Int. J. Nonlinear Anal. Appl. 8 (2017) No. 1, 291-299
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2017.10744.1522

Similarity measurement for describe user images in social media

Alireza Tavakoli Targhi

CS group of Mathematics department, Shahid Beheshti University, Tehran, Iran

(Communicated by M. Eshaghi)

Abstract

Online social networks like Instagram are places for communication. Also, these media produce rich metadata which are useful for further analysis in many fields including health and cognitive science. Many researchers are using these metadata like hashtags, images, etc. to detect patterns of user activities. However, there are several serious ambiguities like how much reliable are these information. In this paper, we attempt to answer two main questions. Firstly, are image hashtags directly related to image concepts? Can image concepts being predicted using machine learning models? The results of our analysis based on 105000 images on Instagram show that user hashtags are hardly related to image concepts (only 10%of test cases). Second contribution of this paper is showing the suggested pre-trained model predicate image concepts much better (more than 50% of test cases) than user hashtags. Therefore, it is strongly recommended to social media researchers not to rely only on the user hashtags as a label of images or as a signal of information for their study. Alternatively, they can use machine learning methods line deep convolutional neural network model to describe images to extract more related contents. As a proof of concept, some results on food images are studied. We use few similarity measurements to compare result of human and deep convolutional neural network. These analysis is important because food is an important society health field.

Keywords: Similarity Measurement Web mining; Health Topics; Computer Vision; Machine Learning Models.
2010 MSC: Primary 68T45; Secondary 68P05.

Email address: a_tavakoli@sbu.ac.ir (Alireza Tavakoli Targhi)

Received: February 2016 *Revised:* December 2016

1. Introduction

Social networks provide a creative way for people to produce, share and exchange ideas over the internet. For example, in every minute of the day, people share over 2.4 million pieces of content in the Facebook, Twitter users tweet about 277 thousand times, and users post near 216 thousand posts in Instagram. In addition to helping people to strengthen their relationship, these media produce rich metadata for web mining purposes and digital health analysis. Moreover, with the huge growth of metadata, it is crucial to use learned models to accelerate analyzing, and decision making in researches.

Some examples of applied model usages in social net-work analysis are community detection, information diffusion pre-diction, sentiment analysis and predicting Tweet hashtags in Twitter [17, 13, 10, 18]. Likewise, thanks to the help of these machine learning models, image labels can be predicted in Instagram by the study of image contents and the number of followers and followings [8].

In the other research on Instagram images, it is proven that images which contain faces are more liked by other users [2].

Moreover, analyzing the number of metadata produced by some photo sharing social media like Instagram, may invoke some controversial questions which are worth asking. Some examples of rich metadata are user hashtags which could answer many questions in mental health field.

In this paper some of these questions would be answered. For example, a question is whether or not Istagram image hashtags can be predicted. In other words, are there any direct correlation between an image content and user hashtags or user hashtags only clarify users feelings in that moment? What are the most hot food topics in social media? How images and hashtags are related?

In a recent research work [4], 30 images are considered as a case study to ask expert people to tag them. It is proven that an average of 55% of the participants chosen hashtags are similar with those suggested by the photo owners. Also it is shown only 30% of the hashtag for 30 test images are appropriate and are related to the content of the image. However, it brings into question that how the results would be for a large number of images. The second question is that, can a machine learning model predict hashtags better than human? In the following section, this question will be answered.

Another question is what the most frequent health topics are in social media? Also in [16, 11] health topics are explored using microblogging websites like Twitter and Sina Weibo. Moreover, in Instgaram, images are rich data for health topics classification. For example, in [5] Instagram images are classified into 8 common classes in which food is one of them with 10% of overall distribution.

In this paper, most frequent food topics extracted from image labels are categorized by user hashtags and machine generated hashtags to compare their performance in this task. This research work shows hashtags of images generated by users are not related to the image content in most cases. Additionally, it is shown a machine learning approach based on deep-learning has superior performance to annotated images which are more reliable than user annotations

2. Image labels and hashtag relationship

There are some steps to analyze and compare correlation of images contents and user hashtags. First, images are split into ones containing face and ones not. Then, the most probable labels for images are extracted using a pre-trained model of deep Convolutional Neural Network (CNN). After image labeling, users hashtags are preprocessed using natural language toolkit (NLTK) for obtaining the WordNet ID (wnid). Then, for the comparison part, similarity for each image is formulated using a bipartite graph. At last, it is shown that the model labels images better than users. Each part is describes in the next sections.

2.1. Dataset

By the help of Instagram API, over 105 thousand U.S.A Instagram images and their metadata including image likes, comments, hashtags, etc. are gathered. By analyzing extracted labels using the considered pre-trained model in this paper, it is recognized that image hashtags which containing faces are extremely irrelevant to images.

It is logical because, in most cases, people do not use hashtags like "face" or "human" for their faces. Otherwise, users use hashtags like "yay", "workflow", "selfie", etc. Moreover, some images have abstract concepts, and their labels could not be predicted by the model. Figure **??** show three types of defined images.

To extract image labels better, the images are separated into images. containing faces and which not. To separate images, the Viola-Jones cascade object detector in MATLAB is used. This detector uses different features like HOG, Haar-like features and a cascade of classifiers trained using boosting [3, 15]. About 60% of separated images were not containing faces and 40% of them have at least one face in them. This result is as same as results in the previous paper [5].

Figure 1: First defined category is face set in which pre-trained model labels are hardly related to users hashtags.

Figure 2: Second defined category is abstract image set in which pre-trained model labels are hardly related to users hashtags.

Figure 3: Third defined category is other image set like scenery, objects, etc. in which pre-trained model labels are.

2.2. Image labeling

After getting image data, for saving computation time, one of the most accurate pre-trained model, GoogleNet which is available inCaffes Model Zoo, is used for image classification [1, 14]. The model is the Caffe replication of Googles GoogLeNet.

GoogLeNet, the winner of the Large Scale Visual Recognition Challenge 2014 (ILSVRC2014) classification task, obtains a top-1 accuracy 68.7% (31.3% error) and a top-5 accuracy 88.9% (11.1%

Figure 4: Predicted labels for the image by pre-trained model.

error) choosing within 1000 classes on the validation set on the ILSVRC2012 dataset. The goal of ILSVRC is to estimate the content of photographs for the purpose of retrieval and automatic annotation using a subset of the large hand-labeled ImageNet dataset as training. ImageNet is an image database organized according to the WordNet hierarchy [7].

WordNet is a large lexical database of English in which Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept [9]. As shown in Figure 1, by the use of GoogleNet, top-5 most probable labels are extracted for each dataset image. Each synset label is identified with WordNet Id (wnid). In the next session, how WordNet ID of user hashtags are produced will be discussed.

2.3. Hashtag preprocessing

After image labeling, only images which containing at least a user hashtag are selected. 40% of images do not have hashtags. Thanks to the help of Natural Language ToolKit (nltk), for each user hashtag, a set of synsets are extracted using WordNet noun tree with is-a relationship. Each synset is uniquely identified using WordNet ID (wnid). However, some hashtags are not nouns and their synsets are not recognizable. Thus they were omitted too.

2.4. Similarity comparison

After above filter about 22 thousand images remained for later analysis. To find similarity score between two images concept synsets and hashtags synsets, firstly, the similarity measurement between two WordNet ID is described. Then, the overall similarity between image labels and users hashtags is formulated using a bipartite graph.

Similarity Measurement: After extracting wnid for image labels and hashtags, these two sets of WordNet IDs are compared using a similarity measurement. The first one is Jiang-Conrath similarity (JCS) [6]. According to papers, this similarity is closer correlating with human judgment [12]. The formulation is like below:

$$classes(w) = \{c | w \in words(c)\}$$
$$freq(c) = \Sigma_{w \in words(c)} \frac{freq(w)}{|classes(w)|}$$

$$P(c) = \frac{freq(c)}{N}$$
$$IC(c) = \log^{-1} P(c) \tag{2.1}$$
$$Sim(s_1, s_2) = (1/IC(s_1)) + IC(s_2) - 2 * IC(lcs),$$

where IC (s_1) is Information Content of synset s_1, and lcs is the Least Common Subsumer of two synsets. The Information Content is measure of specificity for concept. The higher values of Information Content, the more specific the concepts are [9]. In this formula, if the denominator becomes zero, the result would be infinite. To handle this issue, if the denominator result is less that 0.5 the similarity score is considered (2.1).

2.5. Similarity formulation

To formulate over all similarity, a bipartite graph G=(U,V,E) where U is the set of image concept synsets, V is the set of user hashtag synsets and E are weighted edges resembling similarity scores described in above section. The proposed overall similarity formula is maximum of geometric mean described below:

$$AllSim(G) = MaxCm * ({W_{m1}}^n + {W_{12}}^n + \ldots + {W_{mn}}^n)^{(1/n)} \tag{2.2}$$

where n is the number of user hashtag synsets, and m is the number of image concept synsets. $C_i$is the probability of i th synset which model predicted, and $W_{ij}$is the above calculated similarity for two image and hashtag synsets. Another similarity is based on maximum of two synset similarities:

$$AllSim(G) = MaxW_{ij} \tag{2.3}$$

where i is between 1 to n and j is 1 to m as described above, and $W_{ij}$is similarity$Sim(s_i, s_j)$ which s_i is i th synset of user and $s_j$is j th synset of user.

In this paper, 7 thousand images have the top-1 predicted labels above 0.5. Only 4% of these images have similarity above 0.5 with similarity calculated with formulae (2.2) and 10% with formulae (2.3). This shows that images concept and user hashtags are hardly related.

3. Image labels and hashtag reliability

By choosing 100 random images from images which the above overall are less than 0.1, it is discovered that the pre-trained model better labels images. Only in 10 % of selected images user hashtags are better descriptors of images' concept. Figure 5.a is the example of image that users hashtag, "surfing", is better than the model label, "doormat". On the other hand, in more than 40 % of images model correctly label images. As Figure 5.b shows, the top model label is "laptop" and the user hashtag is "dollop" for the image, and the laptop is clearly is visible in image. In the rest of images, in which most of images are abstract or written captions, none of the labels exactly describe images content. The user hashtag is "happy Tuesday", and the image label is "envelope". Figure 5.c shows this situation.

4. Hot health topics in social media

Counting images shows that about 45% of 22 thousand image posts are labeled as food either by model or users. Table 1 shows the 20 most frequent model predicted labels in which hashtag synsets have

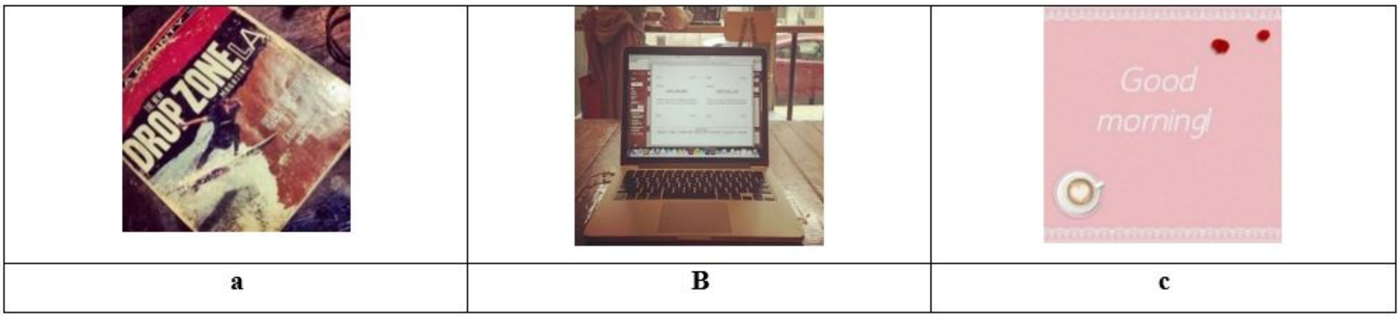

Figure 5: a. The example of image that users hashtag, "surfing", is better than model label, "doormat". Figure 5.b. The top model label is "laptop" and the user hashtag is "dollop" for the image, and the laptop is clearly visible in the image. Figure 5.c. Abstract or written captions, none of the labels exactly describe images' content. The user hashtag is "happy Tuesday", and the image label is "envelope."

Table 1: The 20 most frequent model predicted labels in which hashtag synsets have a "food" synset ancestor in the WordNet hierarchy. The percentage and frequency are calculated in 22 thousand images.

Food Labels	mole	meatloaf	ice_cream	menu	eggnog	choclate_sauce	pizza	hotdog	bagel	red_wine
Frequency	1124	803	749	545	409	405	390	376	321	266
Percentage	10.77%	7.69%	7.17%	5.22%	3.92%	3.88%	3.74%	3.60%	3.07%	2.55%
Food Labels	mashed_potato	trifle	dungeness_crab	plate	crayfish	french_loaf	confectionery	puff	pretzel	board
Frequency	260	259	243	235	208	207	190	178	173	173
Percentage	2.49%	2.48%	2.33%	2.25%	1.99%	1.98%	1.82%	1.70%	1.66%	1.66%

Table 2: The 20 most frequent user hashtags in which hashtag synsets have a "food" synset ancestor in the WordNet hierarchy. The percentage and frequency are calculated in 22 thousand images.

Food Labels	food	coffee	sweet	lunch	pizza	beer	water	dessert	cocktail	chicken
Frequency	570	391	216	211	179	174	127	115	114	103
Percentage	8.50%	5.83%	3.22%	3.15%	2.67%	2.60%	1.89%	1.72%	1.70%	1.54%
Food Labels	coloring	whiskey	doughnut	vintage	taco	caffe_latte	french_fries	cake	dinner	ice
Frequency	96	84	79	75	66	64	63	59	57	57
Percentage	1.43%	1.25%	1.18%	1.12%	0.98%	0.95%	0.94%	0.88%	0.85%	0.85%

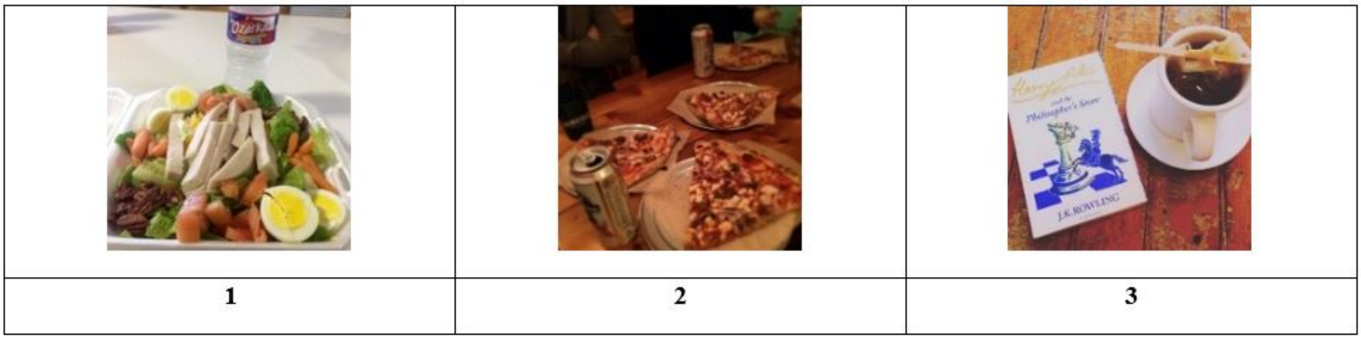

Figure 6: Food images (images with at least one food related image label) which correctly contain at least one food

a "food" synset ancestor in the WordNet hierarchy. The percentages and frequencies are calculated in 22 thousand images.

Table 2 is user hashtags percentages and frequencies calculated as above. Comparing these two sets of labels shows there is a low correlation between these two sets. This comparison leads to another question whether or not food related hashtags are used properly on food images. To answer

Table 3: The 20 most frequent model predicted labels in which hashtag synsets have a "food" synset ancestor in the WordNet hierarchy. The percentage and frequency are calculated in 22 thousand images.

Image Predicted Labels	User Hashtags
['cucumber', 'cuke', 'zucchini', 'courgette', 'plate', 'Granny_Smith', 'corn']	['eatinghealthy', 'ozarka', 'salad']
['pizza', 'pizza_pie', 'butcher_shop', 'meat_market', 'waffle_iron', 'French_loaf', 'lipstick', 'lip_rouge']	['boilerroom', 'pbr', 'logansq', 'pizza']
['espresso', 'chocolate_sauce', 'chocolate_syrup', 'cup', 'coffee_mug', 'pretzel']	['tea', 'coffee', 'windycitybloggers', 'harrypotter']

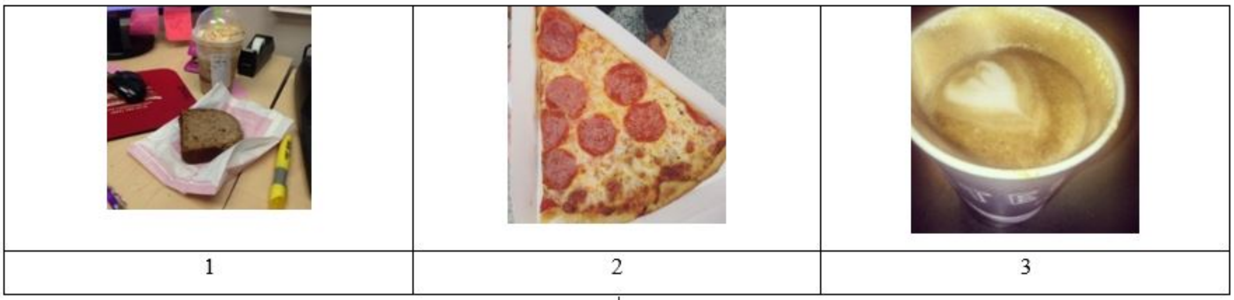

Figure 7: Food images without any food hashtags.

Table 4: The 20 most frequent model predicted labels in which hashtag synsets have a "food" synset ancestor in the WordNet hierarchy. The percentage and frequency are calculated in 22 thousand images.

Image Predicted Labels	User Hashtags
['ice_cream', 'icecream', 'plate', 'packet', 'spatula', 'burrito']	['sorryjellybean', 'mommyneedscoffee', 'work', 'inventory']
['pizza', 'pizza_pie', 'tray', 'trifle', 'bakery', 'bakeshop', 'bakehouse', 'meat_loaf', 'meatloaf']	['illinois', 'chicago']
['espresso', 'cup', 'coffee_mug', 'eggnog', 'soup_bowl']	['intelligentsia']

this question, food related images are divided into three categories. First category is consisted of food images (images with at least one food related image label) which correctly contains at least one food related user hashtag.

This category is about 29% of food related posts in 22 thousand images. Figure 6 shows some samples. The second category is about food images without any food hashtags. This category is about 32% of food related posts. Figure 7 shows some samples of it.

The last category which is really important is consist of not food images (images without any food related image label) which own at least one food related hashtag. This category is about 39% of food related posts. Figure 8 shows samples of it. Considering the correlation between user hashtags and image labels on image posts labeled as food either by model or users, shows user hashtags will lead us to not food images in 57% of cases. So gathering images using user hashtags are not suitable for learning data in deep learning algorithms. As told in previous sections, image labels extracted by using a pre-trained deep CNN are better defining images Thus, for finding relevant content, these labels are more reliable.

5. Conclusion

Social networks like Instagram include metadata which can be used by researchers in many fields like cognitive science, web mining and health. It is obligate to ensure that these metadata are reliable.

One of these studies is on Instagram images and their hashtags. In this paper, it is shown that 40% of Instagram images contain at least a human face which hashtags are not strongly related to image con-tent, so they cannot contribute to producing qualified knowledge.

According to [2], images with human face received more likes. However, our results are shown this is not due to the human face popularity, but because of majority of uploaded images are containing a

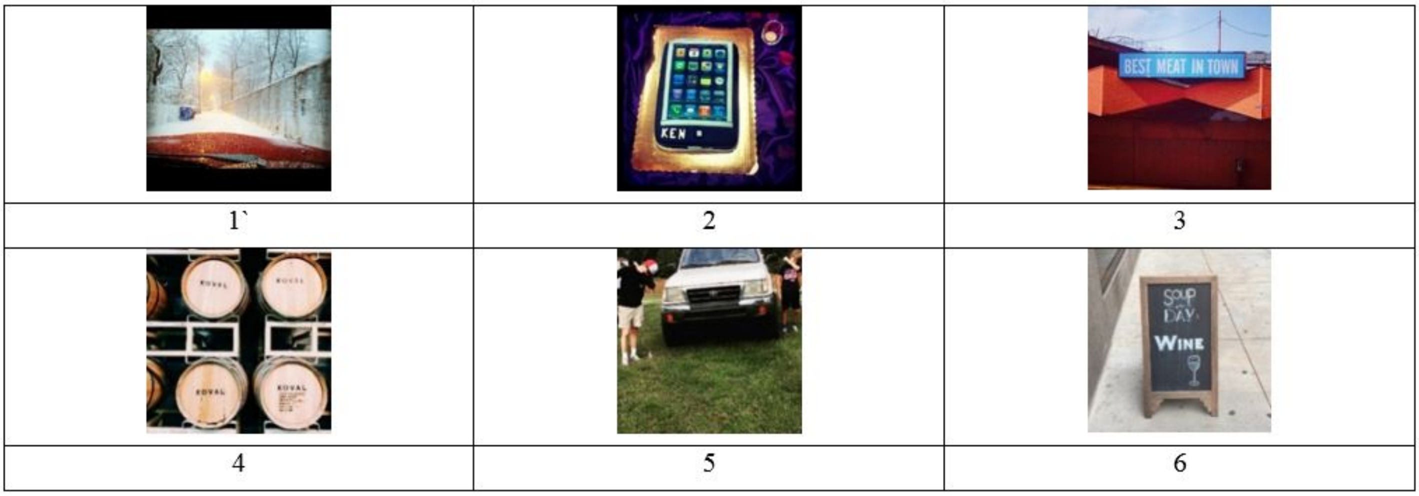

Figure 8: Food images (images without any food related image labels) which own at least one food related hashtag.

Table 5: The 20 most frequent model predicted labels in which hashtag synsets have a "food" synset ancestor in the WordNet hierarchy. The percentage and frequency are calculated in 22 thousand images.

Image Predicted Labels	User Hashtags
['snowplow', 'snowplough', 'chainlink_fence', 'barn', 'fountain', 'traffic_light', 'traffic_signal', 'stoplight']	['sandwich', 'love', 'foodporn', 'food', 'loveit', 'eggs', 'instapic', 'bacon', 'foods', 'good', 'insta', 'instagreat', 'best', 'instafood', 'chicago', 'instagood', 'bagel', 'instapics', 'yum', 'instabacon', 'awesome']
['cellular_telephone', 'cellular_phone', 'cellphone', 'cell', 'mobile_phone', 'iPod', 'remote_control', 'remote', 'slot', 'onearmed_bandit', 'payphone', 'paystation']	['cake', 'weddings', 'iphone']
['street_sign', 'cinema', 'movie_theater', 'movie_theatre', 'movie_house', 'picture_palace', 'barbershop', 'bakery', 'bakeshop', 'bakehouse', 'traffic_light', 'traffic_signal', 'stoplight']	['enoughsaid', 'meat', 'texas', 'dallas']
['barrel', 'cask', 'drum', 'membranophone', 'tympan', 'face_powder', 'Petri_dish', 'rain_barrel']	['whisky', 'vscocam', 'koval', 'dyk']
['minivan', 'beach_wagon', 'station_wagon', 'wagon', 'estate_car', 'beach_waggon', 'station_waggon', 'waggon', 'croquet_ball', 'car_wheel', 'police_van', 'police_wagon', 'paddy_wagon', 'patrol_wagon', 'wagon', 'black_Maria']	['sauce']
['street_sign', 'mailbox', 'letter_box', 'ashcan', 'trash_can', 'garbage_can', 'wastebin', 'ash_bin', 'ashbin', 'ashbin', 'dustbin', 'trash_barrel', 'trash_bin', 'brass', 'memorial_tablet', 'plaque', 'doormat', 'welcome_mat']	['soup', 'friday', 'weekend', 'wine']

human face. Additionally, more than 40% of uploaded images do not have any hashtags, this means these images cannot be used on most of recent analysis. Moreover, most of user hash tags (about 57%) are not related to the image contents. Consequently, they are not informative and can mislead any research based on user hashtags. The results are shown by statistics and few visual samples to address these short comes.

In this research, a pre-trained model is used to automatically generate hashtags for uploaded images. A food sample is described as an example. User hashtags and model image labels are used to categorize different food types in the shared images and their metadata. As expressed above, machine generated tags are more related to the contents of images

References

[1] https://github.com/bvlc/caffe/wiki/model-zoo.

[2] S. Bakhshi, D.A. Shamma and E. Gilbert, *Faces engage us: Photos with faces attract more likes and comments on instagram*, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM 2014.

[3] N. Dalal and B. Triggs, *Histograms of Oriented Gradients for Human Detection*, IEEE Comput. Soc. Conf. Comput. Vision Pattern Recogn. 1 (2005) 886–893.

[4] S. Giannoulakis and N. Tsapatsoulis, *Instagram Hashtags as Image Annotation Metadata. Artificial Intelligence Applications and Innovations*, Springer International Publishing, 206–220, 2015.

[5] Y. Hu, L. Manikonda and S. Kambhampati, *What we instagram: A first analysis of instagram photo content and user types*, Proceedings of ICWSM. AAAI, 2014.

[6] J.J. Jiang and D.W. Conrath, *Semantic similarity based on corpus statistics and lexical taxonomy*, arXiv preprint cmp-lg/9709008, 1997.

[7] A. Krizhevsky, I. Sutskever and G.E. Hinton, *Imagenet classification with deep convolutional neural networks*, Adv. Neural Inform. process. syst., 2012.

[8] J. McAuley and J. Leskovec, *Image labeling on a network: using social-network metadata for image taxonomy*, arXiv preprint cmp-lg/9709008, 1997.

[9] G.A. Miller, R. Beckwith, C. Fellbaum, D. Gross and K.J. Miller, *Introduction to WordNet: An on-line lexical database*, Int. j. lexicography 3 (1990) 235–244.

[10] A. Pak and P. Paroubek, *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*, LREC. Vol.10, 2010.

[11] M.J. Paul and M. Dredze, *Discovering health topics in social media using topic models*, (2014): e103408.

[12] T. Pedersen, *Information content measures of semantic similarity perform better without sense-tagged text*, Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 2010.

[13] M.G. Rodriguez, J. Leskovec, D. Balduzzi and B. Schlkopf, *Uncovering the structure and temporal dynamics of information propagation*, Network Science 2 (2014) 26–65.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, *Going deeper with convolutions*, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9, 2015.

[15] P. Viola and M.J. Jones, *Rapid Object Detection using a Boosted Cascade of Simple Features*, Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vision Pattern Recogn. 1 (2001) 511–518.

[16] S. Wang, M.J. Paul and M. Dredze, *Exploring health topics in Chinese social media: An analysis of Sina Weibo*, AAAI Work World Wide Web Public Heal Intell, 2014.

[17] J. Yang, J. McAuley and J. Leskovec, *Community detection in networks with node attributes*, Data Mining (ICDM), 2013 IEEE 13th international conference on. IEEE, 2013.

[18] L. Yang, T. Sun, M. Zhang and Q. Mei, *We know what @ you# tag: does the dual role affect hashtag adoption?*, In Proceedings of the 21st international conference on World Wide Web, pp. 261–270. ACM, 2012.