

Web 画像マイニング: Web からの画像知識の獲得 と その応用

柳井 啓司 (電気通信大学)

Web Image Mining : Acquisition of Visual Knowledge from the Web and its Application

Keiji Yanai (The University of Electro-Communications)

1 はじめに

今日, World Wide Web 上には大量の画像とそれに付随する HTML テキストが存在している. 我々はそうした Web 上に大量に存在するテキスト情報付き画像データを収集, 分析, 利用する「Web 画像マイニング」を提案し [1, 2], 研究を行っている. 本稿では, 我々の「Web 画像マイニング」に対する取り組みについて述べる.

2 Web 画像マイニングとは?

人工知能研究においては, システムに大量の知識を手で与えることに限界があるという知識獲得のボトルネックの問題を解決するために, 近年, Web からの知識獲得 (Web マイニング) の研究が盛んに行われている. Web 画像マイニングもそれと同様に, 実世界画像とその意味内容との対応の知識 (ここでは画像知識と呼ぶ) を Web から自動的に獲得し, その知識を実世界画像分類や自動キーワード付与などに応用することを目的とする. ただし, 従来からの Web からの知識獲得の研究では主にテキストデータのみを対象にしていたのに対して, Web 画像マイニングでは画像データを主な対象として, それに付随するテキスト情報も扱うという違いがあり, テキスト処理だけでなく画像処理も含まれることが Web 画像マイニングの大きな特徴である.

Web 上の画像は, 従来の画像データベース中の画像とは違って, 単独で独立に存在していることは通常なく, その画像を参照している HTML テキストと共に存在し, Web ページを表現する情報の一部となっている. Web ページは, 様々な人が様々な目的のために構築するものであり, 多様な画像を含んでいる. 中には, テキストに有用な情報が含まれていなかったり, まったく不正確な情報が含まれている場合もあるが, 統計的に見ると画像の内容と関連の深い内容のテキストを含んでいることが多い. 例えば, 「動物のライオン」の画像を含んでいるページは, 「ライオン」という単語が含まれていることが多く, 含まれていない場合でも動物に関連するテキスト情報が含まれていることが多い. 一般には, HTML テキスト中には, 画像へのリンクタグ (IMG SRC もしくは A HREF) の周辺に関連する画像に関連する情報が多く含まれているので, タグの前後のテキストを画像の「周辺テキスト」と呼び, 画像内容を推定するためのテキスト情報として使われる.

Web 画像マイニングを実際に行うためには, Web からの画像知識の収集システムと, 収集した画像知識を利用する応用システムの 2 つが必要である. Web からの画

像知識の収集のためには, その起点として Web 検索エンジンを用いるのが一般的である. 特に, Google Image Search や Yahoo Image Search などの商用 Web 画像検索エンジンを利用すると手軽に画像収集が可能となる. 一般には, まず初期処理としてこれらの Web 画像 (もしくはテキスト) 検索を利用して, 知識として収集する候補の画像とそれを参照している HTML テキストと共に収集し, 画像の周辺テキストや画像特徴量を解析することによってそれらの候補の中から適切な画像を絞り込む.

なお, Web 画像マイニングにおける画像知識収集では, 画像知識抽出のために大規模 (100 枚 ~ 100 万枚以上) な自動画像収集を行う必要があり, 少量の画像を検索する Web 画像「検索」とは異なるという意味で, Web 画像「収集」という用語を意図的に用いている.

周辺テキストに対するテキスト解析を併用することによって, 従来の画像データベース中の画像の利用では考えられなかった様々な応用が, Web から収集した画像知識の応用例として期待できる. 例えば,

- (1) 画像分類・認識の学習データとして Web 画像を利用.
- (2) ある特定の画像の大量収集と画像特徴分析.
- (3) 単語概念の視覚性の分析.

などである. 元々, 我々は (1) のために Web 画像の知識としての利用を始めたが, 最近では他にもいろいろな可能性があることが分かってきた. 以下では, これら 3 つの応用例について簡単に説明する.

3 一般画像認識への適用

画像の意味内容を計算機に自動的に認識させる「一般物体認識 (generic object recognition)」に対する要求は高いものの, 辞書に出ているほどの多くの物体に関する視覚的知識を持っている人間には遠く及ばないのが現状である. そこで, 我々は一般物体認識のための知識を, 多種多様な画像が存在する World Wide Web から収集することを提案した [1, 2]. これが我々の Web 画像マイニング研究の始まりである.

図 1 に提案したシステムを示す. [1, 2] では, Web 画像収集システムと画像分類システムを組み合わせることによって, Web 画像マイニングによる一般画像の自動分類を試みた. 従来からそれぞれのシステムは存在していたが, これらを組み合わせることは新しい試みであった. 大量性, 多様性という 2 つの大きな特徴を持つ Web 画像を画像分類の学習画像, つまり画像知識として用いることによって, 従来のシステムでは難しかった一般の多種多様な画像を自動分類を実現することを目的とした. ただし, 実験で

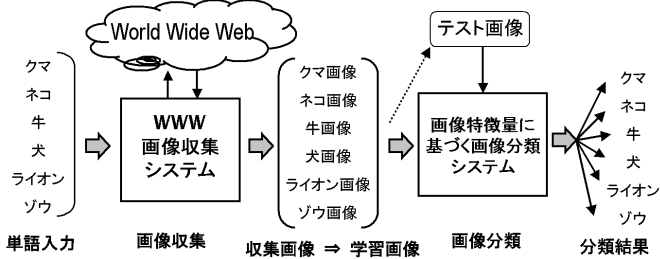


図 1. Web 画像マイニングによる一般画像分類システム .



図 2. ラーメン画像. ラベル付き領域が“ラーメンらしい”部分である .

は 20 種類の画像分類で分類率が 35%前後で、高い分類精度とはいえない結果であった .

これを改善するために、現在我々は (1)Web から精度よく画像を多くの画像を収集する手法、(2) 精度が 100%でないノイズを含む学習データからの学習方法、の両方を研究中である . (1) の方については確率モデルを用いた新しい手法を 2005 年に提案した [3, 4] .

他にも、Web 上の画像を用いた画像認識の研究は、Google Image Search の出力結果からの画像確率モデルの学習 [5]、Web 画像アルバムサイトの画像と注釈情報を用いた一般画像へのアノテーション [6] など、近年行われるようになっており、今後同種の研究は増加していくことが予想される .

4 特定画像の大量収集と画像特徴分析

[3, 4] において、我々は画像選択のために収集対象画像の種類毎に確率モデルの構築を行った . この確率モデルは、収集した画像のうち周辺テキスト解析によって正しい可能性が極めて高いと判定された画像を基にして自動的に学習したものであり、その対象らしさを表す確率モデルであるといえる . そこで、[3, 4] の実験で用いられた「ラーメン」について考えると、確率モデルを用いることで、領域分割された画像中の各領域について「ラーメンらしさ」を求めることが可能となる .

図 2 に「ラーメンらしい」領域にラベルのついたラーメン画像を示す . ここで示した 3 枚の画像にはいずれもどんぶりの縁の部分とチャーシュー部分にラベルがついており、ラーメンにとって重要な部分は「どんぶり」と「チャーシュー」であるということが Web 画像マイニングによって明らかになったといえる .

他の研究では、UC Berkeley の T.L.Berg らによる Web ニュース画像からの有名人画像の抜き出し [7] や、Web からの動物画像の収集 [8] など特定種類画像の Web 画像マイニングの一種であるといえる .

5 単語概念の視覚性の分析

一般物体認識の目的の一つは、限定のない一般的な画像にその画像が表す内容に対応する単語を付与すること

であるが、その際に用いる単語を選ぶ方法は従来は人手によるアドホックなもので、客観的に単語を選ぶ方法は存在しなかった . 例えば、物体を表す名詞の中には「ライオン」「山」というように視覚的特徴に直接対応しているものと「動物」「乗り物」の様に様々な動物、乗り物が存在するために視覚的特徴と単語概念を直接結び付けることが困難であるものが存在し、画像に認識によって直接付与する単語は前者の様な視覚的特徴と結び付きの強いものであることが望まれる .

そこで我々は [9, 10] において、単語が対応する「概念」がどの程度、「視覚性 (visualness)」を持ち合わせているかを定量的に測定する尺度「画像領域エントロピー」を提案し、一般物体認識において画像に付与する単語として適したものを自動的に選び出すことを行った . 画像領域エントロピーを計算する際には、確率的領域選択手法によって各単語に関連が深い画像領域を推定し、背景などの単語の無関係な画像の領域を影響を取り除く . 実験では、150 語の形容詞について対応する画像を各 300 枚程度 Web より収集し、それらについて画像領域エントロピーを求め比較を行った . その結果、dark や yellow など画像特徴に直接関係する色、明るさを表す単語のエントロピーが小さくなり、一方 famous, religious, artificial など画像特徴に関係がなさそうな単語のエントロピーが大きくなるのが分かった .

6 おわりに

本稿では、Web 画像マイニングの解説を行い、応用例を 3 つ紹介した .

今後は、Web 画像マイニングによる一般画像認識の実現のための高精度な大規模画像コーパスの構築を行うと共に、静止画だけでなく動画画像も含めた Web 上の画像知識の新たな利用方法について模索していくつもりである .

参考文献

- [1] Yanai, K.: Generic Image Classification Using Visual Knowledge on the Web, *Proc. of ACM International Conference on Multimedia 2003*, pp. 67–76 (2003).
- [2] 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).
- [3] Yanai, K. and Barnard, K.: Probabilistic Web Image Gathering, *Proc. of 7th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pp. 57–64 (2005).
- [4] 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).
- [5] Fergus, R., Fei-Fei, L., Perona, P. and Zisserman, A.: Learning Object Categories from Google's Image Search, *Proc. of IEEE International Conference on Computer Vision* (2005).
- [6] Wang, X.-J., Zhang, L., Jing, F. and Ma, W.-Y.: AnnoSearch: Image Auto-Annotation by Search, *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 1483–1490 (2006).
- [7] Berg, T. L., Berg, A. C., Edwards, J., Maire, M., White, R., Teh, Y.-W., Learned-Miller, E. and Forsyth, D. A.: Names and Faces in the News, *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 848–854 (2004).
- [8] Berg, T. L. and Forsyth, D. A.: Animal on the Web, *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 1463–1470 (2006).
- [9] Yanai, K. and Barnard, K.: Image Region Entropy: A Measure of “Visualness” of Web Images Associated with One Concept, *Proc. of ACM International Conference on Multimedia 2005*, pp. 420–423 (2005).
- [10] 柳井啓司, Barnard, K.: 一般物体認識のための単語概念の視覚性の分析, *情報処理学会コンピュータビジョン・イメージメディア研究会報告 CVIM2005-152-1*, pp. 1–8 (2006).