

## WEB上のデータを用いた画像の自動タグ付け

東将央 †

佐藤晴彦 †

小山聡 †

栗原正仁 †

北海道大学大学院情報科学研究科

## 1 研究背景

近年インターネットやデジタルカメラの普及により個人で大量の画像を所持していることが多くなった。この大量の画像を管理するため、画像それぞれに対してタグと呼ばれる内容をよく表す単語をつける方法が用いられている。しかし、対象の画像が多くなってしまおうとひとつひとつタグをつけていくことは非常に手間のかかる作業となってしまう現実的ではない。よって画像に対して自動的にタグ付けを行う方法が検討されている。

## 2 研究目的

本稿では、WEB上にある画像を一種の大規模なデータベースとして利用し、入力画像に対し自動的にタグ付けを行う方法について検討する。近年普及し始めた類似画像検索などを用いることによって、大量の画像とその掲載ページのペアが容易に大量に取得できるようになった。画像の掲載されているページにはその画像を説明するための有効な単語が含まれている場合が多く、文章から画像に含まれる内容を推定することができる。ただし、全ての文章が画像を説明しているわけではなく、タグとしては意味を成さない単語も多く含まれるため、どの単語が有効であるかについてスコア付けを行い適切なタグを付与することを目標とする。

## 3 類似研究

画像へのタグをつける方法としてデータベースを用いて入力画像へと自動タグ付けを行う方法 [1] が研究されている。これは類似した画像には同じ内容のタグが付与される可能性が高いということを用いており、画像の類似度を比較することによりタグ付けを行っている。

また、WEBを利用する方法として、単語の共起を用いて画像に対して重要な説明文を抜き出す研究 [2] や、HTML文章内の重要語を用いて自動的にタグ付けを行う手法 [3] も検討されている。

## 4 提案手法

本章では、類似画像検索の結果を用いた画像へのタグ付けの手法について説明する。まず、Googleの類似画像検索を用いて、WEB上にある画像の中からタグ付けの対象となる入力画像と類似した画像およびその掲載ページのペアを上位50件分取得する。

その後、掲載ページに登場する単語に対して、TF-IDFにより各ページ内での重要度を表すスコア付けを行う。各ページ内の単語の抽出は形態素解析器を利用し、そのうちタグとして有用であると考えられる名詞のみを取り出す。

ページ  $j$  にある単語  $i$  のタグスコア  $tfidf_{i,j}$  は以下のようになる。

$$tfidf_{i,j} = TF_{i,j} \times IDF_i \quad (1)$$

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (2)$$

$$IDF_i = \log \frac{|D|}{|\{d : d \ni i\}|} \quad (3)$$

$n_{i,j}$  は、単語  $i$  のページ  $j$  における出現回数。  $|D|$  はWEBページ全体のページ数、  $|\{d : d \ni i\}|$  は単語  $i$  で検索を行った際のヒット数である。

次に画像の類似度を算出する。画像の特徴量としてColor HistogramとSIFTを使ったBag-of-Visual-Words法(BoVW)の2つを用いる。Color Histogramは、画像中に出現する色の出現数をヒストグラム化したものである。今回はRGBを各4色まで減色処理を行い、計64色でヒストグラムを構成する。SIFTは、128次元のベクトルを持つ画像の局所的な特徴量であり、画像の回転・拡大縮小・明暗の変化に対して非常にロバストな性質をもつ特徴量である。ただし、SIFTは画像ごとに特徴ベクトル数が大きく異なるため、特徴ベクトルに対してBoVW法を用いる。BoVW法は、文書を単語の出現回数でベクトル化するBag-of-Words法(BoW)の画像への応用であり、特徴ベクトルの出現回数で画像全体をベクトル化するものである。この2つの特徴量を用いて、入力画像とデータセット内の画像の類似度を計算する。

Color Histogramを用いた色類似度  $Sim_{CH}$  とSIFTをBoVWによってベクトル化した形状類似度  $Sim_{SIFT}$  を

Automatic tagging of images using web data

†Nobuo AZUMA, Haruhiko SATO, Satoshi OYAMA, Masahito KURIHARA  
Graduate School of Information Science and Technology, Hokkaido University

以下のように定義する .

$$Sim_{CH}(H_1, H_2) = \frac{\sum_{i=0}^{64} \min\{H_1(i), H_2(i)\}}{\sum_{i=0}^{64} H_1(i)} \quad (4)$$

$$Sim_{SIFT}(\vec{V}_1, \vec{V}_2) = \frac{\vec{V}_1 \cdot \vec{V}_2}{|\vec{V}_1| \times |\vec{V}_2|} \quad (5)$$

さらに, この2つの類似度を用いて, 画像  $i$  と画像  $j$  の類似度を以下のように定義する .

$$Sim(i, j) = Sim_{CH}(H_i, H_j) \times Sim_{SIFT}(\vec{V}_i, \vec{V}_j) \quad (6)$$

最後に画像類似度と単語への TF-IDF を用いて入力画像に対するタグのスコアリングを行う . 類似度が大きい画像に対して付いているタグ候補は入力画像に対しても内容を説明するタグとなっている可能性が高いため, 画像類似度を重みとしてスコアの補正を行う . また, 同じ単語が複数のページにわたって出現することが考えられるため, 同じ単語のスコアを合計することにより最終的なタグのスコアとする .

$$Score(i) = \sum_{j \in D} \{Sim(img_{in}, img_j) \times tfidf_{i,j}\} \quad (7)$$

ただし,  $img_{in}$  は入力画像,  $img_j$  はページ  $j$  の画像である .

## 5 実験

画像に対して提案手法を適用しタグ付けを行い, その結果に対して, 画像類似度を利用した場合と利用しない場合を比較して評価を行う . 画像は WEB 上から取得したものをを用いた . その際, その画像が出現するページはデータセットの対象から除外し, その画像が掲載されているページの情報は用いないこととした . また, 画像に付加するタグに関してはつける人の主観が大きく関わり, WEB サイト上に登場しない単語をつけてしまう可能性が高いため, 上位 12 件のうち入力画像のタグとして有用なものがどの程度含まれているかで評価を行った . このときタグは情報を充実させるという目的の下, 「写真」や「画像」などといった画像の内容を示していないものでも許容した . 評価には 11 点平均適合率を用いた .

	Q1	Q2	Q3	Q4	Q5	Q6	Q7
TF-IDF	0.58	0.54	0.55	0.69	0.64	0.47	0.35
提案手法	0.70	0.60	0.50	0.69	0.60	0.45	0.45
タグの個数	4	4	4	4	5	3	2

表 1: 出力結果と平均適合率

この結果を見ると大きく改善したものもあるが, 若干悪化してしまったものもある . 全体としてタグとして利用できる単語が少なく, またスカイツリーが「ス

カイ」と「ツリー」に分割されてしまうなどの問題が確認できた . これは WEB ページから単語を抽出する際に, 形態素解析器を利用しているため, 名詞として登録されていないものは推定を行ったり, 連続しているのに分割されてしまうことが原因として考えられる .

また, Q7 などはタグとして「画像」「写真」というタグしか含まれず, 内容を示しているものは全くなかった . これは, 類似画像検索の結果の時点で, 同じ内容が含まれる画像の数が少なく, スコアが高くなるような単語が少なかったためと思われる . また, 画像の類似度に関しても対象物がひとつのみ写っているものであれば類似度も高くなりやすいが, 様々なものが写っていたり, 草むらや壁のようなテクスチャが画像の大部分を占めるような画像であった場合には類似度が逆に低くなってしまっても確認できた . これは, 今回用いた画像特徴量が物体特定によく用いられるものであり, 様々なものが映っている写真に対してそのまま用いるのは不適切であったものと考えられる . よって, 一般的な写真に適した画像特徴量に対して考察していく必要があると考えられる .

## 6 まとめと今後の課題

本研究では画像へ自動的にタグ付けを行うため, 画像の類似度を考慮して類似画像検索の結果から重要な単語を抽出する方法を検討した . 画像の類似度として Color Histogram と SIFT を使用した Bag-of-Visual-Words 法を用い, 単語の抽出法として形態素解析および tfidf を用いた . 実験の結果, 現在の手法では画像に対する有用なタグが得られる場合が少ないことが判明した . 今後改善すべき点として, 類似度判定の精度向上と HTML 文書からの単語抽出法の改善が挙げられる .

### 参考文献

- [1] タグの共起と画像の類似性を利用したタグ付け支援システム, 瀬崎 直人, 電気学会論文誌 (C) Vol.129 No.3 pp.475-480
- [2] 類似画像とキーワードを利用した Web 画像の説明文抽出, 竹内謹治, 2006, 情報処理学会研究報告. 自然言語処理研究会報告 pp.7-12
- [3] HTML テキストの重要文を用いた画像ラベリング手法, 相良直樹, 2004, 電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理 Vol.87 No.2 pp.145-153