

確率モデルを用いた Web 画像マイニングによる画像認識

Probabilistic Image Recognition Employing Web Image Mining

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

Current approaches to image classification require training images prepared by hand. In this paper, we describe experiments on image classification using images gathered from the Web automatically as training images. To gather images from the Web, we use the probabilistic method we proposed before. In the method, we build a generative model which is based on the Gaussian mixture model (GMM) from imperfect training images gathered from the Web in order to distinguish relevant images from irrelevant ones. In this paper, we propose applying the model built during Web image gathering process to generic image classification task.

1. はじめに

近年のデジタルカメラの普及や WWW(World Wide Web) 上の画像の爆発的な増大によって、デジタル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)」に対する要求が高まっている。そのため、近年、一般物体認識の研究が盛んに行われるようになってきており、画像データベース中の画像に対して自動的にキーワードを付けたり [森 01, Barnard 03], 意味的に類似している画像を検索する研究 [Wang 01, Li 03], 一般的な手法によって多くの種類の対象を検出する研究 [Fei-Fei 04] などが行われている。これらの研究では、事前に単語と画像の対応付けを学習する必要があるため、学習データとして予めキーワードが付与されている画像が必要である。そのため、現在、行われている実世界画像に対する一般画像認識 (Generic Object Recognition) の研究では、対象とする物体の種類が多くても 101 種類に留まっている。例えば、マルチメディア系のビデオ映像解析の国際的ベンチマークコンテストの TRECVID の High-level feature extraction task (対象検出) では、与えられる学習データの物体の種類が 39 種類、コンピュータビジョン系の物体認識コンテスト PASCAL Challenge では 10 種類の物体のみが認識対象になっている。

一般的な物体認識を実現する場合には、実世界における多くの種類の物体やシーンを認識できるようにする必要がある。そのためには、人間と同じようにとても多くの「物体」に対する知識が必要である。つまり「山」「自動車」「ライオン」「机」など世の中のあらゆる対象についての画像中での見え方に関する知識が必要で、それらを収集しデータベース化することは「一般画像認識」の実現のためには不可欠である。

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 [Fergus 05] 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り、我々はそのために確率モデルを用いた Web 画像収集システム Probabilistic Image Collector [柳井 05, Yanai 05] を提案している。このシステムでは、テキスト検索エンジンを用いて、キーワードに関係しそうな Web ページを大量に収集し、テキスト解析によって可能性の高い画像を推定し Web から収集する。そして、テキスト解析によって特に可能性が高いと判定された画像を初期学習画像として、画像の領域特徴量に関する確率モデルを構築し、それを用いて画像を選別し、最終的に残った画像をキーワードに対応する画像として出力する。前半はテキスト処理、後半は画像処理となっている。確率モデルの構築自体は supervised であるが、処理全体としては unsupervised になっていることが特徴である。画像選択のための確率モデルは、未知の画像がキーワードに対応する画像であるかどうかの推定するためのモデルで、画像認識のための確率モデルとなっている。そこで、本研究では、確率的 Web 画像収集の処理の中で構築される確率モデルを利用して、一般画像分類を試みる。

2. 確率的 Web 画像収集の方法の概要

まずは、確率モデルを用いた Web 画像収集システム Probabilistic Image Collector [柳井 05, Yanai 05] を簡単に説明する。

最初に与えられたキーワードに関係する 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。

その後に画像特徴量に基づく、画像選択を行う。我々は EM アルゴリズムを用いた確率的画像選択法を提案し用いた。初めに A にランクされた画像と、予め用意してあるランダムに収集されたキーワードと無関係な画像を領域分割し、ある一定以上の大きさのすべての領域から画像特徴ベクトルを抽出し、EM アルゴリズムによってクラスタリングを行う。そして、

A: 柳井 啓司 電気通信大学情報工学科 〒182-8585 東京都調布市調布ヶ丘 1-5-1 E-mail: yanai@cs.uec.ac.jp

キーワードに関係する画像（正解画像）に含まれる領域のクラスタをAにランクされた画像を用いて推定し、それに基づいて領域の特徴ベクトルの生成モデルを構築する。さらにそのモデルに基づいて、A, B各画像が正解画像である確率を求め、その結果を用いて、正解領域のクラスタの推定を再度繰り返し行う。予め決めた回数を繰り返したら終了とし、確率の高い順に収集結果を出力する。なお、画像特徴量は色、テクスチャ、形状の3通りのものを組み合わせて用いた。詳しくは[柳井 05, Yanai 05]に説明がある。

3. 画像分類実験の方法

Web画像収集での画像選択時に構築した領域の特徴量 r_j に関するキーワード w_i の確率モデル $P(W = w_i | r_j)$ (ただし $W \in \{w_i, \neg w_i\}$) を、未知画像に適用することによって、未知画像 I がキーワード w_i に関係する確率 $P(W = w_i | I)$ を求めることができる。

そこで、予め n 個の分類対象のキーワード w_1, \dots, w_n に対する確率モデルをそれぞれ独立に構築しておいて、未知画像 I がそれぞれのキーワードに対応する確率 $P(W = w_i | I)$ を計算する。そして、確率が最も大きい分類クラスにその未知画像を分類することとする。

また同様の方法で、分類対象のキーワードに対する確率モデルをまとめて一つ、つまり確率モデル $P(W' | r_i)$ ($W' \in \{w_1, w_2, \dots, w_n\}$) を求める方法も考えられる。実際には、こちらの方が自然であるが、今回の予稿には実験が間に合わなかったため、実験結果は載せていない。これは今後の課題とする。

4. 実験

表1に示す6種類のキーワード（りんご、海岸、花、ライオン、日暮れ、滝）について、分類実験を行った。学習データは、(1) Webから収集した画像そのまま、つまり、誤りも含んだ学習画像データセット (*raw*) と、(2) 人手で正解画像のみを選んで作り出した正解画像のみの学習画像データセット (*all_ok*) の2通りを用意した。一方、評価データは、(1) Corel画像データ集から各キーワード100枚ずつ選び出して作ったデータセット (*Corel*) と、(2) 商用の画像検索エンジンを用いてWebから人手で各キーワードにつき100枚ずつ、学習用画像と同じものを含まないように注意して集めた評価用データセット (*Web*) の2通りを用意した。学習データセット2通り、評価データセット2通りを組合せて実験を実施し、合計4通りの分類実験を行った。分類実験の評価は、再現率 ((そのクラスに正しく分類された画像の枚数)/(テストに用いたそのクラスの画像の枚数))、適合率 ((そのクラスに正しく分類された画像の枚数) / (そのクラスに分類された画像の全枚数))、およびそれら2つの値の調和平均であるF値 (F-measure) を用いた。

表1に4通りの実験の結果を示す。学習データが *raw* で、評価データが Corel 画像の実験1では、分類結果の平均のF値が36.7であった。一方、学習データが正解のみの場合の実験2の結果は平均F値は48.9で、実験1に比べて大幅に結果が良くなっている。これは、実験1の学習データにノイズが含まれていたためである。「滝」を除いて個々の結果も改善されている。

一方、実験3は、学習画像は *raw* で、評価画像に Web 画像を用いた。結果のF値は43.0で、実験1よりも6.3向上している。これは、Webから収集した画像で学習したために、Corel画像よりもWeb画像の方が、学習画像と評価画像の画像の傾

向が似ていたためである。学習画像に *all_ok* を用いた実験4では、50.0のF値が得られた。これも同様に実験2よりもよい値になっている。

表 1: 実験1~4の結果

クラス	実験1 (学習:raw, 評価:Corel)			実験2 (学習:all_ok, 評価:Corel)		
	再現率	適合率	F値	再現率	適合率	F値
apple	36.4	5.7	9.9	27.1	18.6	22.0
beach	29.8	25.8	27.6	60.9	72.2	66.0
flower	39.5	18.1	24.8	38.3	19.1	25.5
lion	55.1	27.3	36.5	100.0	67.7	80.7
sunset	34.1	79.4	47.7	42.3	84.5	56.4
waterfall	42.2	49.5	45.6	36.2	21.2	26.7
平均値	39.5	34.3	36.7	50.8	47.2	48.9

クラス	実験3 (学習:raw, 評価:Web)			実験4 (学習:all_ok, 評価:Web)		
	再現率	適合率	F値	再現率	適合率	F値
apple	50.0	2.0	3.8	31.7	37.3	34.2
beach	33.3	60.8	43.1	41.5	52.9	46.5
flower	31.9	29.4	30.6	35.0	41.2	37.8
lion	61.8	41.2	49.4	100.0	72.5	84.1
sunset	51.3	76.5	61.4	54.7	56.9	55.8
waterfall	39.2	39.2	39.2	50.0	27.5	35.4
平均値	44.6	41.5	43.0	52.2	48.0	50.0

5. おわりに

Web画像収集システムの画像選択のために以前に提案したガウス混合分布 (GMM) による生成モデルを用いた確率的画像選択方法で、処理途中に構築した確率モデルを利用して、一般画像分類を試みた。学習画像が *raw* の場合、Corel画像に対しては36.7、Web画像に対しては43.0のF値が得られた。学習画像が *all_ok* の場合、Corel画像に対しては48.9、Web画像に対しては50.0のF値が得られた。すべて正解画像の学習セットよりも *raw* を学習セットとした方が得られたF値は低いですが、2割程度の低下に留まっており、今度さらに改良を加え、より多くの種類を分類実験することによって、Web画像を学習画像とすることで十分実用的な性能を期待できると考えられる。

今後の課題としては、分類精度の向上と、分類クラス数の数百、数千クラスへの大規模化が挙げられる。特に分類精度の向上のためには、分類対象のキーワードに対する確率モデルをまとめて一つ、つまり確率モデル $P(W' | r_i)$ ($W' \in \{w_1, w_2, \dots, w_n\}$) を求める方法を導入することが挙げられる。これは、translation model[Barnard 03] とほぼ等価であり、再度モデル構築をする必要があるものの、こちらの方がいい結果が得られることが期待される。また、画像特徴量に関して、領域の色、テクスチャ、形状の特徴以外に、[Fei-Fei 04, Fergus 05] でよい成果が得られている特徴点オペレータに基づく局所特徴量表現の利用も有望である。

参考文献

- [Barnard 03] Barnard, K., Duygulu, P., de Freitas, N., Forsyth, D., Blei, D., and Jordan, M.: Matching Words and Pictures, *Journal of Machine Learning Research*, Vol. 3, pp. 1107-1135 (2003).
- [Fei-Fei 04] Fei-Fei, L., Fergus, R., and Perona, P.: Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories, in *Proc. of IEEE CVPR Workshop of Generative Model Based Vision* (2004).
- [Fergus 05] Fergus, R., Fei-Fei, L., Perona, P., and Zisserman, A.: Learning Object Categories from Google's Image

- Search, in *Proc. of IEEE International Conference on Computer Vision* (2005).
- [Li 03] Li, J. and Wang, J. Z.: Automatic linguistic indexing of pictures by a statistical modeling approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 9, pp. 1–14 (2003).
- [Wang 01] Wang, J. Z., Li, J., and Wiederhold, G.: SIMPLiCity: Semantics-Sensitive Integrated Matching for Picture Libraries, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 9, pp. 947–963 (2001).
- [Yanai 03] Yanai, K.: Generic Image Classification Using Visual Knowledge on the Web, in *Proc. of ACM International Conference on Multimedia 2003*, pp. 67–76 (2003).
- [Yanai 05] Yanai, K. and Barnard, K.: Probabilistic Web Image Gathering, in *Proc. of 7th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pp. 57–64 (2005).
- [森 01] 森, 高橋, 岡: 単語群つき画像の分割クラスタリングによる未知画像からの関連単語推定, *電子情報通信学会論文誌 D-II*, Vol. J84–D–II, No. 4, pp. 649–658 (2001).
- [柳井 04] 柳井啓司: 一般画像自動分類の実現へ向けた 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).
- [柳井 05] 柳井啓司: 実世界画像コーパス作成のための高精度 Web 画像収集, *画像の認識・理解シンポジウム (MIRU 2005)* (2005).