

Web ページからの説明付き画像の選択

芝野 博誠 黄瀬 浩一 松本 啓之亮

大阪府立大学大学院工学研究科情報工学分野

e-mail : shibano@ss.cs.osakafu-u.ac.jp, {kise, matsu}@cs.osakafu-u.ac.jp

1 はじめに

近年インターネットの普及により、テキストや画像など数多くのメディアが存在するようになった。これらのメディアから必要な情報を検索する手法として、複数のメディアを扱うクロスメディア検索が注目されている。

クロスメディア検索の代表例としては、キーワードによる Web 上の画像検索が挙げられる。このような画像検索では、ユーザはシステムに対し検索質問をテキストで提示し、システムはユーザに対し検索結果として画像を提示する。しかしテキストと画像のデータの形式が異なるため、これらを直接比較することはできない。そこで何らかの方法により、画像と周囲のテキストとの対応関係を定義づける必要がある。

画像と周囲のテキストを対応付けるためには、まずテキストと対応付けられる画像(説明付き画像)を選択しなければならない。方法としては人手による方法と機械学習による方法が考えられるが、人手による方法は対象画像が増加するにつれ膨大な労力が必要となり、事実上困難である。

よって本稿では、機械学習器として近年注目を集めている Support Vector Machine(SVM)[1]を用いて、識別器を構成し、説明付き画像の選択する手法を提案する。また Web ページ 100 枚を対象とした実験結果の報告と考察を行う。

2 説明付き画像の選択

2.1 画像選択の問題点

クロスメディア検索とは、テキストと画像などのように種類の異なるメディアを横断的に検索する技術である。特にテキストと画像の組み合わせに関する研究は、ここ数年テキスト、画像ともに数が飛躍的に増加していることやそれぞれの検索技術が向上していることなどを背景に盛んに行われている。そこで本研究では、テキストと画像を対象とする。

従来のテキストと画像を対象としたクロスメディア検索では、対象となる画像を選択する際、人手により選択したり、選択基準を人があらかじめ設定し、その基準に適合しない画像を対象から外すなどの方法が行われていた。

しかしこれらの手法にはそれぞれ問題点がある。人手による選択は、Web 上の対象画像が膨大なため、人手で全てを調べることがほぼ不可能である。

選択基準を設定する方法には、基準を画像のサイズなどの画像特徴量から設定する方法 [2] と画像の存在するページのテキスト情報から設定する方法 [3] がある。しかしこれらの方法では、画像特徴量またはテキスト情報の一方しか考慮しておらず、対象画像の選択方法としては改善の余地がある。

本研究ではこの点に着目し、テキスト情報と画像情報の両方を考慮した説明付き画像の選択手法を提案する。

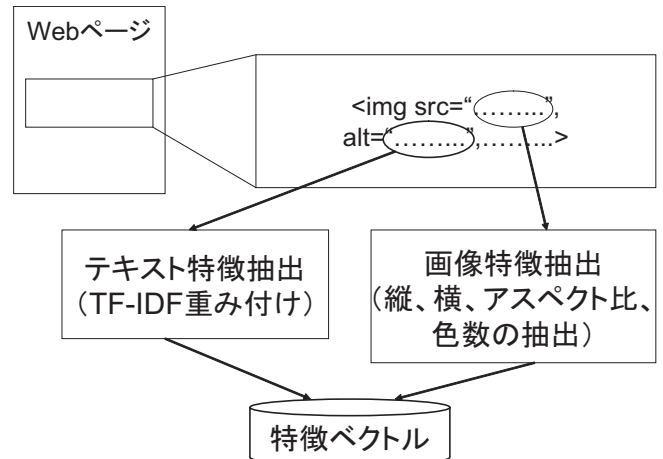


図 1: 特徴抽出

2.2 提案手法の流れ

以下では、英文のページを対象として説明付き画像を選択する手法について説明する。処理は大きく分けて特徴抽出と学習・識別がある。

図 1 に特徴抽出の処理の流れを示す。

まず、画像特徴抽出について述べる。抽出する特徴は画像の縦と横の長さ、両者の比率であるアスペクト比、画像に含まれる色の数である。この特徴を画像ごとに集め、画像の各要素ごとに正規化を行う。その結果を画像の特徴ベクトル v_I とする。

次に、テキスト特徴抽出について述べる。テキスト特徴は各画像の alt 属性のテキストを調べて抽出する。まず alt を含めた Web ページ全体から、冠詞や前置詞などテキスト特徴として適当ではないと考えられる不要語を除去する。その後、以下の方法で特徴ベクトル v_T を抽出する。

$$v_T = (v_1, \dots, v_m), v_i = \frac{l_i g_i}{p} \cdot a_i$$

$$l_i = \log(1 + f_i), g_i = \log \frac{n}{n_i}, p = \sqrt{\sum_{i=1}^m (l_i g_i)^2}$$

$$a_i = \begin{cases} 1 & \text{単語 } w_i \text{ が対象画像の alt に存在するとき} \\ 0 & \text{それ以外} \end{cases}$$

ここで f_i は現在注目している Web ページにおける単語 w_i の出現頻度、 n は Web ページの総数、 n_i は単語 w_i を含む Web ページ数、 m は Web ページ全体の単語の総数である。なお alt にテキストが存在しない場合は、 v_T は零ベクトルとなる。

最後に得られたベクトル v_I, v_T を連結し、特徴ベクトル $v = (v_I, v_T)$ を得る。

表 1: 実験結果

	従来法	提案手法
再現率	96.8%	83.7%
精度	52.2%	68.2%
F 値	67.8%	75.1%

表 2: 実験結果の分布

		提案手法	
		正	誤
従来法	正	261	17
	誤	63	52

得られた特徴ベクトル v に、説明付き画像である場合を 1、それ以外を -1 とするラベル y をつけ、学習データ (v, y) を作成する。このデータに SVM を用いて識別器を構成し、未知の入力 v に対するラベル y を推定して選択を行う。

3 実験

提案手法の有効性を検討するため比較実験を行った。まず、Google.com で検索対象を英語のみに限定して “river” というキーワードで画像検索を行い、そのうち上位 100 件の Web ページをダウンロードして、実験の対象ページとした。また対象画像は対象ページ中に含まれる JPEG 形式の画像とした。この方法により、説明付き画像 125 枚、説明の付かない画像 268 枚の計 393 枚を得た。

比較対象としては、文献 [2] の手法を従来法として実装した。従来法は、画像サイズが 55×55 以上でアスペクト比が 3.8 倍以下 [2] のものを説明付き画像とするものである。

提案手法は次のように用いた。まず、正解データを画像検索の検索順位の奇数と偶数で二つに分けた。このデータに対し、一方のデータから提案手法による画像識別器を構成し、もう一方のデータで構成された画像識別器のテストを行う 2 分割交差検定を行った。SVM としては TinySVM[4] を用いた。カーネル関数は 5 次の多項式カーネルである。

評価尺度としては再現率、精度、 F 値を用いた。 F 値とは再現率 $R = |C|/|A|$ と精度 $P = |C|/|B|$ の調和平均で表される尺度であり、 $F = 2RP/(R + P)$ により求められる。ここで、 $|A|$ は正解の画像数、 $|B|$ は検索された画像数、 $|C|$ は検索された画像中の正解の画像数である。

結果を表 1 に示す。従来法は、再現率が非常に高いのに対して精度があまり高くないので、 F 値があまり高くないことがわかる。このことは画像のしぼりこみが十分には行われていないことを示している。一方、提案手法は、従来法に対して精度と F 値が向上していることがわかる。これは再現率は低下しているものの精度が向上したためである。

次に提案手法によって結果が改善された例と逆に誤って識別した例を具体的に挙げ、その識別器の誤りの傾向を示す。

図 2 は結果が改善された画像の例である。この画像のサイズは 49×70 (pixels) であるため、従来法では説明付き画像とはならない。しかしこの画像は使われている色数が多く、alt のテキストも本文中に十分出現していたため、提案手法では説明付き画像として識別されている。

反対に、図 3 は誤って識別された例である。この画像のサイズは 120×66 (pixels) であるため、従来法では説明付き画像となる。しかし使われている色数が少なく、alt のテキス

alt= “Sloop”
出現回数 3 回

図 2: 結果が改善された例

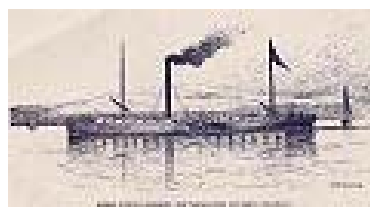
alt= “Clermont”
出現回数 0 回

図 3: 誤って識別された例

トが本文中のテキストと全く一致していなかったため、提案手法では誤って識別されている。

表 2 に、2 手法での判断の一致数を示す。総合的には精度が向上していることから分かるように従来法が判断を誤ったものを提案手法が正しく判定する例が多く見られた。これは、従来法ではサイズによって説明付き画像であると誤って判断された画像が、提案手法では正しく判断されたためと考えられる。また両手法とも細長い画像やサイズの大きい画像を中心に識別を誤っている傾向が見られた。これは、縦横とアスペクト比が他の要素に比べて強い影響を与えているためと考えられる。

4 おわりに

本稿では、テキストと画像の情報を取り出し、SVM を用いることで説明付き画像を自動的に分類する手法について提案した。本手法の特徴は、画像特徴とテキスト特徴を同時に用いて学習器を構成している点である。今後の課題としては、実際の画像とテキストとの対応付けへの応用や使用するデータや属性を増やすことでより精度を向上させることが挙げられる。

謝辞 本研究は日本学術振興会科学研究費補助金基盤研究 (C) (No.14580453) の補助による。

参考文献

- [1] 前田英作, “痛快! サポートベクトルマシン”, 情報処理学会誌, Vol.42. No.7, pp.676-683(2001).
- [2] 嶋田和孝, 伊藤哲郎, 遠藤勉, “表層的な情報を用いた画像の内容特定と分類”, 第 139 回自然言語処理研究会, 2000-NL-139-8(2000).
- [3] 柳井啓司, “キーワードと画像特徴量を利用した 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.SIG10-008, pp.79-91(2001).
- [4] URL : <http://cl.aist-nara.ac.jp/~taku-ku/software/TinySVM/>