

画像の文脈情報の Web からの抽出と提示

Extraction and Visualization of Image Contexts from Web

是津 耕司[†]田中 克己[†]

Koji ZETTSU

Katsumi TANAKA

Web 上には多種多様な画像が公開されており、それらをキーワードで検索したいという要求が高まってきている。しかし、ユーザが質問として与えるキーワードは、必ずしも検索対象の画像の中身を表しているとは限らず、その画像が用いられている状況や分野・側面を表すことも多い。このようなキーワードは、その画像の周辺に出現することが多い。そのため、検索範囲を画像の周辺コンテンツにまで広げて質問を評価するようにしなければならない。本論文では、この周辺コンテンツを、検索対象の“文脈情報”と呼び、Web ページの文書構造とハイパーリンクに基づいて定義する。文脈情報は、ある画像が Web 上でどのようなコンテンツと一緒に使用されているかを表す。本論文では、画像の文脈情報を Web から自動的に抽出し検索に用いる方法を提案する。これにより、ユーザが検索対象の画像の中身に含まれるキーワードを正確に指定できなくても、その文脈情報から所望のコンテンツを特定することができるようになる。

The Web contains wide variety of images associated with Web contents by Web document structure and hyperlinks. Supposing that a user retrieves images from the Web by keywords, it often happens that some of the query keywords represent usage context of a target image rather than the image content. The usage context indicates subjects, categories and/or aspects in which the image is used. These keywords appear surroundings of the image. Therefore, it is necessary for an image retrieval mechanism to evaluate surrounding contents of images. In this paper, we propose an approach for retrieving images based on their usage contexts. We propose methods for extracting, retrieving and visualizing usage context of images from the Web. Our approach enables for a user to retrieve images based on their usages, even if he/she does not have enough knowledge about the images.

1. はじめに

Web 上には、様々な Web ページの一部として大量の画像が公開されている。近年、これらの画像を、Web ページと同様、キーワードで検索したいという要求が高まってきている。従来の画像検索では、画像の物理的特徴量（色や形状など）や内容記述に基づいて、特定の中身を持つ画像を検索するこ

とが行われていた。しかし、不特定多数の画像を対象とする Web の画像検索では、個々の画像を特定する確かなキーワードを指定することは困難である。一方、これらの画像は Web ページの一部として使われており、画像の使用箇所の周辺を見れば、各画像がどのような目的や話題で用いられているのかが分かる。我々は、このような情報を**画像の文脈情報**（以下、単に文脈）と呼ぶ。画像の文脈は、画像の周辺に存在するコンテンツ集合（単語や他の画像集合）によって表され、その画像が Web 上でどのようなコンテンツと一緒に使われているかを示している。文脈情報を手がかりに画像を検索することができるようになれば、画像の中身を正確に知らなくても、その画像が使われている Web ページの内容や画像に関連付けられた話題から所望の画像を特定することが可能になる。

本論文では、画像の文脈の基本概念を説明するとともに、画像の文脈を Web から抽出する方法、および抽出結果を可視化する方法について述べる。また、文脈の類似性の評価に基づき、類似した文脈で使われている画像を検索する方法についても述べる。

2. 画像の文脈情報

画像の文脈は、Web 上で画像の周辺に存在するコンテンツ（単語や他の画像）の集合として定義される。明確な文法を持つテキストと異なり、画像は、物理的周囲に存在するコンテンツだけでなく、画像の使用箇所と文書構造的に隣接するコンテンツや、画像の使用箇所を参照しているコンテンツとも関連付けられている。我々は、従来、物理的な周囲のみを指していた“周辺”の概念を拡張し、以下の3種類の文脈を定義する：

Surrounding context：画像の物理的な周囲に存在するコンテンツ集合によって、画像がどのようなコンテンツと一緒に使われているかという文脈を表す。

Structural context：文書構造に基づき、画像を含む段落とその上位段落・下位段落のコンテンツ集合によって、画像がどのような話題の中で使われているかという文脈を表す。

Referential context：リンク構造に基づき、画像を直接あるいは間接的にリンクしているコンテンツの集合によって、画像がどのような内容からどのように参照されているかという文脈を表す。

図 0 に、Web における画像の文脈の例を示す。Surrounding context は、画像の前後に位置する単語や画像の集合 (s_2, s_1, s'_1, s'_2) によって表される。Structural context は、Web ページの文書構造に基づいて、画像を直接含む段落 P_1 、およびその上位段落 (P_2, P_3) ・下位段落に含まれる単語や画像の集合によって表される。Referential context は、画像を含む（インライン参照している）Web ページのタイトル T_i 、このページをリンクしているリンクアンカー A_i 、およびリンクアンカー A_i を介してこのページを参照している Web ページのタイトル T_2 によって（各々に含まれる単語や画像の集合によって）表される。

3. 文脈の抽出

Web から画像の文脈を抽出するためには、抽出対象となる周辺範囲を特定しなければならない。我々は、画像の周辺を特徴付けるのに十分なコンテンツを含む範囲を文脈として特定する。ある画像の周辺は、以下のようなコンテンツに

[†] 正会員 京都大学大学院情報学研究科社会情報学専攻
{zettzu, tanaka}@dl.kuis.kyoto-u.ac.jp

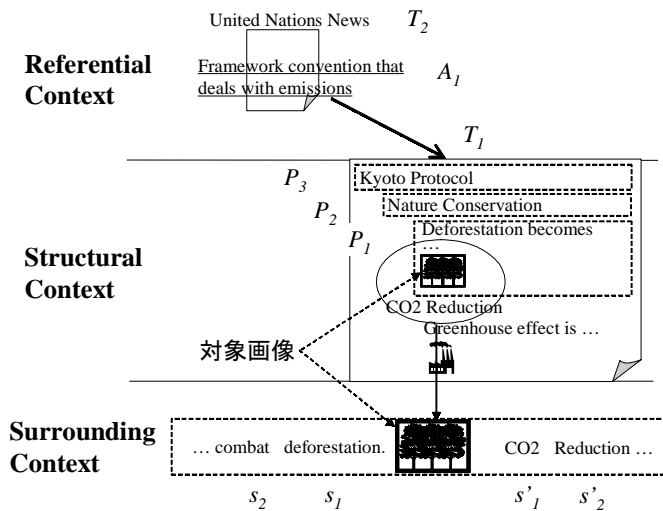


図 0 Web における画像の文脈
Fig.1 Contexts of an image in the Web.

よって特徴付けられる：

- その画像の周辺にのみ頻出し、ほかの画像の周辺にはあまり出現しない単語や画像
- その画像の近くに出現する単語や画像

与えられた画像 I に対し、各種類の文脈ごとに、以下の条件を満たす周辺範囲 R を特定する：

1. 画像 I の文脈 $C(I)$ に含まれるすべての周辺コンテンツは、画像 I を含む周辺範囲 R の中に出現する。
2. 周辺範囲 R の中に出現する周辺コンテンツの文脈貢献度の合計が既定の閾値を超えている。
3. 周辺範囲 R のどの部分範囲 R' も上記 1 および 2 を満たさない。

周辺コンテンツ o の文脈 $C(I)$ に対する文脈貢献度は、周辺コンテンツ o と同じ単語や画像が文脈 $C(I)$ にのみ集中して出現する程（他の画像の文脈にはあまり出現しない程）高くなり、また周辺コンテンツ o が画像 I の近くに出現する程高くなるように評価される。この文脈貢献度は、以下の式で求められる：

$$ccontrib(o, C(I)) = adensity(o, C(I)) \frac{b}{distance(o, I)}$$

ここで、 $adensity(o, C(I))$ は周辺コンテンツ o の文脈 $C(I)$ に対する出現密度であり、 $distance(o, I)$ は周辺オブジェクト o と画像 I の距離である。 b は定数である。周辺コンテンツ o の文脈 $C(I)$ に対する出現密度は、周辺コンテンツ o と同じ単語もしくは画像の全出現頻度（検索対象の全ての画像の文脈での出現頻度）に占める文脈 $C(I)$ での出現頻度の割合を示す。また、周辺コンテンツ o と画像 I との距離 $distance(o, I)$ は、各種類の文脈ごとに以下のように定義される：

Surrounding context の距離： 画像 I と周辺コンテンツ o の間に存在する単語および画像の数。

Structural context の距離： 画像 I を含む段落から周辺

コンテンツ o を含む段落に至るまでに辿る親子関係の数。

Referential context の距離： 画像 I から周辺コンテンツ o を含む Web ページ（タイトル）もしくはリンクアンカーに至るまでに辿る Web ページおよびリンクアンカーの数。

4. 文脈の可視化

抽出された文脈を、以下の目的で可視化する：

- 画像と各周辺コンテンツ、および周辺コンテンツどうしのつながりを連続的に表示し、画像がどのような内容とどの程度強く関係しているのかを直感的に把握できるようにする。
- 複数の文脈を同時に比較可能な形で表示し、それらの類似性や差異に基づいて個々の文脈の特徴を的確に把握できるようにする。

我々は、テキストの文脈表示方法としてよく知られている KWIC (KeyWord In Context)[1]をマルチメディアに拡張した表示方法を提案する。提案手法では、文脈の種類ごとに、抽出された各文脈に含まれる周辺コンテンツを画像を中心に一列に並べて表示し、それらをリストして一覧表示する。各文脈は、文脈の種類ごとに以下の方法で表示される：

Surrounding context の表示： 画像の前後に位置する周辺コンテンツを、画像の前語にそれぞれ表示する。

Structural context の表示： 文書構造に従って、画像の上位段落に位置する周辺コンテンツを画像の前に表示し、画像の下部段落に位置する周辺コンテンツを画像の後ろに表示する。

Referential context の表示： リンク構造に従って、画像に向かって参照してくるリンク系路上に存在する周辺コンテンツを画像の前に参照順に表示する。

図 0 に、図 0 の文脈を表示した例を示す。この表示方法により、画像を中心に左右を見れば周辺コンテンツとのつながりが一目で把握することができる。また、画像を中心に複数の文脈を一覧表示することで、周辺コンテンツの比較が容易に行える。更に画像の物理的特徴量に基づいて文脈一覧をソートすることにより、見た目の類似した画像の文脈を近くに配置し、類似画像の多様な文脈を把握し易くする。

図 0 に、Web 画像検索エンジン(Google)を使って検索した画像（キーワード “warning sign”）の文脈を表示した例を示す。図 0 は、同じ画像に対する 3 種類の文脈を表示している。先行の画像の文脈に着目すると、surrounding context からは、この画像が毒物に関する警標 (“venom”, “warning sign”) であることが分かる。一方、structural context からは、この画像が毒物学 (“toxinology”) の中でオーストラリアのクラゲに関する話題の中で使われていることが分かる。さらに、referential context からは、この画像が個人のホームページと snake dancer に関する Web ページからリンクされていることが分かり、この画像がどのような Web ページで使われているかを異なる側面から表している。このように、画像の文脈情報からは画像に関する様々な背景情報が得られ、ユーザは、従来のように画像キーワードを試行錯誤して再検索しなくても、文脈情報を基に画像を絞り込むことができる。

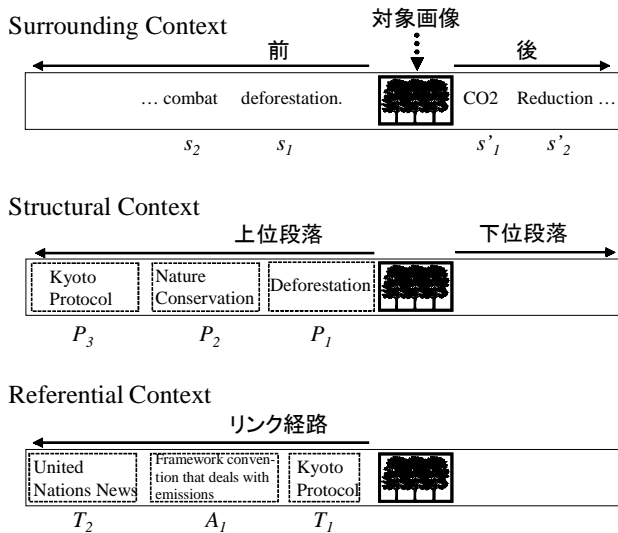


図 0 各種類の画像の文脈の表示
Fig.2 Visualizing each type of image context.

5. 文脈の類似性に基づく画像検索

ある Web ページのある箇所を使用するのにふさわしい画像を検索する場合、使用箇所の内容と画像のキーワードが必ずしも一致しているとは限らず、従来は様々なキーワードを試行錯誤しながら画像検索を行わなければならなかった。この問題に対し、我々は、画像の文脈に基づき、使用箇所と類似した文脈で既に使われている画像を Web から検索する方法を提案する。

画像の検索は、以下の手順で行われる。

1. ユーザが、ある Web 文書中で画像を使用したい箇所

を指定する。

2. 検索メカニズムは、指定された箇所の周辺から文脈情報を抽出し質問を生成する。
3. 検索メカニズムは、検索対象の各画像の文脈と質問の文脈の類似度を評価し、既定の閾値以上の類似度を持つ文脈に対応する画像を検索結果として返す。検索結果は、文脈の類似度に基づいてランキングされる。
4. ユーザは、検索結果の中から、文脈の内容と画像の中身に基づいて使用すべき画像を選択する。

文脈の類似度は、ある 2 つの文脈が類似した周辺コンテンツによってどの程度同じ様に特徴付けられているかを表す。画像 I_i の文脈 $C(I_i)$ と画像 I_j の文脈 $C(I_j)$ の類似度は次のようにして求められる。まず、 $E = \{e_1, e_2, \dots, e_n\}$ を、文脈 $C(I_i)$ と文脈 $C(I_j)$ に含まれるキーワードおよび画像の集合として表す。次に、文脈 $C(I_i)$ を、 E の各要素を成分とする特徴ベクトル $f(C(I_i)) = (v_1, v_2, \dots, v_n)$ で表わす。 v_k は、文脈 $C(I_i)$ の中で e_k と同一のキーワードもしくは画像を表わす周辺オブジェクトの文脈貢献度の合計である。同様にして、文脈 $C(I_j)$ も特徴ベクトル $f(C(I_j))$ で表わす。最終的に、文脈 $C(I_i)$ と文脈 $C(I_j)$ の類似度は、対応する特徴ベクトル $f(C(I_i))$ と $f(C(I_j))$ のコサイン相関値[2] として算出される。

$$\text{similarity}(C(I_i), C(I_j)) = \frac{\mathbf{f}(C(I_i)) \cdot \mathbf{f}(C(I_j))}{|\mathbf{f}(C(I_i))| |\mathbf{f}(C(I_j))|}$$

図 0 に、プロトタイプによる画像検索の様子を示す。図 0 には、結果画像の文脈の一覧が表示されている（文脈表示の詳細については第 4 章を参照）。検索結果の画面では、ユーザが指定した使用箇所の文脈に含まれる単語と同じ単語に下線が引かれている。先頭行に示された文脈では、“Warning” と “Motorists” が指定された箇所と結果画像の両方の文脈に出現し、かつ “Motorists” は他の結果画



図 0 画像の文脈の表示例
Fig.3 An example of visualizing usage contexts of images.



図 0 類似文脈による画像検索
Fig.4 Query images by similar contexts.

像の文脈には殆ど出現しない(この文脈に対する出現密度の高い)。従ってこれらの単語の文脈貢献度は共に高く、それらによって特徴付けられる双方の文脈の類似度も高くなっている。このように、各結果画像の文脈が指定した箇所の文脈とどの様に類似しているのかを確認しながら、最もふさわしい画像を選択することができる。

6. 関連研究

従来の Web 検索における文脈[4,5]とは、検索する範囲のことを指し、Web ページの閲覧履歴や検索履歴などからユーザの興味に関するキーワードを抽出し、検索対象を絞り込むことを行っていた。一方、本研究における文脈とは、検索対象(画像)の周辺コンテンツを指し、検索対象がどこでどの様に使われているかという情報に基づいて検索を行うために文脈を用いている。

Google や Altavista などの Web 検索エンジンでは、Web ページ内で画像の近傍にあるテキストを使ってキーワードによる画像検索を実現している。また、[3] では、画像を含む Web ページのリンク元ページから画像のキーワードを抽出して検索することを行っている。これらは、本研究と同様、周辺コンテンツを用いて画像の検索を行っているが、検索結果の画像と周辺コンテンツとの関係がユーザに明示されないため、検索結果の適合性を評価するために各画像の実際の使用箇所を逐一確認しなければならない。これに対し、本研究では、3 種類の文脈によって画像と周辺コンテンツの関連性を明確に規定し、文脈情報の可視化によってそれらをユーザに明示することでこの問題を解決している。

7. まとめと今後の課題

本論文では、Web コンテンツの構成(構造化文書およびリンク)に沿って 3 種類の周辺を定義し、それぞれの周辺から画像の使われ方を特徴付けるのに十分なコンテンツを含む周辺範囲を文脈として抽出し、それらを可視化する方法を提案した。また、抽出された文脈に基づいて、ある特定の文脈に類似した文脈で使われている画像を Web から検索する方法についても述べた。今後の課題としては、文脈に基づく画像検索の評価方法の確立とスケーラビリティの向上が挙げられる。

【謝辞】

本研究は、一部平成 15 年度科研費特定領域研究(2)「Web の意味構造に基づく新しい Web 検索サービス方式に関する研究」(課題番号:15017249, 代表:田中克己)による。ここに記し謝意を表します。また、本研究は、一部 21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」(代表:上林弥彦)による。ここに記し謝意を表します。

【文献】

- [1] H. P. Luhn. Keyword in context index for technical literature (kwic index). American Documentation, No. 11, pp.288-295, 1960.
- [2] G. Salton and M. McGill. Introduction to modern information retrieval. In McGraw Hill, 1983.
- [3] V. Harmandas, M. Sanderson, and M. D. Dunlop. Image retrieval by hypertext links. In Proceedings of the ACM SIGIR '97 Conference on Research and Development in Information Retrieval, pp. 296-303, 1997.
- [4] Lawrence, S.: Context in Web Search, IEEE Data Engineering Bulletin, Vol. 23, No. 3, pp.25-32 (2000).
- [5] Finkelstein, L., Gabilovich, E., Matias, Y., Rivlin, E., Solan, S., Wolfman, G. and Ruppim, E.: Placing Search in Context: The Concept Revisited, ACM Transactions on Information Systems, Vol. 20, No. 1, pp. 116-131 (2002).

是津 耕司 Koji ZETTSU

1992 年東京工業大学工学部情報工学科卒業。1992 年日本アイ・ビー・エム株式会社。2003 年 独立行政法人通信総合研究所専攻研究員。現在、京都大学大学院情報学研究科博士後期課程に在学中。マルチメディアデータベース、情報検索の研究開発に従事。日本データベース学会、情報処理学会、ACM 各会員。

田中 克己 Katsumi TANAKA

1974 年京都大学工学部情報工学科卒業。1976 年同大学院修士課程修了。1979 年神戸大学教養部助手、1986 年同大学工学部助教授。1994 年同大学工学部教授(情報知能工学科)。1995 年同大学大学院自然科学研究科情報メディア科学専攻専任教授、2001 年京都大学大学院情報学研究科社会情報学専攻教授、現在に至る。工学博士。主にデータベースの研究に従事。日本データベース学会、人工知能学会、日本ソフトウェア科学会、IEEE Computer Society、ACM 各会員。