

科学研究費助成事業 研究成果報告書

平成 27 年 4 月 30 日現在

機関番号：14401

研究種目：挑戦的萌芽研究

研究期間：2012～2014

課題番号：24650039

研究課題名(和文) ウェブスケール画像例に基づく概念間の距離

研究課題名(英文) Inter-Concept Distances Based on Web-scaled Image Instances

研究代表者

馬場口 登 (Babaguchi, Noboru)

大阪大学・工学(系)研究科(研究院)・教授

研究者番号：30156541

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：画像検索の高度化を阻害する要因の一つに、検索したい画像集合を適切なキーワードで表現するのが困難であるというインテンションギャップ問題がある。本研究では、インテンションギャップの軽減のため、Web空間に存在するタグ付き画像集合を用いてキーワードが表す概念間の距離を定義し、その性質を調べた。具体的な成果として、Web空間からタグ付き画像を収集するシステムの開発、それを用いたタグ付き画像データセットの構築、複数の視覚特徴量とタグ情報を統合した概念間距離算出手法の具体化などが挙げられる。これらの成果は、本分野に加え物体認識や画像検索などの応用分野の発展にも資するものと考えられる。

研究成果の概要(英文)：One of the major problems in keyword-based image retrieval is the intention gap between user's search intent and input keywords. To relax the intention gap, in this research, a couple of distance measures between concepts described by textual keywords were defined and experimentally analyzed, using a large scale of tagged-images. A system for automatically collecting tagged-images in the web space and storing the images into the database, and algorithms of measuring inter-concept distances were developed. These research results contribute to the advancement of object recognition and image retrieval.

研究分野：メディア情報学

キーワード：概念間距離 画像例 ウェブ画像 Word Net 画像特徴

1. 研究開始当初の背景

Web の発達やネットワークの高速化、大容量化に伴って、画像、映像、音声などマルチメディアの検索に関する技術の高度化が要請されている。

最も一般的な画像検索法であるテキスト・キーワードをクエリとする手法では、検索の意図 (インテンション) をキーワード単独あるいはキーワード列で表すのが難しいという、「インテンションギャップ」と呼ばれる問題がある。この問題に対する解決策として、検索システムに順次入力されたキーワードに対し、それらが表す概念の間に距離を与えることができれば、その距離を基に適切なクエリ候補をユーザに提示でき、インテンションギャップを克服もしくは軽減できると期待される。

近年、Web 上には日々膨大な数のタグ付き画像がアップロードされている。これらの画像は、タグ名が表す概念の画像例と捉えることができ、概念と概念の間に妥当な距離を与えるための糸口になり得ると想定していた。

2. 研究の目的

キーワードが表す概念間の距離を、Web 空間に存在する大量のタグ付き画像から定義し、時変性等の性質を実験的に調べる。そのために、Web 空間からのタグ付き画像の収集法、画像例からの視覚特徴量の抽出法、概念間距離の算出法を確立する。

3. 研究の方法

以下に述べる課題を通して研究目的の達成を図る。

- (1) Web 空間から画像例の収集
- (2) 画像例からの視覚特徴量抽出
- (3) 視覚特徴量による概念の数値的表現と概念間距離の定義
- (4) 複数種類の視覚特徴量の統合
- (5) 視覚特徴量とタグ情報の統合

以降では、本研究で定式化する画像例に基づく概念間の距離を IIBD (Image Instance Based Distance) と表記する。本研究では、単語意味に基づいて構築されたキーワード階層木 (オントロジー) である WordNet を対象に定められる、概念間の階層木に基づく距離 (Hierarchic Tree Based Distance; HTBD) に着目し、IIBD を HTBD と比較することにより IIBD の性質を調べ、またその妥当性を評価する。なお、HTBD は人間の感覚との親和性が比較的高いとされているが、上述の WordNet のようなオントロジーに収録されていない単語が表す概念は扱えない点、オントロジーに時変性がないため社会動向の変化に伴う人間の感覚の変化を捉えられない点などに課題がある。

4. 研究成果

(1) Web 空間から画像例の収集

画像共有サイト Flickr を対象に画像例を収集するシステムを構築した[学会発表④]。このシステムは、Flickr に投稿された画像を <画像例、タグ> というマルチメディアデータとして取り扱い、タグ名をキーとして画像例を収集するものである。時間経過に伴って次々に投稿される画像を追加的に収集し、投稿日時とともに蓄積する機能を有しており、Web 空間上の画像集合の変化という形で人間の感覚が持つ時変性を捉えることができる。

また、上述のシステムを用いて、400 種類の概念について各々 1000 枚程度の画像を有する、総画像枚数約 40 万程度のタグ付き画像データセットを構築した[学会発表②]。このデータセットは、概念間距離の解析に止まらず、物体認識や画像検索等にも応用可能であり、視覚情報処理分野の研究の発展に資するものと期待される。

(2) 画像例からの視覚特徴量抽出

画像には色、テクスチャ、形状など様々な属性がある。各々の属性に対応する特徴量を広く検討するために、色に関する特徴であるカラーヒストグラム、画像の局所的な形状に関する特徴量である SIFT (Scale Invariant Feature Transform)、テクスチャに関する特徴量である同時生成行列 (Gray Level Co-occurrence Matrices; GLCM)、画像の全体的な構図に関する特徴量である GIST など、複数の視覚特徴量を取り上げ、これらを画像例から抽出するシステムを開発した[学会発表③]。

(3) 視覚特徴量による概念の数値的表現と概念間距離の定義

概念を表現する画像例は多様であり、また大量に存在する。それらの画像一つ一つから視覚特徴量を抽出することにより、各概念について視覚特徴量の集合が得られる。この集合に基づいて各概念を数値的に表現する方式を開発した[学会発表③]。なお、本項の内容は単一種類の視覚特徴量のみを用いる場合を対象としたものである。複数種類の視覚特徴量を用いる方式への拡張は次項で述べる。

最も単純には、視覚特徴量の集合からその平均を計算する方式が考えられるが、平均のみで多様な画像例から抽出される視覚特徴量の集合の性質を十分に表現することは難しい。そこで開発方式では、以下の手順で視覚特徴量の分布そのものを計算し、これにより各概念を数値的に表現する。まず、各画像例から抽出された視覚特徴量を特徴空間にプロットする。次に、クラスタリング手法を

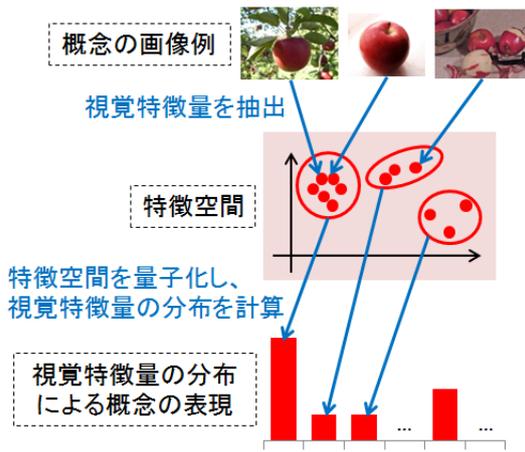


図 1：視覚特徴量による概念の数値的表現

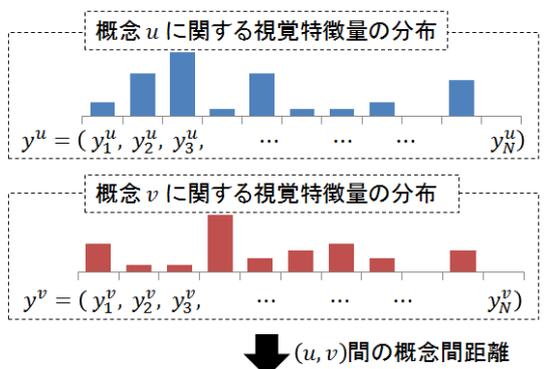


図 2：視覚特徴量分布に基づく IIBD の算出

用いて特徴空間を複数の領域に分割し、領域ごとにその内部にある視覚特徴量の数を数える。その後、各領域内の視覚特徴量の個数を画像例の総数で割ることにより、各概念を視覚特徴量の離散分布として表現する。以上の手順を図 1 に示す。

IIBD は、以上により求まる視覚特徴量分布の間の距離として定義した。距離算出のためのメトリックとしてはユークリッド距離やマハラノビス距離などが一般的であるが、分布間の距離（非類似度）を評価する場合、これらのメトリックよりも情報ダイバージェンスと呼ばれる基準の方が適している。中でも、Jensen-Shannon (JS) ダイバージェンスと呼ばれる基準は、対称性を満たすことからメトリックとして妥当であり、本研究ではこの基準を用いて IIBD を算出する方式を開発した[学会発表③]。その具体的な算出式を図 2 に示す。

(4) 複数種類の視覚特徴量の統合

上記の(3)は、単一種類の視覚特徴量のみを用いる場合の IIBD の算出法である。これを、複数種類の視覚特徴量に基づく方式へと

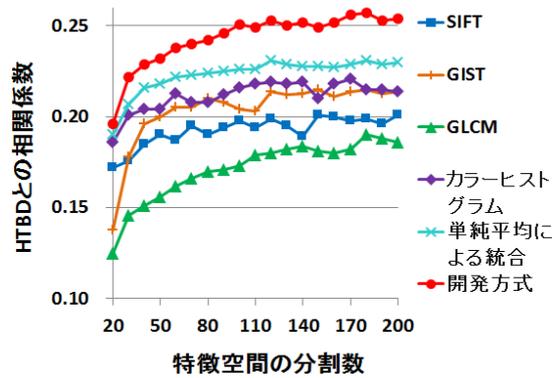


図 3：複数の視覚特徴量を用いる方式の評価

拡張した[学会発表①]。具体的な拡張方法は次の通りである。

まず、各視覚特徴量について、(3)の方式により各々を単独で用いた場合の IIBD を算出する。次に、それらの加重和（線形和）として最終的な IIBD を算出する。加重和の計算に際して、統合前の各 IIBD には、その値が小さいものほど大きい重みを与えられるように設定する。これは、人間の感覚の特性に倣ったものである。概念と概念の類似性を評価する観点には色、形状、機能など様々なものがあるが、ある概念対が何らかの観点において類似していれば、他の観点では類似していないとしても、人間は直感的にその概念対が互いに類似していると判断するものと考えられる。これに倣い、本統合法でも、算出される距離の値が小さい視覚特徴量をより重視し、どれか一種類の視覚特徴量において距離が小さく算出された場合、統合後の IIBD の値も小さくなるように重みを設定する。この重みを概念対ごとに適応的に定めるところに本方式の特色がある。

図 3 は、(1)で収集したタグ付き画像集合を用いて(2), (3), (4)の手法により IIBD を算出し、その結果を HTBD と比較したものである。図中の縦軸は算出した IIBD と HTBD の相関係数を、横軸は(3)で述べた特徴空間分割における分割数を表している。また、「単純平均による統合」の項目は、上述の加重和の計算に際して任意の項に等価な重みを固定的に与えた場合の結果を表している。図 3 の結果から、複数種類の視覚特徴量の統合によって、より妥当性の高い概念間距離を定義できることが分かった。また、統合のための加重和の計算に際して、固定的な重みを与える方式よりも、重みを適応的に与える(4)の方式の方が、算出される概念間距離の妥当性の向上に寄与することが分かった。

本研究課題の着手以降、国内外の他の研究機関においても画像例に基づく概念間距離の算出法が研究され始めているが、人間の感覚の特性に倣って複数種類の視覚特徴量を適応的に統合した方式は他にない。このような方式の有効性を示したことは本研究課題の成果の一つであり、後発の類似研究や画像

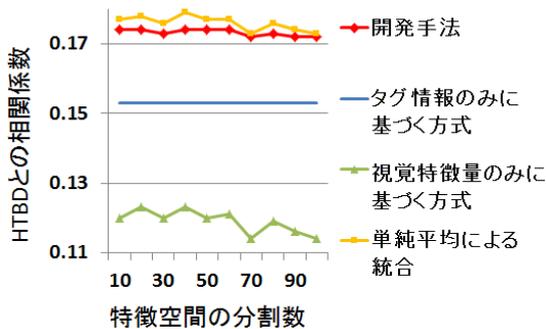


図 4：タグ情報を統合した方式の評価

検索などの関連分野に対し重要な示唆を与えるものである。

(5) 視覚特徴量とタグ情報の統合

ある概念の画像例に対し、その概念を表すタグに加え別のタグが同時に付与されている場合、その別タグもまた注目概念を特徴づける情報となり得る。概念には、例えば「夏休み」などのように、視覚情報だけで表現することが必ずしも容易でないものも存在する。そのような概念について IIBD を算出する際には、上記のような別タグこそが重要な情報となる。そこで、画像例に付与されているタグに基づいて概念間の距離を算出する方式を開発し、これを(4)までで確立した方式と統合することにより、より妥当性の高い IIBD の算出方式を開発した[学会発表②]。

以降では、同じ画像例に二つのタグが同時に付与されているとき、それらのタグ（もしくはそれらのタグが表す概念）は「共起」していると表現する。タグ情報に基づく概念間距離の算出に際し、まず各概念が他の概念との程度共起しているか、すなわち共起頻度をタグ付き画像集合に基づいて求め、その共起頻度のベクトルにより各概念を数値的に表現する。次に、共起頻度ベクトルの間の距離をコサイン距離として求め、これをタグ情報に基づく概念間の距離とみなす。

タグ情報に基づく概念間距離と(4)までの方式による概念間距離の統合は、各々の距離の加重和として定義した。この方式では、視覚情報による表現が容易な概念ほど(4)の方式による概念間距離を重視し、そうでない概念ほどタグ情報に基づく概念間距離を重視するように加重和計算の際の重みが設定される。視覚情報による概念の表現は、画像例から抽出される視覚特徴量の集合が多様性に富んでいるほど困難であると考えられる。このことから、本方式では、(3)において計算される視覚特徴量分布のエントロピーが大きいほど視覚特徴量による概念の表現は困難であると判断する。視覚特徴量による概念表現の容易性に基づいて適応的な重みを与える点に本方式の特色がある。

図 4 は、(1)で収集したタグ付き画像集合を用いて本方式による IIBD を算出し、その

結果を HTBD と比較したものである。図 3 と同様、縦軸は算出した IIBD と HTBD の相関係数を、横軸は(3)で述べた特徴空間分割における分割数を表している。また、「単純平均による統合」は、上述の加重和の計算に際して任意の項に等価な重みを固定的に与えた場合の結果を表している。図 4 の結果から、視覚特徴量とタグ情報の統合によって、より妥当性の高い概念間距離を定義できることが分かった。しかし、開発方式は、統合のための加重和の計算において固定的な重みを与える方式よりも妥当性の面でやや劣る結果となった。これは、タグ情報のみに基づく方式に比べ視覚特徴量のみに基づく方式の妥当性が有意に劣っていることが原因と推察される。この点について詳細な考察を加えた結果、視覚特徴量に基づく概念間距離の妥当性を低下させる主要因は、タグが表す概念と画像例の意味内容が一致していない「ジャンク画像」にあることが分かった。以上の結果から、ジャンク画像除去の重要性が示唆され、今後、そのための手法に関する検討が急務である。

5. 主な発表論文等

[学会発表] (計 4 件)

- ① Kazuaki Nakamura and Noboru Babaguchi: “Inter-Concept Distance Measurement with Adaptively Weighted Multiple Visual Features,” Lecture Notes in Computer Science, Vol. 9010, pp. 56-70, 2015. (査読有)
DOI: 10.1007/978-3-319-16634-6_5
- ② 増広のぞみ, 中村和晃, 馬場口登: “画像情報とテキスト情報の重み付き統合による概念間距離尺度の改善,” 電子情報通信学会技術研究報告, PRMU 2013-103, MVE2013-44, pp. 131-136, 大阪大学, 2014 年 1 月 23 日.
- ③ 中村和晃, 大利綾香, 馬場口登: “画像特徴の適応的選択による画像例ベース概念間距離の算出,” 電子情報通信学会技術研究報告, IE2013-10, PRMU 2013-3, MI2013-3, pp. 13-18, 愛知工業大学, 2013 年 5 月 24 日.
- ④ 大利綾香, 中村和晃, 馬場口登: “多数の Flickr タグを用いた画像例ベース概念間距離の解析,” 電子情報通信学会 2013 年総合大会, D-12-17, p. 110, 岐阜大学, 2013 年 3 月 19 日.

6. 研究組織

(1) 研究代表者

馬場口 登 (BABAGUCHI, Noboru)
大阪大学・大学院工学研究科・教授
研究者番号: 30156541