Visual Analysis on Relations between Nouns and Adjectives Using a Large Number of Web Images

Yuuya Kohara Keiji Yanai
Depeartment of Informatics, The University of Electro-Communications,Tokyo
1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182-8585 Japan
{kohara-y,yanai}@mm.cs.uec.ac.jp
Keywords: generic object recognition, entropy, tag, large number of images, attribute

Abstract

In recent years, due to the wide spread of photo sharing Web sites such as Flickr and Picasa, we can put our own photos on the Web and show them to the public easily. To make the photos searched for easily, it is common to add several keywords which are called as "tags" when we upload photos. However, most of the tags are added one by one independently without much consideration of association between the tags. Then, in this paper, as a preparation for realizing simultaneous recognition of nouns and adjectives, we examine visual relationship between tags, particularly noun tags and adjective tags, by analyzing image features of a large number of tagged photos with mutual information. As a result, it was turned out that mutual information between some nouns such as "car" and "sea" and and adjectives related to color such as "red" and "blue" was relatively high, which showed that their relations were stronger.

1 Introduction

In recent years, due to the wide spread of digital cameras and mobile phones with cameras , the amount of images on the Web has increased explosively. At the same time, because photo sharing sites such as Flickr have become common where users post their images with tags, there are so many tagged images on the Web. These tag information are used as a keyword on image search. However, most of the tags are added one by one independently without much consideration of association between the tags. This sometimes causes irrelevant results when we do AND-search with multiple keywords. For example, we obtain a photo showing blue sky and a red car for the query with "blue AND car". To remove such a photo and to obtain only the photos including blue cars, simultaneous image recognition of multiple tags such as "blue cars" is needed. If we can automatically eliminate irrelevant images, search results for multipe keywords become more correct. In addition, dataset can be created easily with less noise. In order to perform more accurate image acquisition and image search, it is necessary to take into account the relationship between the tags and to focus more on the contents of the image.

Then in this paper, we analyze the visual relationship between nouns and adjectives using a large number of tagged images on the Web. To do that, we use entropy and mutual information based on visual features extracted from images.

Moreover, the result of analysis in this time can be used in simultaneous recognition which recognizes a certain object by a noun and recognizes the state of the object by an adjective further like "there is a car and the color of the car is red".

In the rest of this paper, we describe relate work in Section 2. We explain the overview in Section 3 and the detail of the proposed system in Section 4. We show experimental results in Section 5, and discuss about the results in Section 7. In Section 8, we conclude this paper.

2 Related work

In the field of image recognition, it is increasing the recognition of the attribute such as an adjective. Here, we introduce such research that focused in the attribute.

T.L.Berg et al.[1] focused on the attribute of color, shape, and texture. T.L.Berg et al. extracted a word associated with the attributes from the text which were listed in the shopping site, and labeled to local region represented the attribute most in the image corresponded text description. In addition, the local regions were expressed in blocks of 75 * 75 pixels.

D.Parikh et al.[2] focused on the attributes from the perspective of "nameable". The "nameable" showed whether human can understand and represent by language. D.Parikh et al. discovered attribute of "nameable" by an interactive approach by using Amazon Mechanical Turk.

Both papers [1],[2] recognized a single attribute. However, we limit an attribute to an adjective, and also focus on the relationship between a noun and an adjective.

Next, we introduce the previous works. Here, we cite the papers of Yanai et al.[3], Akima et al.[4], and Kawakubo et al.[5].

Yanai et al. proposed the entropy as a way to quantify the relation of visual concept, and referred to the visual relation about 150 adjectives. We use the method of quantifying the visibility of the word by entropy and calculation of entropy.

Akima et al. built a database with hierarchical structure of between the concepts from distance relationship and hierarchical relationship. They used the entropy and calculate the distribution of the image when they determine the hierarchical relationship. In addition, tag information which given to the images was also used.

Kawakubo et al. searched the visual and geographical distribution in word concepts. In this research, they calculated the image distribution of the class of concepts such as a noun or an adjective to obtain visibility by using calculation of entropy and region segmentation.

The difference between this research and previous works is that we classify the class of combination of two words, and search the visual relation of the combination of nouns and adjectives.

3 Overview

In this paper, the visual relation between a nouns and an adjective is represented by the width of distribution of the image. And, we determine that there is a high relation if visual distribution is narrow enough. Width of the distribution is quantified using the concept of entropy. Entropy is used to calculate the local features obtained from the image. Mutual information is the difference between entropy of noun and entropy of combination with an adjective and a noun. Mutual information becomes higher when visual relation between a noun and an adjective becomes higher.

Execution procedure in the experiment in this paper is shown below.

Execution procedure

1. Image acquisition by the tag-Search
2. Image segmentation
3. Feature extraction and creating BoF
4. Positive region determination
5. Calculation of Feature distribution in each positive region by PLSA
6. Calculation of entropy
7. Calculation of similarity by tag co-occurrence
8. Analysis of the relationship between the tags

In this paper, we calculated similarity by co-occurrence of tags by the Normalized Google Distance (NGD) for comparison with the visual relation by entropy.

4 Proposed method

In this chapter, we describe the methods used in the experiment. In this experiment, we calculated the entropy to refer to visual relation between an adjective and a nouns. In addition, we calculated the similarity by co-occurrence of tags for comparison.

4.1 Calculation of entropy

The entropy was calculated using the probability obtained by the pLSA. The entropy increases when the distribution of local feature vectors representing the positive regions becomes wider. And the entropy decreases when the distribution becomes narrower. Therefore, the size of the entropy represents the width of the image distribution belonging to the class concept which is the combination of a noun and an adjective. That is, calculating the entropy leads to searching for visual relation each the combinations.

4.1.1 Entropy

We calculated the entropy using the $P(z_k|d_i)$ determined by pLSA. First, we calculated

$$P(z_k|w_j) = \frac{\sum_{i=1}^{I} P(z_k|d_i)}{|I|} \tag{1}$$

for each latent topic variables. Then, we calculated the entropy by

$$H(P) = -\sum_{k=1}^{K} P(z_k|w_j)\log(P(z_k|w_j)) \tag{2}$$

for each images using $P(z_k|w_j)$.

4.1.2 Mutual information

The mutual information is a value represented by the difference of the entropy, and indicates the relation between the tags. We calculated the mutual information as

$$MI(X;Y) = H(X) - H(X|Y) \tag{3}$$

where $H(X)$ is the entropy of one class, and $H(X|Y)$ is the entropy of the class combined two classes. If image distribution becomes narrow by combining the tag X with the tag Y, we judge the visual relation become higher from the increase of mutual information.

4.2 Calculation of similarity by co-occurrence of tags

The calculation of similarity by co-occurrence of tags is calculated using the Normalized Google Distance (NGD)[6]. The formula is

$$NGD = \frac{\max\{\log f(x), \log f(y)\} - \log f(x,y)}{\log N - \min\{\log f(x), \log f(y)\}} \tag{4}$$

where x is a noun, y is an adjective, $f(x)$ and $f(y)$ are the image number of tag search by a noun and an adjective in Flickr, and $f(x,y)$ is the image number of AND-search by combination of a noun and an adjective. Moreover, N is the number of all images in Flickr. However, we assume N is 50 billion since it is very difficult to understand the exact number.

5 Experiments

5.1 Dataset

Images were acquired using the API from Flickr. When retrieved, we acquired 800 negative images and 200 positive images under the restriction which we acquire only one image from the same author. In addition, we retrieved positive images in order from the top in ranking of search of Flickr. Negative images were selected from among the images obtained at random from Flickr, which does not have the tags of nouns and adjectives of a particular class. In this experiment, we selected 20 nouns in Table 1 and 15 adjectives in Table 2. Thus, we calculated the entropy about 20*15 classes which are the combination of each noun and each adjective and 20 classes which are only noun.

Table 1: The 20 nouns used in experiment

beach	bird	boat	bridge	car
cat	cloud	cup	dog	flower
fruit	house	people	sea	sky
snow	sun	tower	train	tree

Table 2: The 15 adjectives used in experiment

red	blue	green	black	white
circle	square	morning	night	winter
summer	new	old	beautiful	cool

5.2 Experimental Method

Using the dataset in the previous section, experiments were conducted as follows.

5.2.1 Region Segmentation

Segmentation was used JSEG[7] as maximum number of regions is 10. However, since this number is the maximum number of division, there are many images which the number of regions becomes smaller than 10. Moreover, we also have a post-processing to integrate a small area. In this case, we adjusted the parameters, so that the region where the relative size to the whole picture exceeds 0.075 is not integrated.

5.2.2 Visual Feature Representation

First, we created the codebook which size is 1000 to create the BoF. Then, Color-SIFT features were extracted from the positive and negative images.And we created the BoF of dimension 1000 using a codebook from each region.

5.2.3 Positive region determination

Positive region determination was performed using the mi-SVM. We used SVM-light[8] as the program of SVM. In this experiment, we estimated the positive region by repeating five times in the training and test by SVM. So we judged that the remaining in the final 200 is positive region.

5.2.4 Calculation of feature distribution

In this experiment, in order to calculate of the pLSA in each class using the fold-in heuristics, first, the distribution was determined based pLSA. It was used to determine the distribution base that a BoF of 20,000 randomly selected from all regions in BoF of positive

noun/adjective	−	red	blue	green	black	white	circle	square	morning	night	winter	summer	new	old	beautiful	cool
beach	5.383	0.198	0.099	−0.009	0.027	−0.059	0.018	0.181	0.338	0.305	0.101	−0.058	0.037	−0.045	0.075	0.011
bird	5.478	0.147	0.193	0.182	0.029	−0.045	−0.009	0.115	0.321	0.034	0.103	0.212	−0.023	−0.012	0.063	0.082
boat	5.398	0.193	0.123	−0.065	0.110	−0.034	−0.045	0.122	0.440	0.297	0.095	0.020	0.065	−0.050	0.197	−0.053
bridge	5.466	0.071	0.354	0.161	0.232	0.078	−0.018	0.151	0.336	0.143	0.003	0.042	0.085	−0.028	0.016	−0.022
car	5.486	0.139	0.105	0.003	0.130	0.118	0.131	0.035	0.101	0.129	0.049	0.044	0.150	−0.003	0.018	0.039
cat	5.521	0.003	0.061	0.046	0.145	0.117	0.061	0.092	0.032	0.083	0.069	0.064	0.046	0.044	0.070	0.048
cloud	5.334	0.078	0.066	−0.020	0.154	−0.024	0.030	0.217	0.220	0.135	0.063	−0.064	0.069	−0.005	0.086	0.014
cup	5.431	0.105	0.137	0.100	0.121	0.150	0.073	0.096	0.169	0.103	0.132	0.013	−0.027	−0.060	−0.015	−0.005
dog	5.522	0.027	0.024	0.069	0.120	0.124	0.144	0.086	0.137	0.211	0.069	0.048	0.050	0.038	0.066	0.008
flower	5.357	0.096	0.185	0.145	0.082	0.055	−0.040	0.175	0.153	0.088	0.011	0.077	0.106	−0.128	0.018	0.030
fruit	5.474	0.112	0.113	0.157	0.242	0.085	0.042	0.113	0.006	0.050	0.117	0.149	0.007	−0.046	0.061	0.018
house	5.536	0.114	0.170	0.163	0.161	0.060	0.040	0.091	0.224	0.078	0.129	0.033	−0.011	0.093	−0.003	−0.001
people	5.519	0.084	0.047	0.024	0.114	0.078	0.035	0.013	0.164	0.153	0.093	0.020	0.134	0.058	0.090	0.040
sea	5.368	0.211	−0.022	−0.038	0.198	−0.030	−0.032	0.108	0.439	0.237	0.198	−0.066	0.056	−0.021	0.077	−0.006
sky	5.387	0.188	0.108	0.016	0.146	0.030	−0.026	0.036	0.287	0.237	0.011	0.048	0.053	−0.022	0.006	−0.002
snow	5.490	0.036	0.261	0.038	0.084	0.044	−0.014	0.167	0.279	0.159	0.054	−0.009	−0.013	0.047	0.084	0.077
sun	5.380	0.278	0.044	0.027	0.069	−0.008	0.176	0.042	0.237	0.248	0.069	0.007	−0.016	−0.067	0.179	0.013
tower	5.473	0.113	0.234	0.051	0.151	0.046	0.022	0.063	0.443	0.101	0.044	0.012	0.043	0.037	0.015	0.015
train	5.535	0.056	0.133	0.054	0.242	0.128	0.149	0.071	0.036	0.145	0.028	0.016	0.040	0.045	0.050	0.023
tree	5.437	0.014	0.137	0.072	0.183	0.058	−0.022	0.173	0.376	0.186	0.164	0.056	0.046	0.056	−0.003	0.011

Fig. 1: Calculation result of mutual information (red: high relation class, blue: low relation class)

noun/adjective	red	blue	green	black	white	circle	square	morning	night	winter	summer	new	old	beautiful	cool
beach	0.678	0.492	0.650	0.646	0.630	0.802	0.983	0.620	0.639	0.669	0.445	0.669	0.715	0.539	0.717
bird	0.587	0.518	0.566	0.550	0.563	0.798	0.960	0.666	0.776	0.602	0.708	0.758	0.757	0.616	0.726
boat	0.618	0.516	1.556	0.683	0.647	0.676	0.954	0.606	0.629	0.725	0.575	0.690	0.593	0.618	0.710
bridge	0.646	0.579	0.616	0.623	0.619	0.665	0.798	0.600	0.487	0.613	0.680	0.584	0.567	0.640	0.728
car	0.508	0.556	0.613	0.523	0.573	0.766	0.918	0.747	0.573	0.704	0.683	0.623	0.425	0.666	0.542
cat	0.666	0.624	0.630	0.462	0.518	0.884	0.934	0.761	0.735	0.754	0.774	0.794	0.759	0.660	0.714
cloud	0.579	0.422	0.552	0.588	0.532	0.666	0.859	0.462	0.616	0.640	0.630	0.731	0.651	0.548	0.617
cup	0.659	0.721	0.720	0.711	0.679	0.671	0.943	0.628	0.853	0.853	0.858	0.700	0.734	0.831	0.770
dog	0.638	0.621	0.646	0.477	0.528	0.828	0.925	0.744	0.785	0.617	0.684	0.746	0.720	0.697	0.708
flower	0.405	0.480	0.379	0.579	0.408	0.724	0.878	0.666	0.730	0.709	0.523	0.739	0.765	0.517	0.707
fruit	0.508	0.687	0.534	0.694	0.663	0.667	0.890	0.699	0.809	0.779	0.647	0.812	0.735	0.707	0.671
house	0.594	0.583	0.555	0.604	0.543	0.722	0.895	0.689	0.597	0.618	0.649	0.521	0.434	0.623	0.657
people	0.597	0.589	0.600	0.525	0.524	0.763	0.788	0.700	0.506	0.640	0.527	0.625	0.576	0.474	0.604
sea	0.541	0.394	0.571	0.579	0.560	0.788	0.917	0.588	0.600	0.614	0.472	0.700	0.615	0.525	0.699
sky	0.463	0.226	0.415	0.535	0.444	0.696	0.806	0.489	0.450	0.504	0.480	0.645	0.599	0.498	0.635
snow	0.633	0.560	0.669	0.644	0.435	0.732	0.922	0.603	0.567	0.157	0.763	0.667	0.717	0.653	0.717
sun	0.495	0.408	0.468	0.530	0.496	0.673	0.871	0.420	0.606	0.515	0.416	0.664	0.582	0.405	0.585
tower	0.679	0.573	0.663	0.666	0.629	0.722	0.728	0.649	0.522	0.683	0.725	0.668	0.557	0.670	0.713
train	0.697	0.694	0.731	0.664	0.681	0.748	0.910	0.690	0.649	0.684	0.759	0.692	0.571	0.742	0.711
tree	0.483	0.447	0.376	0.536	0.483	0.709	0.826	0.541	0.565	0.447	0.601	0.691	0.574	0.558	0.654

Fig. 2: Calculation result of co-occurrence of tags by NGD (red: high relation class, blue: low relation class)

image. Perform clustering using the pLSA for this BoF to determine the feature distribution. The number of clusters in clustering in this case was 300.

5.2.5 Calculation of mutual information

The entropy was calculated using the joint probability $P(z|d)$ which was calculated by pLSA. Then, we calculated mutual information MI using entropy. Calculation result will be published in the next section.

5.2.6 Tag relation

First, we examined the search number by tag-search about tag X, and the search number by tag-AND-search about pair of tag X, Y. Next, we calculated the relation using the NGF from the search number. This calculation results are also posted in the next section.

5.3 Experimental results

We calculated mutual information for each class. Fig 1 shows the calculation result of the mutual information. We posted a decrement from the entropy of the class of nouns only, in the class combined with the adjective. In addition, Fig 2 shows the calculated results on the similarity of NGD using tag co-occurrence.

6 Discussion

From the experimental results, we compare mutual information of each class. Mutual information decreases when the distribution of images in each class spread, and increases when the distribution of images in each class is narrow. Then, we determine the classes which have amount of mutual information have high visual relation between nouns and adjectives. Moreover, we determine the classes which have small NGD have high visual relation between nouns and adjectives.

6.1 Discussion on visual relations

When we pay attention about the class in combination with the adjective about a color, it turn out that

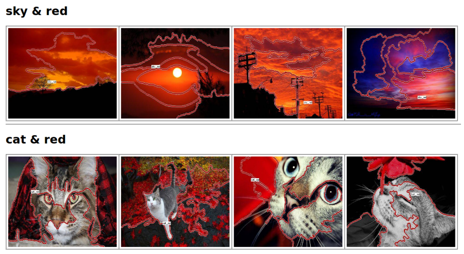

Fig. 3: Positive regions in the class combined with an adjective about color

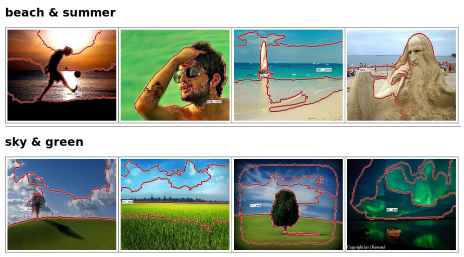

Fig. 4: Positive regions of the class which the tag relation is high, and visual relation is low

mutual information becomes increase in the class that the adjective about color qualify directly to the object being indicated by the noun (see Fig 3). We would mention "red sun" and "red car" class as the example of large mutual information, and "red cat" and "red dog" class as the example of small mutual information in the class which combined with the adjective about a color. When we think about such a class, in the class mutual information is greater, positive region of that class contains the particular color and object. Whereas, in the class mutual information is smaller, positive region of that class don't contain the particular color and object. Therefore, it can be thought that visual relation has been correctly calculated based on intuition.

6.2 Comparison with tag co-occurrence

The class was also present visual relation is low nevertheless the relation by tag are high (see Fig 4). As an example, we cite "summer beach", "green sky" class. It is thought that visual relation became low in "summer beach" class, because there are not only the image of a beach but many images of the people who are doing sea bathing. And, it is thought that the relation by tag became high in "green sky" class, because "green sky" class contains many images of grass, and the coincidence of a tag of sky and a tag like grass which coincides with green easily being high similarly. However, we can show the low level of relation by using the visual relation.

7 Conclusion and Future work

7.1 Conclusion

In this paper, first, we collected images tagged with both particular nouns and adjectives from Flickr. Then, we extracted local features from images, and calculated the distribution of image as the numeric value by the entropy. In addition, we performed comparison and consideration about the visual relation between a noun and an adjective from the change in entropy for each class which combined a noun and an adjective.

As a result, we could obtain the result of mutual information which represents the intuitive visual similarity. Therefore, it turned out that the visual relation in the class which combined an adjective about color is easy to show the relation between tags.

7.2 Future work

We consider creating the new dataset in consideration of visual relation, by using the analysis result of the visual relation searched in the experiments. In addition, we hope that accuracy of classification and training in a field of simultaneous recognition of a noun and an adjective will improve by using that dataset.

References

[1] T. L. Berg, A. C. Berg, and A. J. Shih. Automatic attribute discovery and characterization from noisy web data. In *Proc. of European Conference on Computer Vision*, pp. 663–676, 2010.

[2] D. Parikh and K. Grauman. Interactively building a discriminative vocabulary of nameable attributes. In *Proc. of IEEE Computer Vision and Pattern Recognition*.

[3] K. Yanai and K. Barnard. Image region entropy: A measure of "visualness" of web images associated with one concept. In *Proc. of ACM International Conference Multimedia*, 2005.

[4] H. Kawakubo, Y. Akima, and K. Yanai. Automatic construction of a folksonomy-based visual ontology. In *Multimedia (ISM), 2010 IEEE International Symposium on*, pp. 330–335. IEEE, 2010.

[5] K. Yanai, H. Kawakubo, and B. Qiu. A visual analysis of the relationship between word concepts and geographical locations. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, p. 13. ACM, 2009.

[6] R.L. Cilibrasi and P.M.B. vitanyi. The google similarity distance. *Knowledge and Data Engineering, IEEE Transactions on*, Vol. 19, No. 3, pp. 370–383, 2007.

[7] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 8, pp. 800–810, 2001.

[8] T. Joachims. *SVM ^{light}: Support Vector Machine.* http://svmlight.joachims.org/.