

## Web 画像の自動分類とその評価 AUTOMATIC CATEGORIZATION WWW IMAGES AND ITS EVALUATION

中平 浩二†  
Koji Nakahira

上野 智史†  
Satoshi Ueno

渡井 康行‡  
Yasuyuki Watai

相澤 清晴†  
Kiyoharu Aizawa

### 1. はじめに

ウェブページは主にテキスト情報と画像情報で構成されており、画像は視覚的に重要な情報を我々に提供する。しかしながら、一般の検索エンジンでは、ウェブ情報の解析には画像を用いていない。画像はデータ量が大きいので、ウェブページの解析において、画像を無視し、テキスト情報のみを用いているのである。

写真やロゴ、バナー、タイトル、アイコンなどのさまざまな画像がウェブ上において用いられている。これらの画像には、素早くウェブページの内容や構造をユーザに伝えるという重要な役割がある。ウェブ画像をうまく扱うことにより、既存の検索エンジンでは不可能な新しいことが出来るのではないかと考えられる。

本論文では、ウェブ画像の新しい分類と、検索システムへの利用やネットサーフィンの補助やウェブサイトの要約への利用などの、分類された画像の応用について述べる。画像は、例えば動物や花といった、画像の内容そのものによって分類される事が一般的である。しかし、自然画像とは異なり、ウェブ上の画像はその使われ方によって、各々がその意味や役割を持っている。著者らはまずこの役割に応じた 8 つの分類カテゴリを定義し、ツリー構造上に Support Vector Machine (SVM) を適用した。分類には 3 種の簡単な特徴量を用いている。タグ情報、色などの画像情報、そして画像の位置などのコンテキスト情報である。130 の上場企業のウェブサイトから収集した 20,000 の画像を用いて実験を行った。既に、Google Image Search などウェブ画像を扱った検索エンジンがあるが、それらはテキスト解析のみを用いて画像を収集し、キーワード検索によるシステムである[1]。我々は、画像情報のより深い解析を必要とする、方向性の異なるウェブ画像の利用法について考えている。自然画像の認識率の向上のためにウェブ上の写真を利用するシステム[2]などでは、画像情報を用いた既存の画像検索システムの改良についての研究もなされている。しかし、そのサービスの本質は既存の画像検索システムと変わらない。我々の研究は、ロゴやバナー、写真などの様々な種類の画像を利用してネットサーフィンの向上などを図る初めての試みである。

### 2. ウェブ画像の分類

#### 2.1 SVM を用いた分類

一般に自然画像は、動物や花といった、その画像内容に応じて分類される。しかし自然画像と異なり、ウェブ上の画像はその役割や意味を持つ。たとえば、リンク画像はリンク先のサイトの内容を指し示している。ロゴはそのサイトのシンボルであることが多い。また、文中の図表や絵はその

の文の内容を指し示している。したがって、ウェブ情報を解析するとき、画像を分類化することは非常に重要な情報になると考えられる。そこで我々は Support Vector Machine を用いて、ウェブ画像をその役割や意味に応じて自動分類させる事を試みた。SVM とは二つのグループに分類させる学習システムの一つである。まず 10 の大手サイトにおいて用いられている画像について調べた。そして、その画像の用いられ方において次の 8 つのカテゴリ、“photo figure”、“non-photo figure”、“main-title”、“action-title”、“icon for menu”、“logo”、“aGverting image”、“segmenteGimage”を定義した。より細かな分類カテゴリの定義も可能であったが、階層構造の分類を実現したかったため、まずはその幹としてこれらの 8 つを定義した。SVM を用い、画像群をこれら 8 つに学習分類した。Figure1 に各カテゴリに属する画像の例を示す。

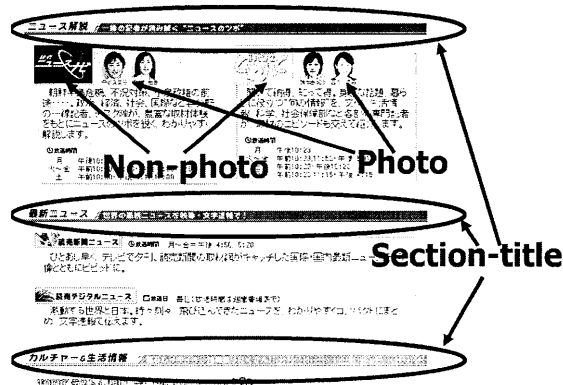


Fig.1. (angle of “action-title” and “photo” and “non-photo”)

以下は分類に用いた特徴量である。画像情報、タグ情報、コンテキスト情報の 9 つの簡単な特徴量を用いた。

“画像サイズ”、“ピクセル数”、“アスペクト比”、“色数”、“ファイルタイプ”、“リンク付きの画像かどうか”、“ページ内における画像の位置”、“画像直後のコメント量”、“同じレイアウトの画像があるかどうか”

ページ内における画像の位置、というのは 0 から 1 までの数値であり、ページの先頭にあるものは 0 となり、最後にあるものは 1 となる。

#### 2.2 実験及び評価

例えば、アイコンはそのサイズと色数は小さく、ファイルタイプは一般に gif や png である。また、“main-title”はアスペクト比が一般に大きく、ページの初めにある事が多い。“menu-icon”は同じレイアウトの画像が同ページに有る事が多い。このように直感的に有効ではないかと考えた 10 数の特徴量に対し、以下のような行列を作成した。

† 東京大学新領域創成科学研究科

‡ 東京大学工学部

Table 1. Matrix of Correlation

	adv	seg	logo	illust	section	photo	icon	title
adv	0.12	0.333	0.22	0.696	0.351	0.731	0.27	0.34
seg	0.33	0.02	0.21	0.887	0.283	1.034	0.11	0.49
logo	0.22	0.205	0.2	0.77	0.307	0.857	0.19	0.39
illust	0.7	0.887	0.77	0.975	0.854	0.9	0.82	0.76
section	0.35	0.283	0.31	0.854	0.393	0.939	0.28	0.5
photo	0.73	1.034	0.86	0.9	0.939	0.699	0.94	0.74
icon	0.27	0.106	0.19	0.821	0.278	0.942	0.12	0.43
title	0.34	0.485	0.39	0.761	0.498	0.744	0.43	0.46

Table1 は特徴量の一つ、”画像サイズ”に対する相関行列である。各セルは、行と列に値するグループに属している要素集団に対し、 $1/(N*N) * \sum \sum |a_i - b_j|$  で得られる値となっている。この特徴量は“advertising image”や“segmented image”を分類するのに有効な特徴量と言えるが、“segmented image”などの分類には不適切な特徴量と言える。この様にして有効な特徴量として選別した結果が前の9つの特徴量であった。上場企業のウェブサイトから130のサイトをランダムに選び、それらに含まれていた19,324の画像を対象とした。1,500の画像を学習サンプルとしてS<sup>S</sup>Mを用いて学習分類実験を行った。Table2 は実験の結果である。平均適合率は0.42となり、また平均再現率は0.76となった。次に、g48 アルゴリズムによる決定木を用いて1,500 ファイルを教師として2,800 ファイルを対象として分類学習を行った。結果をTable3 に示す。平均適合率は0.61となり、平均再現率も0.61となった。また、決定木の葉数は141で、枝数は281となった。決定木を用いた場合と比較して、S<sup>S</sup>Mを用いた場合、適合率が低い結果となった。しかしながら、”Advertising images”以外のカテゴリでは適合率はとても良い値となっている。このカテゴリに割り当てられた画像数がとても多い為に全体の適合率が下がっているが、他のカテゴリに属するべき画像が多数このカテゴリに分類されてしまっていると考えられる。そこでこのカテゴリに再度S<sup>S</sup>Mを適用し、“advertising images”を抽出し、それ以外を分類失敗画像とした。その結果をTable 3 の下部に示す。この結果平均適合率は0.90となったが、平均再現率は0.32となってしまった。しかしながら、この分類失敗画像のグループに対し、再度8つのカテゴリへの学習分類を行えば適合率も再現度も非常に良くなると考えられる。

他のカテゴリの再現率を見るに、分類失敗画像の多くは”figure (photo)”, ”figure (non-photo)”, ”main-title”, 及び”logos”に属するべき画像だと考えられる。色ヒストグラムなどの画像内容に基づく特徴量を用いる事で十分な精度でこれらのカテゴリに分類できるのではないかと考えている。また、決定木を用いた場合では、いくらかの特徴量を加えてもこれ以上精度と再現率の上昇は難しい。したがって、S<sup>S</sup>Mを用いて研究を続ける予定である。

Table 2. Results of experiments using S<sup>S</sup>M

Category	Number (%)	Precision	Recall
Figure (photos)	1521 (7.9%)	0.80	0.15
Figure (non-photos)	337 (1.8%)	0.87	0.10
Icons for menu	3345 (17.4%)	0.97	0.90
Section-title	1761 (9.2%)	0.90	0.87
Main-title	233 (1.2%)	0.92	0.25
Logos	306 (1.6%)	0.76	0.36
Segmented images	791 (3.7%)	0.93	0.77
Advertising images	10642 (55.9%)	0.06	0.98
Advertising images (After a re-classification)	709 (4.1%)	0.82	0.91
Classification failure	9933(51.8%)	—	—

Table 3. Experimental results using the decision tree

Category	Number (%)	Precision	Recall
Figure (photos)	1086 (42.1%)	0.76	0.81
Figure (non-photos)	384 (14.9%)	0.44	0.28
Icons for menu	488 (18.9%)	0.60	0.65
Section-title	247 (9.6%)	0.51	0.37
Main-title	116 (4.5%)	0.40	0.32
Logos	89 (3.4%)	0.20	0.29
Segmented images	129 (5.0%)	0.60	0.92
Advertising images	41 (1.6%)	0.22	0.15

### 3. まとめ

本論文において、テキスト情報と画像情報を用いたウェブ画像の意味や役割に基づく自動分類と、その評価を行った。現在はより深く画像情報を用いる事により、さらに細かい分類を行う事について研究している。また、色情報や形状情報を用いてウェブのマイニングやクラスタリングにも役立てられないか、これらを組み合わせて従来の画像検索エンジンを本質的に改良できないかについても研究している。

### 参考文献

- [1] Sergey Brin and Lawrence Page, “The Anatomy of a Large-Scale Hypertextual Web Search Engine”, Computer Networks and ISDN Systems pp.107-117, vol. 30, 1998.
- [2] S. Sclaroff, M. Cascia, S. Sethi, and R. Ayer, “Refining textual and visual cues for content-based image retrieval on the World Wide Web”, Computer Vision and Image Understanding, 75(1/2):66-98, 1999.