

Web テキストを利用した画像クラスタリングシステム

Image Clustering System using Web Texts

砂山 渡^{*1} 永田 晶子^{*2} 谷内田 正彦^{*2}
 Wataru Sunayama Wataru Sunayama Masahiko Yachida

広島市立大学^{*1} 大阪大学^{*2}
 Hiroshima City University Osaka University

As the internet becomes the basic resource of information, not only texts but images retrieval systems have been appeared. However, many of those supply only a list of images, so we have to seek the expecting images one by one. Although, image labeling is one of the solutions of such a problem, various words are labeled to an image if the words are extracted from only one Web page. Therefore, this paper proposes an image clustering system that labels by words related to a search keyword. This relationships are measured by Web pages. By the experimental results, users were enabled to find the intended images more faster than the ordinal image search system.

1. はじめに

現在、インターネット上にはさまざまな画像検索エンジンが存在する。Google[Google Image]では画像ファイルに付けられている説明文 (IMG タグの ALT 属性) や、本文中のキーワードを利用してラベル付けを行っているが、検索キーワードとは無関係な画像も少なからず検索結果に含まれている。また、インターネット全体を検索する画像検索エンジンの他に、毎日新聞の毎日フォトバンク [毎日フォトバンク]、や画像販売サイト PHOTOS IMAGE SHOP[PHOTOS] などには、自社サイト内の画像データベースを検索する検索エンジンもある。このような限られた範囲内のデータベースであれば、人手によって各画像にラベル付けを行うことや、画像を階層化されたディレクトリに分類して保存することが可能であり、精度良い画像検索が可能になる。

画像検索には大きく分けて、画像に与えられたキーワードに基づく画像検索 (TBIR) と、色情報など画像の内容に基づく画像検索 (CBIR) の2種類があり、これらを組み合わせたシステムがいくつか提案されている。コロンビア大学が研究している WebSEEK[Smith 97] は、ユーザが入力した検索キーワードに関連するカテゴリを文字で提示し、ユーザが選んだカテゴリ内の指定画像に対して CBIR の手法により類似する画像を検索する。WebSeer[Frankel 96] はキーワードだけでなく、画像の大きさ、主要な色、写真か図か、人物が写っているか否かなどの画像特徴をユーザに指定させてから画像の検索を行う。これらのシステムでは、各画像が予めカテゴリ分けされていることが前提となっていることに加え、ユーザによる様々なデータの入力前提となっているため精度は高いが、操作が煩雑で手間がかかるという欠点がある。

また、画像検索の結果を周辺コンテンツから抽出された Web 文脈とともに提示するシステムがある [是津 03]。Web 文脈とは画像の前後にある文や、ハイパーリンクによって画像と関連付けられた周辺コンテンツから抜き出したリンクアンカー、Web ページタイトルなどであり、検索要求と画像との関連性を視覚的に捉えられるという利点がある。しかし、単純に画像前後の文を抜き出しているため、画像と文の関連性が低いことも多く精度の改善が望まれることや、文という単語よりも多く

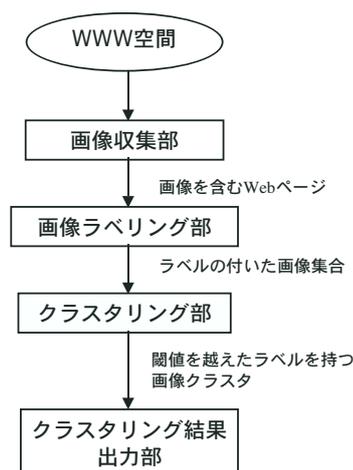


図 1: システムの処理の流れ

の情報を提示しているため、検索結果の適切な視覚化インタフェースが必要と考えられる。

これらをふまえて本研究においては、インターネット上の Web ページに含まれる画像を対象にして、Web ページ中の各画像にテキスト中の重要文を用いたラベル付け手法 [相良 04] を応用して自動的にラベルを付与し、得られたラベル集合に対して独自の手法によってノイズ除去と絞り込みを行った上で、画像のクラスタリングを行う手法を提案する。

2. Web テキストを利用した画像クラスタリングシステム

本研究で提案する画像クラスタリングシステムの処理の流れを、図 1 に示す。本システムは、WWW 空間から画像 URL を取得する画像収集部、各画像にラベルを付与する画像ラベリング部、ラベルに基づいて画像の分類を行うクラスタリング部、結果の表示を行うクラスタリング結果出力部の 4 つのモジュールからなる。

2.1 画像収集部

画像収集部では、Google イメージ検索 [Google Image] を用いて、検索キーワードに関連する画像を含む Web ページ集

連絡先: 砂山渡, 広島市立大学情報科学部,
 731-3194 広島市安佐南区大塚東 3-4-1, 082-830-1705,
 sunayama@sys.im.hiroshima-cu.ac.jp

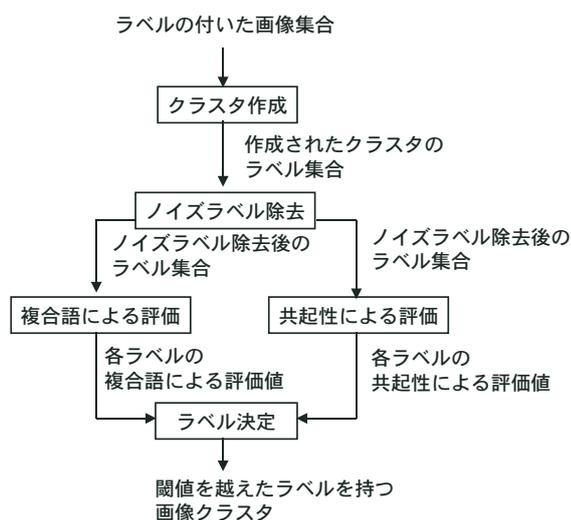


図 2: 画像クラスタリング部の処理の流れ

合を獲得する。その後、獲得した Web ページ集合から、画像へのリンクを抽出して WWW から画像ファイルを収集する。

2.2 画像ラベリング部

画像ラベリング部では、収集した各画像に画像のタイトルを表すラベルを付与する。画像ラベリングには、HTML テキストの重要文を用いた画像ラベリングシステム [相良 04] を用いている。このシステムでは Web ページ中に存在する画像に関連するテキスト領域を抽出し、そのテキスト領域中に存在する各名詞に対して「画像からの距離」「テキスト領域中での出現頻度」「テキスト領域内の重要文」をそれぞれ評価することによって、従来の画像ラベリング手法よりも精度の高いラベル付けを実現している。

2.3 クラスタリング部

クラスタリング部では、画像集合を各画像に付与されたラベルを元にクラスタリングを行う。クラスタリング部の処理の流れを図 2 に示す。

まず、ラベルの付いた画像集合を入力としてクラスタ作成モジュールに与えクラスタ分けを行う。その後、各クラスタのラベルがラベルとして適切であるかどうかを判断する指標を用いて、ノイズラベルの除去およびクラスタラベルの絞り込みを行った上で出力クラスタを提示する。以下で、画像クラスタリング部の詳細を述べる。

2.3.1 クラスタ作成モジュール

クラスタ作成モジュールでは、検索結果として得られた画像に与えられているすべてのラベルを対象にして、各ラベルに対応するクラスタを作成する。すなわち、各クラスタは画像に与えられたラベルをクラスタラベルとしてもち、各画像は与えられている全てのラベルに対応するクラスタに重複を許して分類される。

2.3.2 ノイズラベル除去モジュール

前項で作成したクラスタのラベルには一般的な語も含まれている。そのため、検索キーワードと強く関連があるわけではなく、多くのページに共通に出現する可能性が高い単語によるラベルをノイズラベルとして除去する。

そこで各ラベル $label$ に、式 (1) と式 (2) で表される相対頻度による評価値を付与する。ただしラベルの頻度 $hit(label)$ と

は、ラベルを含む Web ページ数であるとし、本システムでは Google 検索 [Google] のヒット件数で測る。また、式中の L は全ラベル集合を表し、 key は画像検索に用いられた検索キーワードを表す。

$$Pr = \frac{hit(label)}{\sum_{label \in L} hit(label)} \quad (1)$$

$$Pc = \frac{hit(label \cap key)}{\sum_{label \in L} hit(label \cap key)} \quad (2)$$

式 (1) は全ラベル集合における相対的なラベルの出現頻度 (すなわち、ラベル集合における出現確率) を表し、式 (1) は検索キーワード key とラベル $label$ の相対的な共起頻度 (すなわち、ラベル集合における共起確率) を意味する。ここで仮に、検索キーワードがラベルと全く独立に生起するなら、全ラベルの Pc の分布は Pr と同様の分布になるはずである。逆に、検索キーワードとラベルとの間に何らかの依存関係があれば、分布に偏りが生じて Pc と Pr の値が異なると考えられる。そこで、ラベル集合 L における、検索キーワードとの関連の強さを式 (3) で表す。

$$Relation(key, label) = Pc - Pr \quad (3)$$

すなわち、ラベルの単独での出現確率よりも共起確率の方が高いほど検索キーワードとラベルとの間の関連が強いとみなす。具体的には、この式 (3) の評価値 $Relation$ が負になるラベルをノイズラベルとして取り除く。

2.3.3 複合語による評価モジュール

前項による方法でノイズラベルを除去した後、より適切なクラスタラベルを選ぶためにラベルの絞り込みを行う。その 1 つ目の方法として複合語による評価を行なう。

例えば「ひまわり」を検索語として用いた場合に、クラスタラベルとして「畑」があったとする。この時、「ひまわり畑」のように検索キーワードとクラスタラベルの複合語が存在する可能性が高いラベルは、検索キーワードに深く関連すると言える。そこで式 (4) による評価値 $Phrase$ を各ラベルに与える。ただし、 $hit(P_1)$ は複合語「key+label」の hit 数、 $hit(P_2)$ は複合語「label+key」の hit 数である。

$$Phrase(key, label) = \max_{i=1,2} hit(P_i) \quad (4)$$

2.3.4 共起性による評価モジュール

クラスタラベルを絞り込むための 2 つ目の方法として、共起性による評価を行なう。

例えば「ひまわり」を検索語として用いた場合に、クラスタラベルとして「種」があったとする。この時、検索キーワードとクラスタラベルの複合語「ひまわり種」や「種ひまわり」が存在する可能性は低い。一般に、「ひまわり」と「種」には強い関係があると考えられ、多くの Web ページ中で同時に用いられている可能性が高い。そこで、式 (5) による評価値 $Connection$ を各ラベルに与える。

$$Connection(key, label) = \frac{hit(key \cap label)}{hit(label)} \quad (5)$$

2.3.5 ラベル決定モジュール

ノイズラベル除去モジュールでノイズラベルを取り除いた後、複合語による評価モジュールと共起性による評価モジュールでラベルに評価値を付け、その評価値をもとに出力ラベルとクラスタの提示順序を決定する。

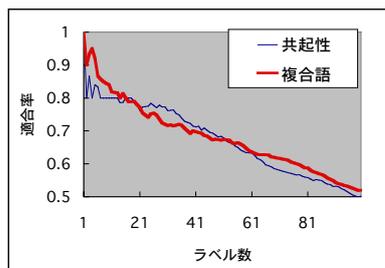


図 3: ラベル評価モジュールの出力ラベル数と適合率の関係

ここでまず、ノイズ除去モジュール、共起性および複合語による評価モジュールの評価精度を調べる予備実験を行った。実験においては、出力されるラベルと検索語がともに出現する画像が想像できるか否かを基準として、5名の大学生および大学院生に5つの検索語「ひまわり」「阪神」「ジュース」「宇宙」「ケーキ」のラベルを評価してもらい、2人以上がノイズと判定したラベルをノイズラベルとし、残りを正解ラベルとした。ノイズ除去モジュールにおいては、合計772個のラベルを428個まで55%に絞り込むと同時に、適合率も37%から54%に改善されるという結果が得られた。

また、共起性および複合語による評価モジュールの評価値順にラベルを出力したときのラベル数と適合率との関係(5つの検索語の平均値)を図3に示す。両評価値とも右下がりのグラフになっており、ラベルに対する正しい評価が与えられていることに加えて、上位30ラベル程度の出力を行う場合の適合率が約70%と高く、クラスタリング結果を利用した画像検索効率の向上が見込まれる。

複合語による評価値 *Phrase* は、上位15ラベルくらいまで共起性による評価値よりも高い適合率が得られている。これは一定のしきい値以上の値をもつラベルは、検索キーワードとの複合語が存在する可能性が高いことを表している。そこで、90%の適合率を与える *Phrase* のヒット件数を調べ、評価値 *Phrase* の閾値を1000と定めた。

また、共起性による評価値 *Connection* の閾値はシステムの出力ラベル数により可変とする。すなわち出力ラベル数を N 、評価値 *Phrase* が1000以上であるラベルの数を N_p とした場合、 $N_c = N - N_p$ 個のラベルを評価値 *Connection* の上位から順に採用する。

2.4 クラスタリング結果出力部

クラスタリング結果出力部ではクラスタラベルとクラスタに属する画像をブラウザに出力する。「ひまわり」を検索語としたときの出力インタフェースの例を図4に示す。出力インタフェースは2段のフレーム構造になっており、左側のフレームにクラスタラベルのインデックスと各クラスタに含まれる画像数を表示する。ユーザが見たいラベルを選択すると、右側のフレームに選択されたラベルを含む画像集合が表示される。

3. 評価実験

本章では、提案する画像クラスタリングシステムにおけるクラスタリングや画像のラベル提示が、ユーザが画像検索を行う際にどのような効果を与えるかを実験により検証する。すなわち、従来型の画像検索システムに比べてすばやく目的の画像



図 4: 出力インタフェース

表 1: 検索キーワードと与えた画像検索課題

検索キーワード	画像検索課題
映画	映画の広告・ポスター
ケーキ	チョコレートケーキ
ジュース	りんごジュース
癒し	癒し空間
ファッション	服がメインで写っている
阪神	阪神タイガースの選手
ひまわり	花のひまわりの絵
宇宙	宇宙空間に地球がある
携帯電話	カメラ付き携帯電話

を獲得できること、また画像検索における文字情報として画像ラベルを出力することの効果を確認する。

比較システムとして、「システム1: 並列提示画像検索システム」「システム2: 本システム(クラスタリングのみ)」「システム3: 本システム(クラスタリング・ラベル)」の3つを用意した。それぞれの出力にはGoogleイメージ検索 [Google Image] による上位500件の画像の中から、システム1は、Googleによる上位100件の画像、システム2とシステム3は2章で述べた方法により出力の上位100件の画像を提示した。

実験は被験者にこれら3つのシステムを用いて、検索結果となった画像集合の中から表1の課題条件に合う画像を見つけるまでに要した時間を各システムで測定、比較することで行った。実験に用いた検索キーワードは「映画」「ケーキ」「ジュース」「癒し」「ファッション」「阪神」「ひまわり」「宇宙」「携帯電話」であり、画像検索エンジンでのヒット数が比較的高い単語とした。課題は画像集合中に複数(5枚以上)あるものという基準で作成した。

これら各課題の条件を満たす目標画像を、画像集合の中から1課題に対し2枚選んでもらい、1枚目を見つけた時間、2枚目を見つけた時間をそれぞれ測定した。被験者はコンピュータ使用経験のある大学生、大学院生の男女18名であり、画像の既視感による誤差をなくすため、被験者は各課題で1つのシステムのみを使用し、各人が均等に3つのシステムを使用するように課題とシステムを割り振った。

本実験の結果として、1画像目を見つけるまでの平均所要時間が図5のように得られた。

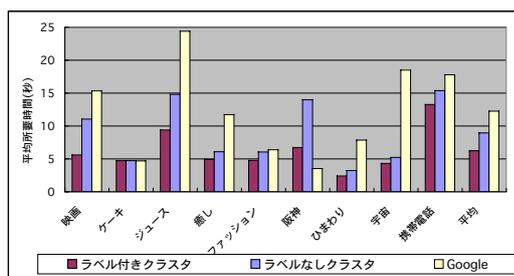


図 5: 1 画像目を見つけるまでの平均所要時間

3.1 クラスタリングの効果に対する考察

システム 1 とシステム 2 の結果を比較することにより、画像検索結果の提示におけるクラスタリングの効果について考察する。図 5 に示す平均所要時間においては、ほとんどの検索キーワードにおいて、クラスタリングを用いた提示方法の方が早く目的の画像が得られている。これは、クラスタ内の画像を 2, 3 枚見た後に、被験者の頭の中で各クラスタに属する画像集合のラベルづけが行われ、目標画像に関連がないラベル付けがなされたクラスタを見ることなく画像を閲覧できたためと考えられる。すなわち、文字による明示的なクラスタのラベル付けがなされていない場合においても、検索結果をクラスタ表示することによって全画像を見る必要がなくなり検索時間を短縮するという効果が確認された。

キーワードごとに結果を分類すると「映画」「ジュース」「癒し」「ひまわり」「宇宙」「携帯電話」においては平均 2 秒以上、システム 2 が早いという結果が得られているが、「ケーキ」「ファッション」においてはあまり差が見られず、「阪神」においてはシステム 1 の Google イメージ検索の方が良い結果となっている。

これはすなわち、多くの画像が提示される中から目標となる画像を探索する場合、同時に提示される画像との視覚的な類似性も、発見に要する時間に影響を及ぼすと考えられる。今回の実験において、ラベルなしクラスタの方が良い結果が得られた「ジュース」「宇宙」「ひまわり」「携帯電話」の課題に関しては、形状や色などが似通っている視覚的に大差がないものの中から目的の画像を探し出す必要があった。このような場合、特に一つ一つの画像をじっくり見る必要があり識別に時間がかかることが予想されるが、クラスタリングによって似通った画像が分類されることによって全画像を見る必要がなくなり、検索時間の短縮につながったと考えられる。逆に、ケーキの画像集合から「チョコレートケーキ」を探す課題は、ケーキの色に特徴があり視覚的に区別しやすく画像の発見が容易であったために、システム間での差が少なかったと考えられる。

また、阪神の画像集合から「阪神タイガースの選手」を探す課題においては、阪神タイガースの選手が含まれると予想される「タイガース」クラスタが作成されたにも関わらず、このクラスタ内に選手の画像が含まれなかったことによる。各選手の画像は選手個人名のクラスタに分類されているため、被験者は下位の選手個人名のクラスタに気付く前に、上位にある阪神タイガースクラスタから画像を探してしまい検索時間がかか

たと考えられる。

3.2 ラベル提示の効果に関する考察

次に、システム 2 とシステム 3 の結果を比較することにより、画像検索結果提示におけるラベル提示の効果について考察する。図 5 に示す平均所要時間においては、ほとんどの検索キーワードにおいて、ラベル提示を行った方が早く目的の画像が得られている。これは、クラスタがラベルなしに提示されただけでは、そのクラスタが何の基準（ラベル）で分類されているのかが分からず、被験者自身がクラスタ内を見比べて何の画像の集まりであるかを判定しなければならなかったことによる。すなわちラベルがないシステム 2 においては、被験者が自身の頭の中でラベル付けを行う時間分だけ、検索に要する時間が長くなったと考えられる。このことから、各クラスタのラベル提示によってユーザの画像集合へのラベル付けを助ける効果があることが確認された。

特に図 5 において、このラベル付けの効果が検索キーワード「阪神」の結果において顕著に現れている。これはラベルが付与されていることで上位のタイガースクラスタに選手の画像が発見できなかった場合においても、下位にある選手の個人名がついたクラスタをすぐに発見できたためと考えられる。

4. 結論

本論文では、画像検索システムの検索結果をクラスタリングを行い、各クラスタのラベルを明示的な言葉として提示するシステムを提案し、既存の画像検索エンジンとの比較実験を行うことで、目的の画像を素早く獲得できることを示した。本システムでは検索キーワードとラベルとの関係性を評価することで、検索キーワードと強い関係を持つラベルを提示し、検索効率を下げることなく使いやすい画像クラスタリングをシステムを実現している。

参考文献

- [Frankel 96] Frankel, C., Swain, M. and Athitsos, V.: Web-Seer: An Image Search Engine for the World Wide Web, Technical Report TR-96-14, University of Chicago, (1996).
- [Google] 検索エンジン Google : <http://www.google.co.jp/>
- [Google Image] Google 画像検索 : <http://images.google.co.jp/>
- [毎日フォトバンク] 毎日フォトバンク <http://photobank.mainichi.co.jp/>
- [PHOTOS] PHOTOS IMAGE SHOP http://www.hmnet.co.jp/photos/shop/search_main.asp
- [相良 04] 相良直樹, 砂山渡, 谷内田正彦: HTML テキストの重要文を用いた画像ラベリング手法, 電子情報通信学会論文誌, Vol.J87-D1, No.2, pp.145 - 153, (2004).
- [Smith 97] Smith, J. and Chang, S.: Visually Searching the Web for Content, IEEE Multimedia, Vol.4, No.3, pp.12 - 20, (1997).
- [是津 03] 是津耕司, 田中克己: Web からの画像の文脈情報の抽出と提示, 第 14 回データ工学ワークショップ, 6-P-05, (2003).