle East	TOTAL
all	791,095	150,550	34,543	434,790	3,784,616
food	14,572	1,229	914	14,115	193,192

TABLE II
List of the representative food categories.

meat	noodles	sweets	rice	bread	beverage
salad	fried-food	seafood	soup	fastfood	stir-fried
egg	curry	flour	Chinese cuisine	cheese	

function $ReLU(x) = max(x, 0)$ to the total value of $a_k^c A^k$ in the channel. By multiplying the mask image L^c and the original image, it is possible to see which part of the image is active.

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (4)$$

$$L_{Grad-CAM}^c = ReLU\left(\sum_k a_k^c A^k\right) \quad (5)$$

To visualize regional features, we train region classifiers on some of the pre-defined food categories. By apply Grad-Cam to the region classifiers, we can detect region discriminative parts of foods. We will shows same examples in the next section.

V. EXPERIMENTS

A. Twitter Geotagged Food Images

In this work, we used the log of the Tweets containing both photos and geo-tags we collected in 2016 for whole a year. With the food/non-food classifier, we selected 220,000 food images from 3.78 million geotagged Twitter images contained in the Twitter log. After that, duplicate images were removed based on simple color histogram image features. Finally, we obtained about 190,000 food images. The number of all images and food images for each region are shown in Table I. In the analysis, we used food images in six regions, East Asia, South-East Asia, North America, South America, Europe and the Middle East excluding three regions, South Asia, Africa and Oceania, because the number of geotagged food images in the three regions were around 1,000 which were not enough as shown in the table.

B. Representative Food Categories

We performed K-means clustering for each of the seven region with the CNN features extracted from the food image, and classified the obtained clusters into the representative food categories. The food categories assigned to the cluster are 17 categories shown in Table II. These categories are decided based on the observation of clustering results of each of the regions. The datasets used for training the food/non-food classifier, UECFOOD100 [7] includes images such as ”ramen” and ”pilaf”, and FOOD-101 includes images such as apple_pie and french_onion_soup. We selected 17 representative food categories so that they roughly covered all the categories of both the datasets.

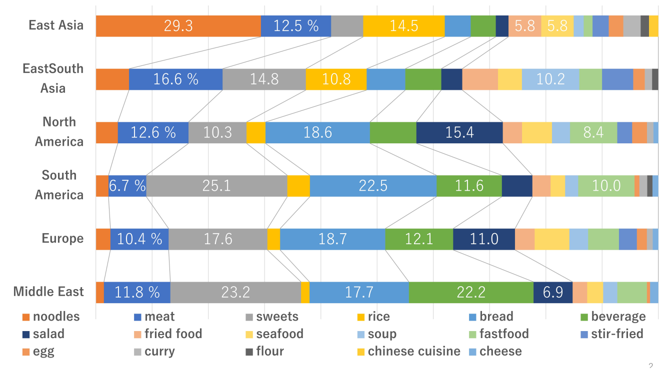

Fig. 3. Food distribution on the six regions. White letters indicate the percentage of the top 5 categories on each region.

Fig. 4. Examples of “noodles” images in East Asia.

Fig. 5. Examples of “soup” images in South-East Asia.

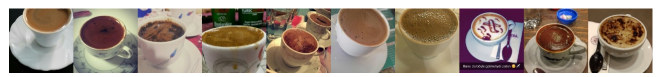

Fig. 6. Examples of region-specific “beverages” in Middle East.

C. Food Trend Analysis by Region

First, we analyzed the ratio of the food categories on each region. The ratios on each region are shown in Fig.3 where the percentages of the top five food categories among 17 categories are shown in bold.

In East Asia, the food categories such as “noodles” and “rice”, which are relatively rare in the other regions, are included at the top. As the possible reason for the high proportion of noodles to 29.3%, there are many images of ”ramen“ (which are very popular noodles in Japan.) as shown in Fig.4, and in addition to ”ramen“, various kinds of “noodle” images such as rice noodles and buckwheat noodles are seen in East Asia. This is a salient characteristics of East Asia.

In South-East Asia, the soup is ranked at the top, and many dishes made of vegetables and meats in a soup like Fig.5 were seen.

In North America, when comparing trends in Asia, the categories for bread and salad tended to be more. On the other hand, when compared with South America and Europe, it turned out that 4 items of the top 5 items of the categories are the same, and two regions had similar to each other regarding food tendency.

In the Middle East, the top five food categories are the same as Europe. However, in the region there are many brown coffee as shown in Fig.6, which is a unique type of coffee to Middle East.

D. Regional Trend Analysis by Food Images

Next, we performed an area analysis in each of the food categories. We calculated the number of images in each

TABLE III
Area distribution of each of the food categories (%).

food	EA	ESA	NA	SA	EU	ME
all	38.77	30.71	12.52	3.72	7.41	6.85
meat	36.32	38.34	11.65	1.87	5.76	6.03
noodles	81.00	12.90	3.44	0.60	1.37	0.70
sweets	18.68	38.37	10.75	7.87	11.00	13.34
rice	57.40	33.98	4.25	1.53	1.71	1.05
bread	18.60	21.92	23.85	8.71	14.41	12.50
beverage	22.92	26.02	13.50	5.72	11.86	19.97
salad	16.21	21.15	35.10	3.72	15.07	8.74
fried food	43.50	37.66	8.13	2.32	5.00	3.39
seafood	45.13	26.45	13.29	1.95	9.25	3.92
soup	14.67	66.26	8.39	1.82	5.23	3.63
fastfood	15.17	30.92	25.71	9.23	10.09	8.87
stir-fried	32.98	50.30	9.71	0.00	7.00	0.00
egg	46.34	29.25	15.0	1.49	6.10	1.79
curry	64.19	21.33	7.56	2.79	4.12	0.00
flour	60.40	35.93	0.00	3.66	0.00	0.00
chinese cuisine	100.00	0.00	0.00	0.00	0.00	0.00
cheese	0.00	0.00	33.72	12.03	23.08	31.16

EA:East Asia, ESA:South-East Asia, NA:North America
SA:South America, EU:Europe, ME:Middle East

region for each food and the ratio of regions for each of the food categories as shown in Table III.

Among them, we explain some food categories that were 10.00% or more regarding region ratio. First of all, the ratio of "noodles" in East Asia was the highest, reaching 81.00%. This is considered to be attributable to the fact that many ramen are eaten in East Asia, and various noodles other than ramen are also eaten.

In South-East Asia, "soup" is the highest proportion, reaching 66.3%. It is thought that Southeast Asian soups show many foods that use other categories of food such as vegetables, meat and noodles, and they are dishes that contain a wider variety of ingredients compared to other regions.

In "beverages", South-East Asia shows the highest proportion. However, looking at the proportion in Middle East, the result is 19.97%, which is larger than the proportion to the whole image of Middle East, and it can be seen that the local tendency in the beverage is affected to some extent.

"Salad", "bread" and "cheese" shows the highest ratio in North America, which are typical Western foods.

E. Visualization of Regional Features

Finally, we visualized region discriminative parts for some food images. We trained a network which classify regions of food images for visualization. The VGG16 [10] network was trained with the output of 6 regions, East Asia, Southeast Asia, North America, South America, Europe, and the Middle East, using all the images of each of the representative categories. The model was used to visualize with Grad-CAM [1]. Due to space limitation, in the paper, we show the visualization results on only meat and bread.

Fig.7 and Fig.8 shows the examples in which regional feature are successfully detected. Fig.7 shows "meat" images in South-East Asia, which show that the meat ball parts are characteristic to the region. Fig.8 shows "sweet" images in Middle East, which is called "baklavaa". Regional features are seen in the green part of puff pastry.

VI. CONCLUSIONS

In this study, we analyzed regional food tendency on only the geo-tagged food images without text data. As a result of the analysis, we could see the unique food tendency on each region and similar food image tendency with other regions, and discover the area tendency on each representative foods. In addition, we found the region discriminative parts that showed unique shapes or appearances of the regional foods.

In this work, we were unable to analyze regional trends in the three regions, South Asia, Africa, and Oceania. This was because there were extremely few food images posted on Twitter in the three areas. For future work, we plan to integrate image data from other photo SNSs such as Instagram and Weibo with Twitter images for more comprehensive analysis. Since we have been collecting Twitter photo logs from 2011 continuously, we also plan to analysis transition of Twitter photo trends for about ten years.

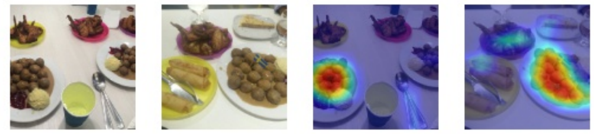

Fig. 7. The Grad-CAM results for "meat" images in South-East Asia, which show that the meat ball parts are region-discriminative parts.

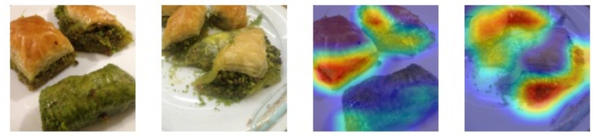

Fig. 8. The Grad-CAM results for "sweet" images in Middle East, which show that the green parts are region-discriminative parts.

References

[1] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In CVPR, 2016.

[2] 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 ICMR, pages 203–210, 2015.

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

[4] T. Nagano, T. Ege, W. Shimoda, and K Yanai. A large-scale analysis of regional tendency of twitter photos using only image features. In MIPR, 2019.

[5] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, B. Philbin, J. Chen, and Y. Wu. Learning fine-grained image similarity with deep ranking. In CVPR, 2014.

[6] K. He, X. Zhang, and J Ren, S.and Sun. Deep residual learning for image recognition. In CVPR, 2015.

[7] Y. Matsuda, H. Hoashi, and K. Yanai. Recognition of multiple-food images by detecting candidate regions. In ICME, 2012.

[8] B. Lukas, G. Matthieu, and V.G. Luc. Food-101 – mining discriminative components with random forests. In ECCV, 2014.

[9] Y. Kawano and K. Yanai. Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. In ECCVWS, 2014.

[10] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

[11] W. Shimoda and K. Yanai. Learning food image similarity for food image retrieval. In BIGMM, 2017.

[12] Y. Jing and S. Baluja. Visualrank: Applying pagerank to large-scale image search. IEEE PAMI, 30(11):1870–1890, 2008.