

Identifying Tags Describing Image Contents

Taka Maenishi, Keishi Tajima (Kyoto University)

BACKGROUND

Some tags do not describe its image contents.



Tags

#flower
#iPhone

Photo of flower taken by iPhone

“iPhone” is not a subject of the photo
(It is a kind of metadata)



Tags

#iPhone
#new

Photo of new iPhone

“iPhone” is a subject of the photo

OUR METHOD

We propose how to distinguish tags describing its image contents.

Subject Specificity: The degree of how much a tag narrows down the subjects shown in the associated images.



A set of images associated with “flower nature” is more similar to “flower” than to “nature”, which means “flower” is likely to be more subject-specific than “nature”.

How to compute similarity between two image sets

Vector representing word frequency: We convert each image set into a vector representation. The vector of an image set associated with a word A , denoted by $V(A)$, is defined as below:

$$V(A) = (\sum_i c_{A,t_i})^{-1} \cdot \langle c_{A,t_1}, \dots, c_{A,t_n} \rangle$$

$$c_{A,t} = \frac{|S_{At}|^2}{|S_A| \cdot |S_t|}$$

t_1, \dots, t_n : Tags that appear in the data set

S_A : Set of images associated with A

S_{At} : Set of images associated with A and t

flower		nature		flower nature	
bloom	0.068	sky	0.028	bloom	0.083
floral	0.062	landscape	0.026	blossom	0.075
botanical	0.047	cloud	0.024	petal	0.061
nature	0.030	green	0.022	plant	0.056
blossom	0.029	outdoor	0.017	floral	0.040
flor	0.025	beautiful	0.016	green	0.039
rose	0.024	forest	0.016	beautiful	0.037
petal	0.023	travel	0.016	spring	0.036
spring	0.022	mountain	0.014	summer	0.034
summer	0.021	grass	0.014	forest	0.032

Top 10 Tags of Flower, Nature, and Flower Nature, collected from 1000 Instagram posts for each vector

Similarity between vectors: We define the similarity between two vectors $V(A)$ and $V(B)$, denoted by $sim(A; B)$, as follows:

$$sim(A; B) = 1 - JSD(V(A)||V(B))$$

$$JSD(p||q) = \frac{1}{2}KL(p||r) + \frac{1}{2}KL(q||r) \quad \text{where } r = (p + q)/2$$

$KL(p||q)$: Kullback-Leibler divergence from p to q

A	B	$sim(A, B)$
flower	flower nature	0.876
nature	flower nature	0.550

Similarity of Flower/Nature to Flower Nature

EVALUATION

We determined which tag is more relevant to the image contents by using our method, and compared the result with a simple baseline method: “A tag appearing earlier in the tag list defined by the user is more relevant”.

The result shows our method is better than the baseline.

Method	Accuracy
Our method	131/170 (= 0.771...)
Baseline	97/170 (= 0.571...)