

D-004

親ページ分類を用いた WEB 画像の実用的な検索方法 The Practical Method of Web Image Retrieval Using Linked Page Classification.

植松 幸生[†]
Yukio Uematsu

竹野 浩[‡]
Hiroshi Takeno

片岡 良治[†]
Ryoji Kataoka

1. 背景

今日、WEB 検索は必要不可欠なインターネットポータルサービスの一つになっている。WEB 検索とはユーザが入力する情報に基づいて、そのユーザが要求していると推測される情報を含む WEB ページを提示するサービスのことである。その検索対象を画像に限定した WEB 検索を WEB 画像検索と呼ぶ。近年のブロードバンド化から、画像をテキストと同じように検索したいという要求が生じ、WEB 画像検索への期待が高まりつつある。このような WEB 画像検索はサービスが開始されたばかりであり、確固たる検索手法はまだ確立されていない。

2. 関連研究

WEB 画像検索の従来技術として 2 つの異なる検索方法がある。1 つは goo, Google 等で既に商用のサービスとして行われているキーワード画像検索である。それらは画像をリンクする WEB ページ (以降 親ページ) のタイトル、アンカーテキストや画像リンク近傍のテキストを画像と関連付け、既存の WEB 検索と同様にテキストを入力して、検索結果となる画像を提示する。キーワードと画像の関連付けは保証されていないため、精度は高くないが、既存の WEB 検索と同じ構成でサービスを提供出来る点が優れている。

もう 1 つの従来技術としては、類似画像検索がある [1]。類似画像検索では画像から事前に抽出した画像特徴量をインデックスとして、入力となった画像との画像特徴量同士を比較し、類似した画像を提示する。検索条件、検索対象がともに画像であるため、直感的な検索が可能である。

また、キーワード画像検索の結果を入力として類似画像検索を行うような 2 つの WEB 画像検索を連携させる方法が近年研究されている [2]。Smith 等は WEB 画像に付与された URI 等から意味的なディレクトリを作成し、キーワード画像検索時にディレクトリ検索を行う。この方法は最初にディレクトリに限定したキーワード画像検索を行う点と、リンク近傍のテキストを利用していない点等から、精度重視の検索方法であると言える。しかしながら、一回目のキーワード入力のみで精度重視の検索する方法ではユーザの多様なニーズに答えることが困難である。

3. WEB 画像検索の 2 ステップモデル

我々はキーワード画像検索の結果として提示される画像の中からユーザが選択した画像を入力として類似画像検索を行うような 2 ステップで実現される新しい WEB 画像検索モデルを提案する。図 1 にその提案モデルを示す。斜線の円がユーザ所望の画像群、白抜きの円が 1 ス

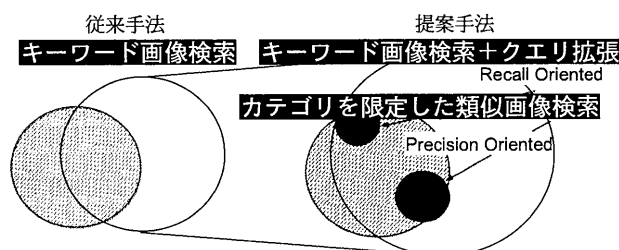


図 1: WEB 画像検索モデル

トップ目のキーワード画像検索の結果として提示される画像群、塗りつぶしの円が我々が提案する WEB 画像検索の 2 ステップ目 (類似画像検索) の検索結果を表す。

キーワード画像検索ではキーワードの多義性から、一回のキーワード入力のみでユーザのニーズを特定することは困難である。また、リンク近傍のテキストと画像との関連付けのみでは画像のメタ情報としてユーザの多様なニーズに答えることは難しい。我々の提案するモデルではキーワード画像検索の段階ではクエリ拡張等を用いて、Recall Oriented な検索を行う。また、検索結果を画像特徴量の観点でクラスタリング表示することで、類似画像検索の入力となる画像を探すことを支援するようなインタフェースを提供する。次にユーザが選択する画像からユーザのニーズを特定し、Precision Oriented な検索結果を返す。下記に我々の WEB 画像検索モデルを実現する 3 つの技術を示す。

1. クエリ拡張等を用いた “Recall Oriented” な検索
2. キーワード画像検索の結果を “クラスタリング” して表示
3. 選択された画像が所属するカテゴリのみを検索対象とした “Precision Oriented” な類似画像検索

4. 実装方法

我々は提案モデルを実用的に構築するために以下の点を考慮する。

1. 検索対象を限定することによる計算量の削減
2. 類似画像検索を導入することによる精度向上
3. 既存ユーザにも使いやすいインタフェース

計算量の削減と類似画像検索の精度向上を実現するためには、画像をカテゴリ分類することで、類似画像検索の検索対象を限定することが重要である。こうした分類を

[†]日本電信電話株式会社 NTT サイバーソリューション研究所
[‡]NTT レゾナント株式会社

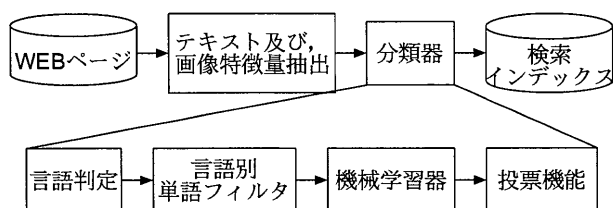


図 2: システム構成図

画像特徴量のみで分類を行うのは、精度、計算量等の面から難しい。そこで、親ページのテキスト情報や、画像特徴量等を利用することで WEB 画像の分類を試みる。本研究ではこうした WEB 画像の分類に対して様々な観点で分類し、その結果を考察する。図 2 は WEB ページを収集し、検索インデックスに登録するまでのフローを示している。まず、クローラが収集した WEB ページを入力として、テキスト情報、及び画像特徴量を抽出する。テキスト情報とは画像をリンクするタグの近傍のテキスト等を指す。次に抽出した情報を入力として分類器に与えると、カテゴリ情報を出力し、その情報を検索インデックスに保存する。

分類器では、まず言語判定を行う。言語判定後その言語に対応するフィルタを通して、機械学習器の入力となるベクトルを作成する。本研究で使用するフィルタはエントロピーと IDF を考慮し、情報量の多い上位 10000 ベクトルを用いた。そのベクトルを入力として機械学習器は親ページがどのカテゴリに属するかを判定し、カテゴリ情報を出力する。投票機能では複数の機械学習器が出力した分類カテゴリの確信度に応じてカテゴリの判定を行う。機械学習のトレーニング時には言語フィルタで出力となった単語ベクトルとその画像に人手で付与されたカテゴリ情報から学習器を生成する。

5. 検証実験

5.1 実験に利用するデータセット

本研究では実際に画像検索の検索クエリとして人気の高いクエリを中心に CJKE(中国語, 日本語, 韓国語, 英語)の言語に対して正解情報を付与したデータセットを作成した。データセットの作成方法としては、被験者に画像を提示し、その画像に対してあらかじめ用意したカテゴリに属するかどうかを判定してもらった。今回使用したカテゴリは大きく分類して(動物, 人, 乗物, 風景, アニメ, 食べ物, 家具)で、その中に階層的なカテゴリを設け、例えば“人”であれば“アイドル”, “政治家”等のサブカテゴリ情報を付与した。結果として 42895 画像にこちらが用意したカテゴリの正解を付与した。

5.2 実験方法

前述したデータセットを用いた画像の分類を行った。データセットに付与されているカテゴリは画像に対して付与したものであるため、親ページにはその画像が持つカテゴリの OR を取って親ページのカテゴリとした。分類カテゴリはなるべく直交性のある(アイドル, 乗物, 風景, 食べ物, 動物)に限定し、複数のカテゴリに属している

表 1: WEB ページ分類精度の検証 (英語)

Method	F_1 -measure(%)	
	NaiveBayes	C4.5
IMF only	34.4948%	40.3645%
SRD only	41.8955%	-%
SRD+Filter	43.3161%	44.3011%
HTML+Filter	64.8084%	64.1115%
HTML+Filter+IMF	66.8453%	64.5671%

正解は省いた。また、付与クラスには“人”という分類が存在したが、顔認識を用いる方法の精度が高いということ、人間が画像にいるという情報よりも、その人間がアイドル等の特定ドメインの画像であることが分かる方が、WEB 画像検索のニーズ上、適していると考え今回の分類には含めなかった。

5.3 実験結果

前章で示した言語判定で英語に分類されたものを対象とした検証実験を行った。分類アルゴリズムは Naive-Bayes, C4.5, SMO を使って 10-fold cross validation を行い、精度比較を行った。表 1 は学習器別に分類を行った精度の検証実験結果である。図中の SRD は画像の近傍テキスト, Filter は前章で示した単語フィルタ, IMF は画像特徴量を示している。画像特徴量としてはカラーヒストグラム, テクスチャ, 画像サイズの全 31 次元を用いた。

本検証実験から画像特徴量のみや、画像の近傍のテキストで分類するよりも、画像の親ページで分類する方が精度高く分類出来ることが分かった。また、画像特徴量とテキスト情報の両方を用いて分類することで若干の精度向上が見込まれることが分かった。結果として約 65%の精度で親ページを分類することが出来ることが確認された。

6. 考察

本稿では WEB 画像検索モデルを提案し、そのモデルを実現するための方法について述べた。また、実現方法の一つである WEB 画像分類に利用する観点別(リンク近傍のテキスト, 親ページ, 画像特徴量)に比較を行った。本検証実験結果は予備的なものであり、大規模なデータや、今回使用しなかった CJK の言語, また画像が 1 つのカテゴリに分類されることを前提にしたが、複数のカテゴリに属する場合についても検証実験をする予定である。また、実際にこのシステムを実サービスとして運用する場合、誤分類した画像をユーザが選択してしまうと、2 ステップ目で極端に Precision が下がることが考えられる。実際は Precision が一定の閾値を満たす条件下での F_1 -measure の比較を行う必要があると考えている。

参考文献

- [1] K. Kushima, M. Satoh, H. Akama, and M. Yamamuro. Integrating hierarchical classification and content-based image retrieval - imagecompass -. In *Conference on Intelligent Information Processing*, pp. 179-187, 2000.
- [2] J. R. Smith and S.-F. Chang. Visually searching the web for content. *IEEE Multimedia*, Vol. 4, No. 3, pp. 12-20, 1997.