

HTML テキストの重要文を用いた画像ラベリング手法

Using Key Sentences of HTML Text to Indexing Method for Images

相良直樹*1 砂山渡*1 谷内田正彦*1
Naoki Sagara Wataru Sunayama Masahiko Yachida

*1大阪大学大学院基礎工学研究科
Graduate School of Engineering Science, Osaka University

In recent years, the use population of the Internet is continuing increasing on a global scale with rapid growth of the personal computer and WWW(World Wide Web). Additionally, transmission of the multimedia information through the WWW get active, and the interest to image retrieval has been increasing greatly. Consequently, this paper provides the indexing method of images collected on the WWW, using key sentences of HTML text. The indexing method is based on relationship between key sentence relevant to image and image. That is, the method uses that key sentence relevant to the image contains the label of image.

1. はじめに

Web 上のテキストを検索する技術は近年目覚しく発展してきた。しかし、Web 上のコンテンツはテキストデータだけではない。テキスト検索で探せない Web コンテンツとして画像があり、これを探る画像検索手法に画像ラベルによる検索手法がある。これまで画像へのラベル付けの手法として、画像からの距離に基づく手法、単純頻度に基づく手法、HTML メタデータを用いた手法 [Munson 01] などが提案されてきたが、どれも完全ではなく程度の差はあるが精度に問題を抱えている。そこで本研究ではこの精度を改善すべく HTML テキストの重要文を用いた画像へのラベリング手法を提案する。

2. Web ページ中の画像とテキストの関係

Web ページ中の画像はそのページ中において重要な役割を担っている。したがって、その画像に関するテキスト領域中の重要文と画像は大きな関わりを持っている。そのため、画像に関するテキスト領域中の重要文中、またはその周辺に画像を表すキーワードが存在している。よって、この重要文に基づく評価手法は画像へのラベル付けを行うにあたって有効であると考えられる。

3. 本システムの処理の流れ

本研究で提案するシステムは図 1 のように 5 つのモジュールから構成されている。以下に各モジュールについて述べる。

3.1 領域抽出モジュール

領域抽出モジュールでは入力として Web ページを受け取り、そのページ中に存在する各画像に関するテキスト領域の抽出を意味的な区切りを表す P タグ、DIV タグ、TABLE タグなどの HTML テキストのタグ情報を用いて行う。

3.2 画像からの距離に基づく評価モジュール

画像からの距離に基づく評価モジュールでは入力として各画像に関するテキスト領域を受け取り、その領域中に存在するすべての名詞 w に対して HTML テキスト領域中での IMG タグ

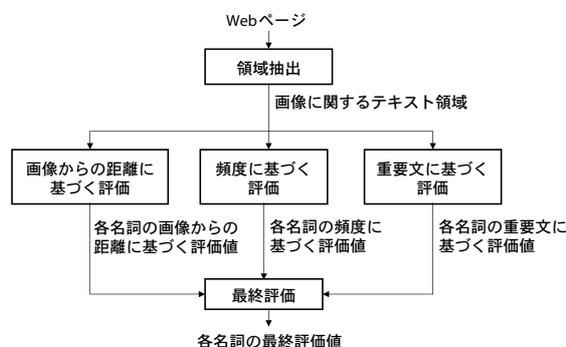


図 1: 本システムの処理の流れ

からの距離に基づいて評価値を付与する。この評価値を与える評価関数は式 (1) に示すように作成した。

$$Dist(w) \stackrel{\text{def}}{=} \max\{1 - P_1 - DIV_1 - BR_1 - PERI_1, 0\} \quad (1)$$

ただし、式 (1) 中の各変数は名詞 w と IMG タグの間に存在する P タグ、DIV タグ、BR タグ、句点の個数が増加するに連れて単調に増加するパラメータである。

3.3 頻度に基づく評価モジュール

頻度に基づく評価モジュールでは入力として各画像に関するテキスト領域を受け取り、その領域中に存在するすべての名詞 w に対してテキスト領域中での出現頻度に応じて評価値を付与する。この評価値を与える評価関数は式 (2) に示すように作成した。

$$Freq(w) \stackrel{\text{def}}{=} \frac{Frequency(w)}{MaxFreq} \quad (2)$$

ただし、式 (2) 中の $Frequency(w)$ は名詞 w の画像に関するテキスト領域中の出現回数を表し、 $MaxFreq$ は領域中でもっとも出現回数が多い名詞の出現回数を表す。

3.4 重要文に基づく評価モジュール

重要文に基づく評価モジュールでは入力として各画像に関するテキスト領域を受け取り、その領域中に存在するすべての名詞 w に対して、テキスト領域中の最重要文と各名詞を含む最

連絡先: 相良直樹, 大阪大学大学院基礎工学研究科, 大阪府豊中市待兼山町 1-3, Tel:(06)6850-6363, E-mail:n-sagara@yachi-lab.sys.es.osaka-u.ac.jp

重要文との距離に基づいて評価値を付与する。この評価値を与える評価関数は式 (3) に示すように作成した。ここで、テキスト領域中からの重要文抽出を行うシステムとしては、高頻度語だけではなく低頻度語にも着目して重要文抽出を行い、さらに情報を採る人間の観点に基づいた重要文抽出も行うことが可能な展望台システムを用いた [砂山 02]。

$$Imp(w) \stackrel{\text{def}}{=} \max\{1 - P_2 - DIV_2 - BR_2 - PERI_2, 0\} \quad (3)$$

ただし、式 (3) 中の各変数はテキスト領域中の最重要文と名詞 w を含む最重要文の間に存在する P タグ, DIV タグ, BR タグ, 句点の個数が増加するにしたがって単調に増加するパラメータである。

3.5 最終評価モジュール

最終評価モジュールでは入力として、これまでに求めた画像に関するテキスト領域中の各名詞に対する「画像からの距離に基づく評価値」、「頻度に基づく評価値」、「重要文に基づく評価値」を受け取り、これらの組合せを行い各名詞 w に最終評価値を付与する。この評価値を与える評価関数は式 (4) に示すように作成した。

$$Label(w) = w_1 \times Dist(w) + Freq(w) + w_2 \times Imp(w) \quad (4)$$

ただし、式 (4) 中の変数 w_1, w_2 は図 2 に示す画像とそれに関するテキスト領域の位置関係によって変化するパラメータである。

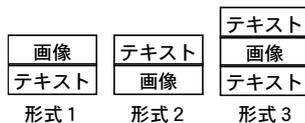


図 2: 画像とテキストの位置関係

4. 本システムの評価実験

4.1 実験方法

画像データベース中の各画像に対して「画像からの距離に基づく評価手法」、「頻度に基づく評価手法」、「画像からの距離」と頻度に基づく評価手法、「重要文に基づく評価手法」、「提案手法」の各手法を実装した各画像ラベリングシステムの出力と人の出力の比較を行う。ここで「画像データベース」、「システムの出力」、「人の出力」の定義を以下に与える。

画像データベース: 検索エンジン Yahoo!JAPAN の画像カテゴリから取得された 200 枚の画像を含む Web ページ集合をデータベースとする。

システムの出力: 各画像に関するテキスト領域中に存在する名詞の中で、各評価手法を実装した画像ラベリングシステムによる評価値が第 3 位までに入った名詞を各システムの出力とする。

人の出力: 被験者 12 名に協力していただき、上記の画像データベース中の各画像に対して、各画像に関するテキスト領域中に存在する名詞のリストから、被験者各人の主観により各画像のラベルとして適当であると判断した名詞 3 つを選択してもらった。その後、被験者によって選択された回数で上位 3 つの名詞を人の出力として決定した。

4.2 実験結果

各評価手法を実装した各画像ラベリングシステムの性能評価を行うために、各システムの出力が「評価値で第 1 位の名詞」、「評価値で第 2 位までの名詞」、「評価値で第 3 位までの名詞」であった場合の各システムの Precision(適合率)と Recall(再現率)を算出し、その結果をグラフ化したものを図 3 に示す。ただし、このグラフ中におけるラベル $Dist$ は画像からの距離に基づく評価手法, $Freq$ は頻度に基づく評価手法, $DistAndFreq$ は画像からの距離と頻度に基づく評価手法, $Importance$ は重要文に基づく評価手法, $Label$ は本提案手法を表す。この図 3 より、同じ Recall では本提案手法がもっとも高い Precision を示していることがわかる。また Recall のみに着目すれば頻度に基づく評価手法が非常に高い値を示しているが、その反面 Precision が低く Recall よりも Precision を重視する画像検索においては優れた手法であるとは言えない。これらのことから本提案手法が従来手法よりも画像ラベリング手法として優れていることが実証された。

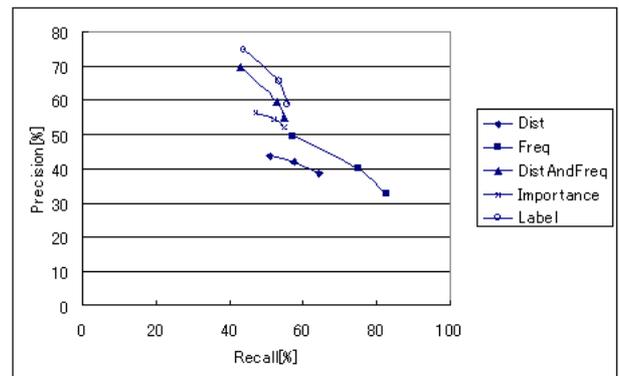


図 3: 各システムの Precision と Recall の関係

5. おわりに

本研究では Web ページ中の画像と、それに関するテキスト領域中の重要文との関係に着目して、HTML テキストの重要文に基づく画像ラベリング手法を提案した。本手法においては「重要文に基づく評価手法」と、従来手法の「画像からの距離に基づく評価手法」、「頻度に基づく評価手法」を組み合わせることによって、従来手法よりも人の感覚に近い、より高精度の画像へのラベル付けが実現できることを評価実験により示した。これにより、本提案手法を画像検索の前段階の処理に用いることによって、よりの確に効率よく必要な画像を検索可能な画像検索システムを構築することが可能となる。

参考文献

[Munson 01] E. V. Munson and Y. Tsymbalenko, "To Search for Images on the Web, Look at the Text, Then Look at the Images", In Proceedings of the First International Workshop on Web Document Analysis, pp.39-42, 2001.

[砂山 02] 砂山渡, 谷内田正彦, "観点に基づいて重要文を抽出する展望台システムとそのサーチエンジンへの実装", 人工知能学会論文誌, Vol.17, No.1, pp.14-22, 2002.