Twitter Food Photo Mining and Analysis for One Hundred Kinds of Foods

Keiji Yanai and Yoshiyuki Kawano

Department of Informatics, The University of Electro-Communications
1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585 Japan
{yanai,kawano-y}@mm.inf.uec.ac.jp

Abstract. So many people post photos as well as short messages to Twitter every minutes from everywhere on the earth. By monitoring the Twitter stream, we can obtain various kinds of images with texts. In this paper, as a case study of Twitter image mining for specific kinds of photos, we describe food photo mining from the Twitter stream. To collect food photos from Twitter, we monitor the Twitter stream to find the tweets containing both food-related keywords and photos, and apply a "foodness" classifier and 100-class food classifiers to them to verify whether they represent foods or not after downloading the corresponding photos. In the paper, we report the experimental results of our food photo mining for the Twitter photo data we have collected for two years and four months. As results, we detected about 470,000 food photos from Twitter. With this data, we made spatio-temporal analysis on food photos.

Keywords: photo tweet, Twitter mining, food photos, food classifier, UEC-FOOD100.

1 Introduction

Twitter is a unique microblog, which is different from conventional social media in terms of its quickness and on-the-spot-ness. Many Twitter's users send messages, which is commonly called "tweets", to Twitter on the spot with mobile phones or smart phones, and some of them send photos and geotags as well as tweets. Most of the photos are sent to Twitter soon after taken. In case of food photos, most of them are taken just before eating on the spot.

In this paper, we focus on food photos embedded in Tweets as a case study on a large-scale Twitter photo analysis. Food is one of frequent topics in Tweets with photos. In fact, we can see many food photos in lunch and dinner time in the Twitter stream.

Then, in this paper, by combining keyword-based search and food image recognition, we mine food photos from the Twitter stream. To collect food photos from Twitter, we monitor the Twitter stream to find the tweets containing both food-related keywords and photos, and apply a "foodness" classifier and 100-class food classifiers to them to verify whether they shows foods or not after downloading the corresponding photos. We used the state-of-the-art Fisher Vector coding with HoG and Color patches for food classifiers which is slightly modified with the rapid food recognition method for mobile environments proposed in [6], and trained them with the UEC-FOOD100 dataset [9][1]

[1] http://foodcam.mobi/dataset/

W.T. Ooi et al. (Eds.): PCM 2014, LNCS 8879, pp. 22–32, 2014.

which consists of 100 kinds of foods commonly eaten in Japan. Since we employ the improved method of the real-time mobile recognition, it takes only 0.024 seconds to recognize one image and it achieved about 83% classification rate within the top five candidates.

In the experiments, we report the results of our food photo mining on 100 kinds of foods in the UEC-FOOD100 dataset from the photo tweet log data we have collected for two years and four months. As results, we detected about 470,000 food photos from Twitter with about 99% accuracy. With this data, we have made spatio-temporal analysis on food photos. In addition, we have implemented the real-time food photo detection system from the Twitter stream.

2 Related Works

In this section, we mention about some representative works on Twitter photo mining.

As a representative work on Twitter mining, Sakaki et al. [12] regarded Twitter users as social sensors which monitor and report the current status of the places where the users are. They proposed a system which estimates the location of natural events such as typhoons and earthquakes. They used geotagged tweets to estimate event locations but no photos attached to tweets.

As early works on tweet photos, Yanai proposed World Seer [13] which can visualize geotagged photo tweets on the online map in the real-time way by monitoring the Twitter stream. Nakaji et al. [10] proposed a system to mine representative photos related to the given keyword or term from a large number of geo-tweet photos. They extracted representative photos related to events such as "typhoon" and "New Year's Day", and successfully compared them in terms of the difference on places and time. However, their system needs to be given event keywords or event term by hand. Kaneko et al. [5] extended it by adding event keyword detection to the visual Tweet mining system. As results, they detected many photos related to seasonal events such as festivals and Christmas as well as natural phenomena such as snow and Typhoon including extraordinary beautiful sunset photos taken around Seattle. All of these works focused on only geotagged tweet photos.

On the other hand, Chen et al. [2] treated with photo tweets regardless of geo-information. They analyzed relation between tweet images and messages, and defined the photo tweet which has strong relation between its text message and its photo content as a "visual" tweet. In the paper, they proposed the method which is based on the LDA topic model to classify "visual" and "non-visual" tweets. However, because their method was generic and assumed no specific targets, the classification rate was only 70.5% in spite of two-class classification. Because we focus on only food photos unlike their work, we use specialized object classifiers and have achieved very high classification accuracy. By using the method for real-time mobile food recognition, we can apply it to more than one million and seven hundred thousand tweet images and implement a real-time food Tweet photo detection system.

3 Overview

In this section, we describe an overview of the proposed method to mine food photos from the stored Twitter logs as well as the Twitter stream. We employ the following three-step processing.

(1) We perform keyword-based search with the names of target foods over a set of photo Tweets.

(2) We apply a newly-proposed "foodness" classifier to the tweet photos selected by the keyword-based search for classifying then into either of "food" or "non-food" photo.

(3) We apply individual food classifiers corresponding to the food names. In the experiments, we prepared multi-class discriminative classifiers trained by SVM with the UEC-FOOD100 dataset in the one-vs-rest manner.

The food classifiers employed in the third step is a slight modification of the method for mobile food recognition proposed in [6], while the foodness classifier is newly proposed for removing non-food photos.

4 Detail of the Proposed Method

4.1 Keyword-Based Photo Tweet Selection

In the first step, we select photo tweets by keyword-based search with the names of the target foods. We search tweet message texts for the words of the target food names.

As the target foods, we used 100 kinds of foods in the UEC-FOOD100 dataset in the experiments. Because the UEC-FOOD100 dataset includes common foods in Japan such as ramen noodle, curry, and sushi, we searched only photo tweets the message texts of which are written in Japanese language. We can easily select them by checking the language attribute of each tweet obtained from the Twitter Streaming API.

4.2 Foodness Classifier

We construct a "Foodness" Classifier (FC) for discriminating food images from non-food images. FC evaluates if the given image is a food photo or not. We use FC to remove noise images from the images gathered from the tweet photos selected by the food names.

We construct a FC from the existing multi-class food image dataset. Firstly, we train linear SVMs [4] in the one-vs-rest strategy for each category of the existing multi-class food image dataset. As image features, we adopt HOG patches [3] and color patches in the same way as [6]. Although HOG patches are similar local features to SIFT [8], HoG can be extracted much faster than SIFT. Both descriptors are coded by Improved Fisher Vector (IFV) [11], and they are integrated in the late fusion manner. We perform multi-class image classification in the cross-validation using the trained liner SVMs, and we build a confusion matrix according to the classification results. In the experiments, we used 64 GMMs for IFV coding and two-level spatial pyramid [7], which is much improved from mobile food recognition [6] in terms of the feature dimension.

Table 1. 13 food groups and their member foods for the "foodness" classifier

type of food group	food categories
noodles	udon nooles, dipping noodles, ramen
yellow color	omlet, potage, steamed egg hotchpotch
soup	miso soup, pork miso soup, Japaneses tofu and vegetable chowder
fried	takoyaki, Japaneses-style pancake, fried noodle
deep fried	croquette, sirloin cutlet, fried chicken
salad	green salad, macaroni salad, macaroni salad
bread	sandwiches, raisin bread, roll bread
seafood	sashimi, sashimi bowl, sushi
rice	rice, pilaf, fried rice
fish	grilled salmon, grilled pacific saury, dried fish
boiled and seasoned	seasoned beef with potatoes simmered ganmodoki seasoned beef with potatoes
sauteed	sauteed vegetables, go-ya chanpuru, kinpira-style sauteed burdock
sauce	stew, curry, stir-fried shrimp in chili sauce

Secondly, we make some category groups based on confusion matrix of multi-class classification results. This is inspired by Bergamo et al.'s work [1]. They grouped a large number of categories into superordinate groups the member categories of which are confusing to each other recursively. In the same way, we perform confusion-matrix-based clustering for all the food categories. We intend to obtain superordinate categories such as meat, sandwiches, noodle and salad automatically. As results, in the experiments, we obtained 13 food groups as shown in Table 1.

To build a "foodness" classifier (FC), we train a linear SVM of each of the superordinate categories. The objective of FC is discriminating a food photo from a non-food photo, which is different from the objective of the third step for discriminating a specific food photo from other kinds of food photos. Therefore, abstracted superordinate categories are desirable to be trained, rather than training of all the food categories directly. The output value of FC is the maximum value of SVM output of all the superordinate food groups.

When training SVMs, we used all the images of the categories under the superordinate category as positive samples. For negative samples, we built a negative food image set in advance by gathering images using the Web image search engines with query keywords which are expected to related to noise images such as "street stall", "kitchen", "dinner party" and "restaurant" and excluding food photos by hand. All the images are represented by Fisher Vector of HoG patches and color patches. SVMs are trained in the late fusion manner with uniform weights.

In the second step, we apply FC for the selected tweet photos and remove non-food photos from the food photo candidates.

4.3 Specific Food Classifiers

In this step, we classify a given photo into one of the prepared food classes.

First, we extract HOG patches and Color patches in a dense grid sampling manner in the same way as the previous step. Then, we apply PCA to all the extracted local features, and encode them into Improved Fisher Vectors. The method to extract features is the same as the previous step including the parameter settings. Next, we evaluate linear classifiers in the one-vs-rest way by calculating dot-product FVs. Finally we output the top-N categories in terms of the descending order of evaluation scores of all the linear classifiers.

In the experiments, we regarded the given tweet photo as a photo of the corresponding food if the food names contained in the tweet messages are ranked in the top five categories by evaluation of 100-kind food classifiers. This is because the top-5 classification rate exceeds 83%, while the top-1 rate is still around 60%.

5 Experimental Results

In this section, we describe the detail of the 100-class food dataset, the results of food photo mining from the Twitter stream, and some analysis on Twitter food data.

5.1 100 Food Categories and Their Classifiers

In the experiments, as target foods, we used 100 foods in the UEC-FOOD100 [9][2], because we employ supervised food photo classification which requires training data to select the target foods in the third step. It contains more than 100 images per category, and all the food item in which are marked with bounding boxes. For training and evaluation, we used only the regions inside the given bounding boxes. The total number of food images in the dataset is 12,905. Figure 1 shows all the category names and their sample photos. As shown in the figure, the dataset consists of common foods in Japan. Then, we restricted tweets from which we mine food photo tweets to only the tweets with Japanese messages, as mentioned in Section 4.1.

In [6], they implemented a mobile food recognition system using the same dataset. Although basically we followed their method for individual food classification in the third step, we extended the parameter setting to improved accuracy. To say it concretely, we doubled the size of GMM for FV encoding from 32 to 64, and added two-level spatial pyramid. As a result, the total feature dimension are raised from 3072 to 35840, which boosted the classification performance evaluated by 5-fold cross-validation as shown in Figure 2. Regarding the processing time, it takes only 0.024 seconds per image to recognize on Core i7-3770K 3.50GHz with multi-threaded implementation optimized for a quad-core CPU.

5.2 Twitter Food Mining

In this subsection, we describe the experimental results on twitter food photo mining. We have been collecting photo tweet logs by monitoring the Twitter stream by using

[2] `http://foodcam.mobi/dataset/`

Fig. 1. 100 kinds of foods prepared in the UEC-FOOD100 dataset. See this figure with magnification in the PDF viewer.

Twitter Streaming API. Here, we used 122,328,337 photo tweets with Japanese messages out of 988,884,946 photo tweets over all the world collected from May 2011 to August 2013 for two years and four months.

From these photo tweets, we selected 1,730,441 photo tweets the messages of which include any of the name words of the 100 target foods in the first step of the proposed processing flow. Then, in the second step, we applied a "foodness" classifier (FC) to all the selected images. After applying FC, we applied 100-class one-vs-rest individual food classifiers. As a result, we obtained 470,335 photos which are judged as food photos corresponding to any of the 100 target food categories by our proposed processing pipeline.

For the 470,335 selected photos as food photos, we evaluate the number of selected photos for each category. Table 2 shows the ranking of 100 food categories in terms of the number of mined tweet food photos. The number of "Ramen noodle" and "curry" photos are the most and the second most with the large margin to the third or less ranked food categories, respectively. In fact, "ramen" and "curry" are regarded as the most popular foods in Japan. "Sushi", "dipping noodle (called as Tsukemen in Japanese)"

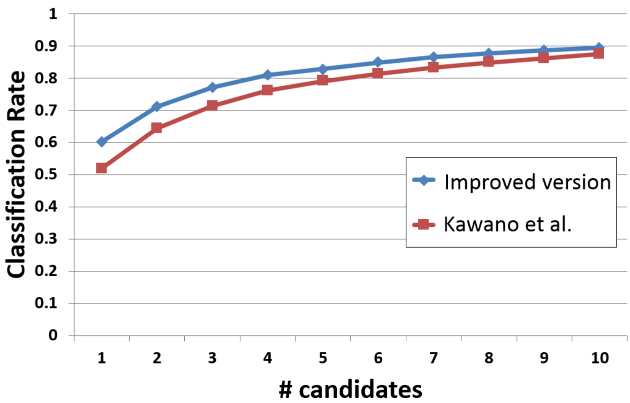

Fig. 2. Comparison on the top-k classification rates with the UEC-FOOD100 dataset evaluated by 5-fold cross validation between [6] and this paper.

Fig. 3. Examples of "omelet" photos. Most of them have drawings drawn by ketchup.

and "omelet with fried rice (called as Ome-rice in Japanese)" are also popular foods in Japan. The results of twitter food image mining reflects food preference of Japanese people. In addition, we found that many of "ome-rice" had drawings or letters drawn with ketchup, as shown in Figure 3. These are estimated to be made at home, while most of "ramen" and "sushi" photos are taken at food restaurants, because there are many ramen noodle and sushi restaurant in Japan. Although "hamburger" and "beef bowl" are also popular in Japan as fast food served at fast-food restaurants such as McDonald and Yoshino-ya, they are ranked at more than twentieth. This is because the foods provided by nation-wide fast-food chain restaurants such as McDonaldo are the same everywhere in the same chain restaurants, and they are not worth posting their photos to Twitter. On the other hand, since there are no monopolistic restaurant chains on ramen noodle and curry in Japan, the foods served at every ramen or curry restaurants have originality and are different from each other.

Next, we evaluated the precision rate of the selected food photos in the each steps regarding the top five foods and two sub-categories of "ramen noodle" and "curry". Table 3 shows the results in case of four types of the combinations of the three kinds of the selection methods, (1) only keywords, (1)+(2) keywords and foodness classifier (FC), (1)+(3) keywords and individual food classifier(IFC), and (1)+(2)+(3) keywords, FC and IFC. Note that this evaluation was done for the 300 random-sampled photos for each cell in the table.

Table 2. The ranking of 100 foods in terms of the number of mined tweet food photos

1	ramen noodle	80021	34	fish-shaped pancake with bean jam	3281	67	dried fish	563
2	curry	59264	35	pork cutlet on rice	3188	68	steamed meat dumpling	561
3	sushi	25898	36	omelet with grilled minced meat	2592	69	french fries	561
4	dipping noodle	22158	37	bibimbap	2368	70	beef ramen noodle	555
5	omelet with fried rice	17520	38	spaghetti	2171	71	sandwiches	551
6	pizza	16921	39	lightly roasted fish	2162	72	cold tofu	517
7	jiaozi	16014	40	seasoned beef with potatoes	2129	73	boiled chicken and vegetables	352
8	Japanese-style pancake	15234	41	natto	2094	74	sirloin cutlet	331
9	steamed rice	14264	42	spaghetti with meat source	1994	75	nanbanzuke	323
10	sashimi	13927	43	steamed egg hotchpotch	1843	76	fried chicken	314
11	hambarg steak	11583	44	egg sunny-side up	1635	77	stir-fried beef and peppers	312
12	beef stake	9503	45	croissant	1579	78	roll bread	288
13	takoyaki	9004	46	udon noodle	1500	79	roast chicken	263
14	fried rice	8383	47	simmered pork	1443	80	macaroni salad	239
15	fried noodle	7905	48	mixed sushi	1371	81	boiled fish	228
16	oden	7453	49	pork miso soup	1229	82	kinpira-style sauteed burdock	225
17	toast	6350	50	ginger-fried pork	1158	83	tempura udon	213
18	cutlet curry	6339	51	potato salad	1150	84	raisins bread	205
19	tempura	5905	52	egg omelet	1146	85	goya chanpuru	198
20	rice ball	5462	53	eels on rice	1071	86	green salad	145
21	gratin	5223	54	egg roll	1058	87	chinese soup	141
22	croquette	4837	55	sweet and sour pork	1049	88	Japanese tofu and vegetable chowder	137
23	stew	4797	56	fried shrimp	1049	89	salmon meuniere	96
24	sashimi bowl	4730	57	sauteed vegetables	1040	90	grilled pacific saury	84
25	chicken-'n'-egg on rice	4513	58	shrimp with chill source	1003	91	chip butty	76
26	tempura bowl	4464	59	cabbage roll	965	92	fried fish	72
27	beef bowl	4285	60	mixed rice	901	93	begitable tempura	71
28	spicy chili-flavored tofu	4081	61	pilaf	891	94	tensin noodle	69
29	yakitori	3829	62	soba noodle	880	95	ganmodoki	34
30	hamburger	3662	63	potage	816	96	grilled salmon	25
31	chilled noodle	3473	64	hot dog	795	97	sauteed spinach	12
32	sukiyaki	3408	65	chicken rice	736	98	teriyaki grilled fish	3
33	miso soup	3295	66	wiener sausage	577	99	grilled eggplant	2
						100	pizza toast	0

Regarding (1), the precision of two sub-categories, "beef ramen noodle" and "cutlet curry", are relatively higher, 94.3% and 92.7%, than "ramen noodle" and "curry". From this results, we can assume that when tweeting detailed food names with photos, the photos probably represent the corresponding foods. After applying both FC and IFC, (1)+(2)+(3), the precision of all the seven foods achieved the best compared to the cases of applying only single methods or only keyword-based search, (1), (1)+(2) and (1)+(3). Except for "sushi", the precision reached 99.0%, which means nearly perfect. This shows the effectiveness of introducing both FC and IFC after keyword-based search. Exceptionally, "sushi" is a difficult food to recognize by object recognition methods, because the appearances of "sushi" varies greatly depending on the kinds of the ingredients on the pieces of hand-rolled rice.

Finally, we describe simple spatio-temporal analysis on Twitter food photos. Figure 4 shows the prevailing-food map where the red marks, the yellow marks and the blue marks represent the areas where "ramen noodle", "curry" and "okonomiyaki" are most popular in terms of the number of food photo tweets, respectively. The left map, the center map, and the right map show the prevailing-food map on all the term (May 2011-Aug. 2013), Dec. 2012 (in winter), and Aug. 2013 (in summer), respectively. From the leftmost map, "ramen noodle" is the most popular over Japan on average through a year. However, compared between the center map and the rightmost map, popularity of "curry" increases in summer, while "ramen noodle" becomes the most popular

Table 3. The number of selected photos and their precision(%) with four different combinations

food category	(1)	(1)+(2)	(1)+(3)	(1)+(2)+(3)
ramen noodle	275652 (72.0%)	200173 (92.7%)	84189 (95.0%)	80021 (99.7%)
beef ramen noodle	861 (94.3%)	811 (99.0%)	558 (99.7%)	555 (99.7%)
curry	224685 (75.0%)	163047 (95.0%)	62824 (97.0%)	59264 (99.3%)
cutlet curry	10443 (92.7%)	9073 (98.0%)	6544 (98.7%)	6339 (99.3%)
sushi	86509 (69.0%)	43536 (86.0%)	48019 (72.3%)	25898 (92.7%)
dipping noodle	33165 (88.7%)	24896 (96.3%)	28846 (93.7%)	22158 (99.0%)
omelet with fried rice	34125 (90.0%)	28887 (96.3%)	18370 (98.0%)	17520 (99.0%)

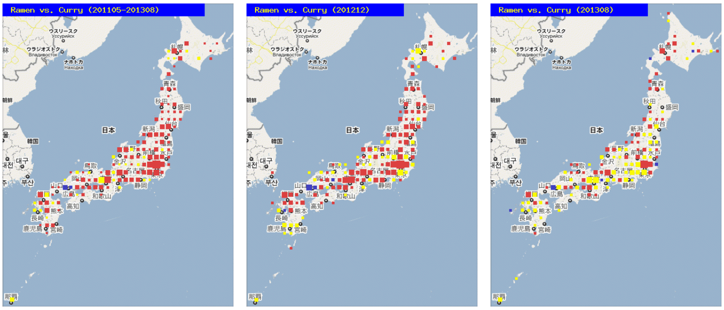

Fig. 4. The prevailing-food map of Japan. See the text.

in winter. Exceptionally, in the area around Hiroshima where the blue marks appear, “okonomiyaki” is always the prevailing food in Twitter food photos, this is partly because Hiroshima has a very popular regional food, “Hiroshima-yaki”, which is a variant of “okonomiyaki”.

As another temporal analysis on the mined food photos, we examined the time when each food are eaten the most frequently over a day. As results, the most frequent time when “ramen noodle” and “curry” are eaten is between 12pm and 2pm, while the most frequent time of “sushi” and “okonomiyaki” is between 7pm and 9pm. This reflects the difference of the characteristic of the foods. As shown in this subsection, the data we collected through Twitter food photo mining is useful for food habit analysis.

5.3 Real-Time Food Photo Detection System

We implemented a real-time Twitter food photo detection system which continuously detects 100 kinds of food photos from the Twitter stream. We detect the photo tweets including any of 100 kinds of Japanese food names about ten times per minute at most. Because the time to download a thumbnail image is about 2 or 3 seconds and the processing time for food recognition for each image is less than 0.1 seconds, we can process all the pipeline on a single machine in the real-time way. The very fast food recognition method which was originally designed for a mobile application made it possible.

As shown in Figure 5, the detected food photos are shown on the map if they have geotags or geo-related words such as place names in their Tweet messages, and on the right side the photos are displayed as the results by online k-means clustering. This system can be accessible via `http://mm.cs.uec.ac.jp/tw/`.

Fig. 5. Detected food photos are displayed on the map when geo-information are available

6 Conclusions

In this paper, we described food photo mining from the Twitter stream as a case study of specific categories of tweet photo mining. To do that, we proposed the three-step processing consisting of keyword-based selection with food category names, classification of food or non-food photos, and visual verification of the correspondence between the extracted food words and the food category of the tweet photo. In addition, we showed the collected data was useful for various kinds of analysis of foods.

Currently we always keep running the real-time food photo detection system and collecting new food photos. For example, we are collecting about 20,000 "ramen noodle" and 15,000 "curry" photos per month. When the number of both food images exceeds one million, we will release them as a large-scale food dataset for research purpose, which we expect enables fine-grained food recognition research.

References

1. Bergamo, A., Torresani, L.: Meta-class features for large-scale object categorization on a budget. In: Proc. of IEEE Computer Vision and Pattern Recognition (2012)
2. Chen, T., Lu, D., Kan, M.Y., Cui, P.: Understanding and classifying image tweets. In: Proc. of ACM International Conference Multimedia, pp. 781–784 (2013)

3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. of IEEE Computer Vision and Pattern Recognition (2005)
4. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: LIBLINEAR: A library for large linear classification. The Journal of Machine Learning Research 9, 1871–1874 (2008)
5. Kaneko, T., Yanai, K.: Visual event mining from geo-tweet photos. In: Proc. of IEEE ICME Workshop on Social Multimedia Research (SMMR) (2013)
6. Kawano, Y., Yanai, K.: Rapid mobile object recognition using fisher vector. In: Proc. of Asian Conference on Pattern Recognition (2013)
7. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proc. of IEEE Computer Vision and Pattern Recognition (2006)
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
9. Matsuda, Y., Yanai, K.: Multiple-food recognition considering co-occurrence employing manifold ranking. In: Proc. of IAPR International Conference on Pattern Recognition (2012)
10. Nakaji, Y., Yanai, K.: Visualization of real world events with geotagged tweet photos. In: Proc. of IEEE ICME Workshop on Social Media Computing (SMC) (2012)
11. Perronnin, F., Sánchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 143–156. Springer, Heidelberg (2010)
12. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proc. of the International World Wide Web Conference, pp. 851–860 (2010)
13. Yanai, K.: World seer: a realtime geo-tweet photo mapping system. In: Proc. of ACM International Conference on Multimedia Retrieval (2012)