

# Semi-Supervised Learning を用いた Web 画像収集システム

## An Web Image-Gathering System Employing Semi-Supervised Learning

柳井 啓司  
Keiji Yanai

電気通信大学 情報工学科

Department of Computer Science, The University of Electro-Communications

We propose introducing a probabilistic method for image selection into the Web image gathering system. The existing system consists of two step processing: (1) gathers HTML files of Web pages related to keywords, analyzes them and fetches only Web images expected to be highly related to the keywords. (2) selects only relevant images from the gathered images based on the image-feature-based clustering. In this paper, we propose constructing a visual generative model based on the GMM (Gaussian Mixture Model) to represent the distribution of image related to the keywords, and using it to select images in the Web image gathering system.

### 1. はじめに

近年, World Wide Web に対する Web 画像検索の研究が盛んになってきている. Web 画像検索は, 現在 Google, Lycos などをはじめとする商用 Web 検索エンジンがその検索サービスの一つとして一般に提供しているが, これらは手軽に利用できる反面, 基本的に画像内容を解析せずにテキスト検索の場合と同様にページに含まれるキーワードやリンク情報のみを手がかりにした検索を行っており, 無関係な画像が検索結果として返されることが多く, 検索精度は高くない.

それに対して, HTML ファイルのキーワード解析に, 画像の特徴量に基づく画像検索 (Content-Based Image Retrieval, CBIR) を組み合わせた Web 画像検索の研究が提案されている. その中でも代表的な研究が, WebSeer[Framkel 96], WebSEEK[Smith 97], Image Rover[Sclaroff 99] の3つである. これらは, 検索時にユーザが欲しい画像に関係するキーワードを入力する点ではほぼ同じで, 検索に用いるキーワードについても「WWW 上の画像は HTML 文書にそのファイル名が記されてリンクされており, 画像を説明するキーワードが HTML 文書に含まれていることが多い」という経験則を利用して, 通常の HTML 文書の検索手法と同等の方法を用いて, 画像をリンクしている HTML 文書からキーワードを抽出して, その画像を検索する場合のキーワードとしている. これらでは, 初めにキーワードのみによって画像を検索し, さらにその検索結果の画像群の中からユーザが望む画像を指定して, さらに CBIR の手法を用いて, 指定画像に類似している画像を検索するという, 2段階の検索が前提となっている.

また, 近年提案された 4 つのシステム [Lempel 01, 獅々堀 04, Feng 04, Cai 04] では, Google Image Search を用いて与えられたキーワードに関連する画像を収集し, それをユーザーからのフィードバック情報, もしくは予め用意したデータベース中の知識を用いて, [Lempel 01] ではリンク情報に基づいて, [獅々堀 04] では画像特徴量のみに基づいて, [Feng 04] や [Cai 04] では画像特徴量に加えて HTML ファイルのテキスト情報, リンク情報も用いて画像の絞りこみを行い, よりユーザに適した検索結果を提示する.

以上述べた Web 画像検索の研究は, 従来の Web 検索エンジンの検索精度を高めることを目的としており, 少量の精度の高い検索結果をユーザに返すことが最終目的になっている. そのためにユーザとのインタラクションによる結果の精錬化を前提としている場合が多く, 処理中のユーザの介入が前提となっ

ている. そのため, 処理の高速性にも重点がおかれていることが特徴である.

一方, 本研究における「Web 画像収集」においては, 少数の検索結果をユーザに提示することを目的とする「Web 画像検索」とは異なり, 画像の意味内容を計算機に自動的に認識させる「一般画像認識 (Generic Object Recognition)」を実現するための画像知識ベース自動構築のために, 多様な画像を多数 Web から自動的に収集することを目的とする.

我々は, Web 上の画像を学習画像として用いて, 一般の実世界画像に対する自動分類を行うことを提案している [Yanai 03, 柳井 04]. Web 上の画像 (Web 画像) は, 様々な人が様々な目的で撮影した画像であり, 類似している画像が多く含まれる商用の画像データベースとは異なり, 実世界の一般的な画像の多様性をそのまま反映していると考えられる. また, Web 画像はそれを含んでいる Web ページの HTML 文書を解析することによって, 画像に関連するキーワードを抽出することが可能であるという特徴を持つために, 目的の画像を自動収集することが可能である. 我々は, この一般画像認識のための Web からの画像収集を「Web 画像マイニング」と呼んでいる. 他にも, Web 上の画像を用いた物体認識の研究 [Song 04] は存在しており, 今後同種の研究は増加していくことが予想される.

「Web 画像マイニング」においては, Web から精度よく画像を収集することが必要である. そこで本研究では, 我々がすでに提案しているキーワードに関係する画像を, 大量に収集することを目的とする Web 画像収集システム Image Collector [柳井 01, Yanai 04] を改良することによって, より精度の高い Web 画像収集の実現を目指す. 具体的には, 確率的手法による画像選択法を提案し, それにより従来研究よりも高精度の画像収集を実現する.

本論文の構成について簡単に述べる. 2章では従来のシステム, および確率的手法を用いた Web 画像収集の方針について述べる. 3章では方法の詳細について述べ, 4章では画像収集実験を行い, 最後に 5章で全体のまとめを行う.

### 2. 方針

本研究では, 主に Image Collector [柳井 01, Yanai 04] の画像特徴量解析による画像選択部分に新たな手法を導入することによって, 画像収集の高精度化を実現する.

Image Collector は, 与えられたキーワードに関係する画像を Web から数百枚程度収集することを目的とする. 具体的には, 処理は 2 つのフェーズからなり, テキスト解析および画像収集フェーズ, 画像特徴量に基づく画像選択フェーズから

成る。(1) まず最初に、テキスト解析フェーズとして、与えられたキーワードに関する Web ページの URL を Google などのテキスト検索エンジンから獲得し、それらすべての Web ページの HTML ファイルを収集する。そして、HTML ファイルから画像ファイルの URL を抽出し、それを HTML 解析によって A, B, C の 3 つにランク付けする。A はキーワードとの関係が強く、B は中程度、C は無関係とみなして、A, B にランク付けされた画像のみを実際に収集する。(2) 次に画像解析フェーズとして、収集したすべての A, B 画像から画像 1 枚ごとに画像特徴を抽出する。そして、A 画像に対して階層的クラスタリングを行い、小さなクラスタを除去し、大きなクラスタの平均画像に近い画像を B 画像から選択し、あわせて最終収集結果とする。

本研究では、(2) の部分に semi-supervised learning に基づく確率的画像選択手法を導入する。従来の階層的クラスタリングの代わりに、EM アルゴリズムによる確率的クラスタリングを導入し、A 画像を正例、予めランダムに収集した非キーワード画像を負例として、正例クラスタ、負例クラスタを選択する。ただし、1 枚の画像を 1 つの画像特徴ベクトルとしていた従来の方法と違って、各画像を領域分割して、それぞれの領域から画像特徴ベクトルを抽出する。そして、正例、負例の領域のそれぞれの生成確率モデルを構築し、それを用いて A 画像、B 画像の各領域が正解領域である確率を求め、画像ごとに上位の確率の領域の平均をとって正画像である確率を求める。ここでは、さらに A, B 各画像の領域の正領域である確率が上位の領域を正例として仮定し、正解領域のクラスタの推定、A, B 画像の各領域が正画像である確率の計算を再度数回繰り返していき、最終的に確率の高い順に画像を出力する。つまり、最初は A 画像のみを正解画像と仮定して、それによって構築されたモデルに基づいて B 画像からも画像を選択し、2 回目以降の繰り返しでは選択された B 画像も正例であると見なし、学習と選択の繰り返しを行う。これは、A 画像をラベル付き画像、B 画像をラベルなし画像と見なし、semi-supervised learning を行っていることと等価である。

### 3. 方法

初めに Image Collector と同様であるテキスト処理フェーズの説明を簡単に行い、次に新しく提案する確率的画像選択法を用いた画像選択フェーズの説明を行う。

#### 3.1 画像収集と HTML 解析

ここは従来システムの Image Collector と同じあるので、簡単に説明する。

最初に収集した画像に関するキーワードをシステムに与える。キーワードには、システムが検索エンジンに渡す検索エンジン用キーワードと HTML 文書の解析時に用いる画像分類用キーワードの 2 種類がある。

- (1) 既存の商用テキスト検索エンジンを利用し、ユーザの与えたキーワードに関する Web ページの URL (Universal Resource Locator) を集める。検索エンジン用キーワードを検索エンジンに送る。
- (2) 集めた URL が示す Web ページにアクセスして、各 Web ページの HTML 文書を獲得する。
- (3) 各 HTML 文書に対して HTML タグに基づく解析し、HTML 文書からリンクされている画像ファイルとキーワードとの関係の強さについての評価を、画像ファイルへのリンクタグやタグ周辺のテキストにキーワードがどの程度含まれているかなど調べることによって行い、評

価の高いものから順に A, B, C にランク分けする。ここでは、画像分類用キーワードを用いる。

- (4) A ランク、B ランクに該当した画像を Web から収集し、それぞれ A 群画像、B 群画像と呼ぶ。C ランクに該当した画像はキーワードと無関係の画像と見なし収集しない。

ここで用いられている HTML タグを利用したキーワードと画像ファイルの関係の強度に関する評価方法は、テキスト検索エンジンにおいてキーワードと HTML 文書との関係の強度の評価に用いられている方法と類似した方法であり、Web 画像検索システムにおいては一般的に用いられている手法である [Rowe 98, Sclaroff 99, Framkel 96]。

#### 3.2 画像解析に基づく確率的画像選択

初めに収集したすべての A 画像、B 画像を領域分割する。ここでは JSEG [Deng 01] を用いた。そして、ある一定以上のサイズの領域に対して特徴量抽出を行う。特徴量は色、テクスチャ、形状の 3 通りのものを組み合わせて用いた。色に関しては  $Lu^{*}a^{*}$  の平均値と分散、テクスチャに関しては 2 種類のガウシンフィルタの 4 方向への平均値、形状に関しては 1 次モーメントの値、および面積を周囲長の 2 乗で割った値、相対サイズ、重心の相対位置など 10 次元を用いた。全部で 24 次元ベクトルとなった。また、後で述べる負例の生成モデルの構築のため、予めランダムに非キーワード画像を収集しておき、同様に各領域の特徴ベクトルを求めておく。

次に、すべての A 画像と、予めランダムに収集した非キーワード画像のそれぞれの領域の特徴ベクトルを一緒にして、EM アルゴリズムを用いて、ガウス混合分布モデル (Gaussian Mixture Model) による確率的クラスタリングを行う。ここでは、ガウス分布の数を  $m_0$  個用いるとする。

次に、A 画像の各領域の特徴ベクトルを正例、予めランダムに収集した非キーワード画像の各領域の特徴ベクトルを負例として、正例クラスタ、負例クラスタの選択を行う。具体的には、 $c_j$  を GMM の  $j$  番目のガウス分布の要素、 $r_i^A$  を A 画像の  $i$  番目の領域、 $n_a$  を A 画像のある一定サイズ以上のすべての総領域数とすると、キーワード  $X$  に対応するクラスタが  $j$  番目のガウス分布要素である確率  $P(c_j|X)$  は、ベイズの定理より、

$$P(c_j|X) = (1/n_A) \sum_{i=1}^{n_A} P(c_j|r_i^A, X) \quad (1)$$

$$= (\alpha/n_A) \sum_{i=1}^{n_A} P(r_i^A|c_j, X) P(c_j) \quad (2)$$

となる。ただし、 $\alpha$  は正規化定数とする。このうち、 $P(c_j|X)$  の上位  $m_1$  個のガウス分布要素をキーワード  $X$  が対応する領域を表すクラスタを表すガウス分布要素とする。つまり、A 画像の各領域を正領域と見なし、その特徴ベクトルを代入し、A 画像の全領域に対してそれぞれのガウス分布要素が  $X$  を表す確率を求めて、上位のガウス分布要素に対応するクラスタをキーワード  $X$  の正例クラスタとする。同様に、非キーワード画像に対しても  $P(c_j|nonX)$  を求め、上位  $m_2$  個のガウス分布要素を求める。

次に選択された正例クラスタ、負例クラスタを用いて、 $X$ ,  $nonX$ , それぞれの生成モデル次のように構築する。A 画像、B 画像の各領域  $r_i$  に対して、以下の式で表される  $p1(X|r_i)$  および  $p1(nonX|r_i)$  を求める。

$$p1(X|r_i) = \sum_{k=1}^{m_1} w_{1,k} \frac{1}{\sqrt{(2\pi)^N |\Sigma_{1,k}|}}$$

$$\exp^{-\frac{1}{2}(X-\mu_{1,k})^T \Sigma_{1,k}^{-1}(X-\mu_{1,k})} \quad (3)$$

$$p1(\text{non}X|r_i) = \sum_{k=1}^{m_2} w_{2,k} \frac{1}{\sqrt{(2\pi)^N |\Sigma_{2,k}|}} \exp^{-\frac{1}{2}(X-\mu_{2,k})^T \Sigma_{2,k}^{-1}(X-\mu_{2,k})} \quad (4)$$

$N$  は特徴ベクトルの次元数,  $w_{1,k}$ ,  $\mu_{1,k}$ ,  $\Sigma_{1,k}$  をそれぞれ,  $X$  に対応する  $k$  番目のガウシアン要素の重み, 平均ベクトル, 分散共分散ベクトルとする. ただし,  $\sum w_{1,k} = 1$  となるように正規化されているとする. 負例についても同様である.

そして,  $p1(X|r_i)$  および  $p1(\text{non}X|r_i)$  から各領域がキーワード  $X$  を表す画像の一部である確率  $P(X|r_i)$  を A 画像, B 画像から抽出した全領域に対して, 以下のように求める.

$$P(X|r_i) = \frac{p1(X|r_i)}{p1(X|r_i) + p2(\text{non}X|r_i)} \quad (5)$$

そして, A 画像, B 画像の全領域の中で,  $P(X|r_i)$  がある一定値より大きい領域を正例として選び出す. 同様に負例の選択も行う. ここまでで, A 画像を正例として生成モデルを学習し, A 画像, B 画像の領域から正例, 負例を抽出したことになる.

ここで, 繰り返し計算の 2 回目として, 最初の処理によって, 正解領域として選択された A, B 画像からの領域と, 予めランダム収集した非キーワード画像のそれぞれの領域の特徴ベクトルを一緒にして, EM アルゴリズムを用いて, ガウス混合分布モデル (Gaussian Mixture Model) による確率的クラスタリングを再度行う. ここでも, 最初の場合と同様に, ガウス分布の数を  $m_0$  個用いるとする.

そして, 再び正例, 負例クラスタの選択を行う. そして, それを用いて, キーワード  $X$  に対応するクラスタが  $j$  番目のガウス分布要素である確率  $P(c_j|X)$  を以下の式から求める.

$$P(c_j|X) = (1/n_A) \sum_{i=1}^{n_{sel}} P(c_j|r_i^{sel}, X) \quad (6)$$

$$= (\alpha'/n_A) \sum_{i=1}^{n_{sel}} P(r_i^{sel}|c_j, X) P(c_j) \quad (7)$$

となる. ただし,  $r_i^{sel}$  を  $i$  番目に選ばれた領域,  $n_{sel}$  を選ばれた領域の総数,  $\alpha'$  は正規化定数とする. そして, 同様に,  $P(c_j|X)$  の上位  $m_1$  個のガウス分布要素をキーワード  $X$  が対応する領域を表すクラスタを表すガウス分布要素として, 正例クラスタの再推定を行う. これを数回繰り返して, 収束もしくはある一定回数以上 (実験では 5 回) 繰り返しを行ったら, 最終的に正例クラスタを確定させる.

ここでは, 最初に A 画像を正例, B 画像を未知のデータとして, 繰り返しによって, 両方のデータを用いて, キーワード  $X$  に対応する領域の特徴ベクトルの生成モデルを学習していることとなる. つまり, 準教師あり学習を行っていることとなる. これは, 大量のテキストを分類する場合に用いられている方法 [Nigam 00] と同様のものである. ただし, [Nigam 00] では教師なしデータが教師付きデータに比べて十分に多い場合を扱っているが, 本研究では教師付きデータと, 正解がある程度含まれることが予め分かっている教師なしデータを用いるという違いがある.

最後に, それぞれの画像がキーワード  $X$  を表している確率  $P(X|I)$  を求める. これは画像  $I$  に含まれる領域の  $P(X|r_i)$  の中の上位 2 つの平均値とする. ここでは, 上位 2 つの領域の  $P(X|r_i)$  が高ければ, 全体としてキーワード  $X$  の画像である

可能性が高いということを仮定する. secondmax を二番目に上位の値を選ぶ関数とすると,

$$P(X|I) = \frac{1}{2}(\max P(X|r_i) + \text{secondmax} P(X|r_i)) \quad (8)$$

となる. そして, 最後に  $P(X|I)$  がある一定値以上の画像について, 収集結果として  $P(X|I)$  の順番に表示, もしくはシステム内の画像データベースに記録する.

#### 4. 実験

表 1 に示す 8 種類のキーワード (ライオン, りんご, 赤ちゃん, ノートパソコン, ラーメン, 新幹線, 金閣寺, 大リーグ野茂) について, WWW からの画像収集実験を行った.

表 1 の左半分には, 比較のために従来システムでの結果を示す. 具体的には, 商用テキスト検索エンジンから取得した HTML ファイルの URL 数, それから抽出した画像ファイルの URL 数, 実際に A 画像, B 画像として収集した枚数と, 括弧内はその主観評価による適合率 ((キーワードに適合している画像の枚数)/(収集枚数)), さらに, その合計の枚数と適合率を示す. 収集画像枚数は 220 枚から最大 2596 枚までと, キーワードによってかなりばらつきがあった. 適合率は 47.6% から 68.9% までとおおむね 5 ~ 7 割程度であった. この適合率を上げるのが, 収集フェーズの後の, 画像選択フェーズの役目である. また, A, B 画像それぞれの適合率をみると, A 画像の平均適合率が 7 割, B 画像が 4 割である. 提案手法では, A 画像を最初, 正解画像と見なして, A, B 画像を選択し, その選択された B 画像も用いて, さらに画像選択を繰り返し行う.

表 1 の第 4 カラムに, 従来のカラーヒストグラムを用いた方法による画像選択の結果を示す. 選択枚数, その適合率に加えて, 再現率 (選択画像中の適合画像の枚数)/(収集した画像中の適合画像の枚数) も示してある. 適合率はおおむね収集画像に比べて, 10 ポイント程度上昇している. 但し, これも種類によっては, 例えば「りんご」の場合の様にほとんど変化のないものもある. また, 再現率は 5 割 ~ 7 割程度であった.

表 1 の第 5, 6 カラムに, 提案手法を用いた場合の画像選択の結果を示す. ガウス混合モデルの要素の数  $m_0$ , 正例クラスタの数  $m_1$  をそれぞれ 100, 200 と 5, 10, 20 に設定した場合の 6 通りの実験のうち, 結果が良かった  $m_0 = 200, m_1 = 5$  と  $m_0 = 200, m_1 = 10$  の結果を示す. なお, 負例クラスタの数は  $m_2 = 0.5m_0$  とした. 表中の上段の初期選択結果は, 繰り返しを行う前の最初の選択結果で, つまりラベル付きとみなした A 画像のみを用いて, モデルを学習して, そのモデルによって A, B 画像から選択した結果である. 下段は 5 回繰り返し処理を行った後の結果である.

この 2 通りの平均の結果を比べると  $m_0 = 200, m_1 = 5$  の場合が適合率に関してはよい結果が得られている. 従来のカラーヒストグラムの場合より, 良い結果である. 但し, 再現率はあまり良くなく, 選択枚数は従来方法の半分以下になってしまった. 実際, 正例クラスタの数が多い  $m_1 = 20$  の場合の方が再現率は高かったが, 適合率は低くなってしまい, トレードオフの関係が見られた.

初期選択結果と 5 回繰り返した後の最終選択結果を比較すると, 2 通りの場合とも 5 ポイント程度再現率が改善された. 繰り返し処理によって選択された B 画像も学習画像に用いることで結果が改善されたことが考えられる. ただし「金閣寺」「野茂」は  $m_0 = 200, m_1 = 5$  の場合, 再現率が上がった代わりに 10 ポイント近くも適合率が落ちている. また, 収集枚数に関しては, キーワードの種類によって, 増減に大きなばらつきがある. このことから, 本手法はまだ安定した手法でなく, 改善の余地が大きいと言える.

表 1: 収集結果 . 収集枚数 (適合率 (%)), 従来法および提案法による選択枚数 (適合率, 再現率 (%))

| キーワード   | 収集 URL 数     |  | 収集画像                                   |  | カラーヒストグラム                                   |  | $m_0 = 200, m_1 = 5$                        |  | $m_0 = 200, m_1 = 10$                       |  |
|---------|--------------|--|--|--|---|--|---|--|---|--|
|         | HTML<br>画像   |  | A<br>B<br>選択結果 (A+B)                   |  | A<br>B<br>選択結果 (A+B)                        |  | 初期選択結果 (A+B)<br>最終選択結果 (A+B)                |  | 初期選択結果 (A+B)<br>最終選択結果 (A+B)                |  |
| ライオン 動物 | 2127<br>1737 |  | 126(90) 200(41)<br><b>326 (60.8)</b>   |  | 113(90) 72(43)<br><b>185 (72.3, 67.5)</b>   |  | 120 (64.7, 40.1)<br><b>165 (66.5, 56.8)</b> |  | 169 (63.3, 55.7)<br><b>98 (71.9, 35.9)</b>  |  |
| りんご     | 2247<br>1185 |  | 268(74) 414(65)<br><b>682 (68.9)</b>   |  | 223(73) 140(65)<br><b>363 (70.1, 54.1)</b>  |  | 201 (74.8, 33.6)<br><b>207 (74.9, 34.5)</b> |  | 119 (67.8, 18.1)<br><b>341 (67.1, 50.8)</b> |  |
| 赤ちゃん    | 2312<br>1501 |  | 529(69) 265(46)<br><b>794 (60.5)</b>   |  | 455(72) 89(53)<br><b>544 (67.6, 76.5)</b>   |  | 152 (56.5, 20.3)<br><b>313 (74.7, 46.3)</b> |  | 375 (54.4, 47.0)<br><b>157 (56.9, 20.7)</b> |  |
| ノートパソコン | 2151<br>4236 |  | 479(66) 945(48)<br><b>1424 (55.7)</b>  |  | 340(69) 462(54)<br><b>802 (60.8, 61.4)</b>  |  | 50 (77.1, 6.6)<br><b>192 (79.7, 21.9)</b>   |  | 416 (54.9, 32.6)<br><b>254 (62.0, 22.6)</b> |  |
| ラーメン    | 2225<br>6286 |  | 901(78) 1695(55)<br><b>2596 (67.0)</b> |  | 699(78) 793(59)<br><b>1492 (71.4, 61.2)</b> |  | 96 (76.7, 8.2)<br><b>618 (82.2, 27.6)</b>   |  | 470 (84.6, 28.6)<br><b>462 (88.0, 21.9)</b> |  |
| 新幹線     | 2181<br>5166 |  | 496(68) 1146(40)<br><b>1642 (54.1)</b> |  | 428(70) 604(43)<br><b>1032 (59.3, 68.9)</b> |  | 186 (57.6, 18.5)<br><b>154 (61.5, 11.2)</b> |  | 269 (59.2, 21.3)<br><b>459 (61.0, 32.9)</b> |  |
| 金閣寺     | 2133<br>9949 |  | 497(69) 521(27)<br><b>1018 (47.6)</b>  |  | 403(72) 181(24)<br><b>584 (56.6, 68.2)</b>  |  | 56 (50.0, 7.7)<br><b>278 (41.1, 24.6)</b>   |  | 431 (44.5, 43.0)<br><b>339 (44.4, 32.4)</b> |  |
| 野茂 大リーグ | 2039<br>1435 |  | 96(72) 124(25)<br><b>220 (46.9)</b>    |  | 83(72) 25(32)<br><b>108 (63.4, 66.4)</b>    |  | 55 (60.0, 34.0)<br><b>100 (49.5, 50.5)</b>  |  | 142 (46.0, 66.0)<br><b>90 (52.8, 48.5)</b>  |  |
| 平均値     | 2177<br>3936 |  | 424(73) 663(43)<br><b>1087 (57.7)</b>  |  | 343(74) 295(47)<br><b>638 (65.2, 66.3)</b>  |  | 114 (64.9, 23.5)<br><b>253 (70.0, 32.5)</b> |  | 298 (60.0, 37.6)<br><b>275 (65.0, 32.0)</b> |  |

## 5. おわりに

画像収集システムの画像選択において, ガウス混合分布 (G-MM) による生成モデルを用いた確率的画像選択方法を提案した. HTML 解析によって評価の高い A 画像を正例, 中程度に高い B 画像を未知画像, それとは別にランダムに収集した画像を負例として, 準教師学習の方法によって生成モデルを学習し, それを用いて A 画像, B 画像に対してキーワード  $X$  を表す確率  $P(X|I)$  を求め, 画像選択を実現した. その結果, 選択枚数は半分以下に減少したものの, 適合率は平均で 5 ポイント改善された. また, 繰り返し処理を行うことで, 初期結果よりも 5 ポイント程度適合率が改善された.

今後の課題としては, パラメータの組合わせの最適化, 利用する画像特徴量の改良, 現在選択処理のみで 5 分程度掛かっている時間の短縮などが挙げられる. 本手法では, 処理の途中で各キーワードに対応する生成モデルを構築しているので, それを利用して未知画像に対する画像分類実験を行うことを検討している.

## 参考文献

- [Cai 04] Cai, D., He, X., Li, Z., Ma, W., and J.R., W.: Hierarchical clustering of WWW image search results using visual textual and link information, in *Proc. of ACM International Conference Multimedia 2004*, pp. 660–667 (2004).
- [Deng 01] Deng, Y. and Manjunath, B. S.: 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).
- [Feng 04] Feng, H., Shi, R., and Chua, T.: A bootstrapping framework for annotating and retrieving WWW images, in *Proc. of ACM International Conference Multimedia 2004*, pp. 960–967 (2004).
- [Framkel 96] Framkel, C., Swain, M. J., and Athitsos, V.: Web-See: An Image Search Engine for the World Wide Web, Technical Report TR-96-14, University of Chicago (1996).
- [Lempel 01] Lempel, R. and Soffer, A.: PicASHOW: Pictorial Authority Search by Hyperlinks on the Web, in *Proc. of the Tenth International World Wide Web Conference*, pp. 438–448 (2001).
- [Nigam 00] Nigam, K., McCallum, A., Thrun, S., and Mitchell, T.: Text Classification from Labeled and Unlabeled Documents using EM, *Machine Learning*, Vol. 39, No. 2/3, pp. 103–134 (2000).
- [Rowe 98] Rowe, N. and Frew, B.: Automatic caption localization for photographs on World-Wide Web pages, *Information Processing and Management*, Vol. 34, No. 1, pp. 95–107 (1998).
- [Sclaroff 99] Sclaroff, S., LaCascia, M., Sethi, S., and Taycher, L.: Unifying Textual and Visual Cues for Content-Based Image Retrieval on the World Wide Web, *Computer Vision and Image Understanding*, Vol. 75, No. 1/2, pp. 86–98 (1999).
- [Smith 97] Smith, J. R. and Chang, S. F.: Visually Searching the Web for Content, *IEEE Multimedia*, Vol. 4, No. 3, pp. 12–20 (1997).
- [Song 04] Song, X., Lin, C., and Sun, M.: Autonomous visual model building based on image crawling through internet search engines, in *Proc. of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pp. 315–322 (2004).
- [Yanai 03] Yanai, K.: Generic Image Classification Using Visual Knowledge on the Web, in *Proc. of ACM International Conference Multimedia 2003*, pp. 67–76 (2003).
- [Yanai 04] Yanai, K., Shindo, M., and Noshita, K.: A Fast Image-Gathering System on WWW Using a PC Cluster, *Image and Vision Computing*, Vol. 21, No. 1, pp. 59–71 (2004).
- [柳井 01] 柳井啓司: キーワードと画像特徴を利用した WWW からの画像収集システム, 情報処理学会論文誌:データベース, Vol. 42, No. SIG10 (TOD11), pp. 79–91 (2001).
- [獅々堀 04] 獅々堀, 小泉, 柘植, 北: 画像知識データベースを用いた WWW 画像検索システムの開発, 電子情報通信学会論文誌 D-I, Vol. J87-D-I, No. 2, pp. 154–163 (2004).
- [柳井 04] 柳井啓司: 一般画像自動分類の実現に向けた World Wide Web からの画像知識の獲得, 人工知能学会誌, Vol. 19, No. 5, pp. 429–439 (2004).