

下位語を利用した単語概念が持つ視覚的多様性の数値化

松尾 賢治[†] 川野 悠[‡] 大島 裕明[‡] 田中 克己[‡]

[†] 株式会社 KDDI 研究所 〒356-8502 埼玉県ふじみ野市大原 2-1-15

[‡] 京都大学大学院 情報学研究科 〒606-8501 京都市左京区吉田本町 36-1

E-mail: [†] matsuo@kddilabs.jp [‡] {kawano, ohshima}@dl.kuis.kyoto-u.ac.jp, ktanaka@i.kyoto-u.ac.jp

あらまし 本研究では、ある与えられた語に対する視覚的多様性を数値化する手法を提案する。「熱帯魚」はその種類によって姿形や模様等が大きく異なり、視覚的に多様であると言えるが、「カエル」は形状や質感が比較的似ており、「熱帯魚」に比べて視覚的多様性が低いと考えられる。本手法では、まず Web 検索エンジンが保有する索引情報から与えられた語に関する下位語を高速に見付け出す。Web 画像検索を利用して、発見した下位語に適合する画像を収集する。収集された画像を特徴量空間上でクラスタリングした結果と下位語による分類との一致度合いから、単語概念が持つ視覚的多様度を算出する。

キーワード 視覚的多様性、下位語抽出、クラスタリング、エントロピー

1. はじめに

近年の画像認識技術の向上により、顔認識や指紋認証など特定の対象を認識する技術が既に実用化されている。一方、対象を特定せず、「人」、「自動車」や「飛行機」等の画像の一般的な名称を認識することを目的とした一般物体認識は、現時点でも困難な課題の一つであり、研究への取り組みが盛んである。一般物体認識では、認識させたい対象が写っている大量の画像データから、その対象に固有の視覚的特徴を学習するアプローチが取られることが多い。新たに認識できる対象を付加的に増やすことができるため、より汎用的な認識が可能となる。この場合、単語概念が持つ視覚的な多様性を考慮に入れることは極めて重要である。例えば、「カエル」はその種類に関わらず形状や質感が比較的似ており、安定した特徴が得られると考えられる。しかし、「熱帯魚」は種類によって体形や模様が大きく異なると考えられ、共通して持ちうる顕著な特徴を学習することは難しい。一方、「熱帯魚」ではなく、下位語に相当する「エンゼルフィッシュ」や「グッピー」、「アロワナ」等であれば、共通する視覚的特徴がより明確となり、それぞれに特有の顕著な特徴量が得られる可能性は高い。このことから、単語概念が持つ視覚的多様性は認識精度にも大きな影響を与えると考えられる。

本研究では、視覚的な多様性が認識精度に影響すると仮定し、一般的な名称からその言語概念が持つ視覚的多様性を数値化する手法を提案する。一般的な名称で指定されたある対象について、画像特徴に基づいたクラスタリング結果が、言語概念上の下位語による画像の分類にどの程度一致しているかで、最終的な視覚的多様度を算出する。前述の例では、「熱帯魚」は視覚的多様性が高く、「カエル」は視覚的多様性が低いと考えた。図 1 に示す

ように、画像特徴量を用いたクラスタリングを行うことで、視覚的多様性の高い「熱帯魚」は、見た目が大きく異なる「エンゼルフィッシュ」や「グッピー」、「アロワナ」等にそれぞれ画像が分類されることが想定される。しかしながら、視覚的多様性の低い「カエル」は、「アマガエル」や「ウシガエル」等の視覚的特徴が似通っていることから、画像特徴量を用いたクラスタリングを行ったとしても、その通りには分類されない可能性が高い。したがって、クラスタリングの結果と言語概念上の分類とが一致していれば視覚的多様性は高く、一致していなければ低いと考えられる。

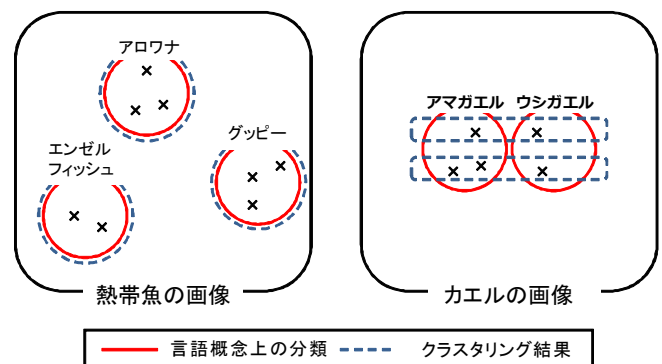


図 1 視覚的多様性によるクラスタリング結果の違い

視覚的多様性の数値化により、一般物体認識の認識精度を向上させる学習が期待できる。すなわち、前述の例では、「カエル」のように視覚的多様性が低いと判定されれば、下位語の「アマガエル」や「ウシガエル」等を別々に学習させて両者間で相互に誤認識を頻出させることを回避するため、元の「カエル」で学習させる。逆に、「熱帯魚」のように視覚的多様性が高いと判定されれば、よ

り下位の階層にある「エンゼルフッシュ」や「グッピー」、「アロワナ」等を区別し、それぞれに固有の特徴を学習させて、認識精度を向上させることを想定している。

2. 関連研究

柳井らはWeb上から収集した画像を用いて単語概念と視覚的特徴との結び付きの強さを評価するための指標として画像領域エントロピーを提案し[1, 2], 川久保らはこれを改良した[3]. まず単語概念についてWeb上から自動収集した画像を領域分割し, 得られた部分領域の中から, 単語概念に関係する正領域を推定する. 正領域の特徴量の分布のエントロピーを画像領域エントロピーとした. 単語概念には, 「ライオン」や「山」のように視覚的特徴に直接対応しているものがある一方, 「動物」「地形」のように視覚的特徴と直接結び付けることが困難であること, また, 画像認識の分類カテゴリを設定する場合, 前者に示す画像特徴と結び付きが強い単語概念であることが望まれることを示した. しかしながら, 結果は領域分割の精度に影響を受ける恐れがある.

一方, 本研究も同様に, 画像認識の学習の際に望ましい名称を決定することを目的とするが, 画像特徴の分布のエントロピーではなく, 下位語による分類と画像特徴に基づく画像集合クラスタリングの結果の一致に関するエントロピーから単語概念が持つ視覚的多様性を求める. 特別な手動操作を必要とせずWeb上のテキストおよび画像データだけを使って, オンデマンドかつ自動的に, 言語概念が持つ視覚的多様度を算出できる点が特徴である.

3. 提案概要

以下の手順で視覚的多様度の算出を試みる. 一般的な名称を指定するだけで, Webに格納されたテキストおよび画像を利用して, 結果を出力できることが特徴である.

- [1] 指定された単語の下位語をWebから発見する.
- [2] 各下位語をクエリとしてWeb画像検索を行う.
- [3] 収集画像を画像特徴に基づきクラスタリングする.
- [4] クラスタリング結果と[1]の下位語による分類を比較し, 一致度合いから視覚的多様性を数値化する.

図2に視覚的多様度の算出のフローを図示した. 手順[1]については, 既に言語間の階層関係が体系付けられたWordnet¹等の概念辞典を使用することなく, Web検索エンジンが保有する索引情報から高速に下位語を発見するための手法を新たに提案する. これにより, 分類学上は

下位語の定義が困難な語についても, Web上の記述で自然に定義された下位語を柔軟に発見できる可能性がある. また, Web検索エンジンが保有する索引情報の特性上, より一般的で注目されている順に下位語が抽出できると考えられる. 手順[1]の詳細を4.1で説明する. 手順[2]については, 既に実用化されているWeb画像検索システムを利用する. 手順[3]については, 既存の階層的クラスタリング手法を適用する. 与えられた画像集合の中から距離の近いもの同士を階層的に統合する. この結果得られるデンドログラムに対して, どの程度離れた特徴量同士を同一クラスとみなすか, しきい値を指定して, クラスタリングの詳細度を変化させることができる. 手順[3]の詳細を4.2および4.3で説明する. 手順[4]については, 既存のクラスタリング結果の評価手法を使用する. 手順[1]で得られた下位語による分類を正解とした場合の一致度合いをエントロピーとして算出する. 手順[4]の詳細を4.4で説明する.

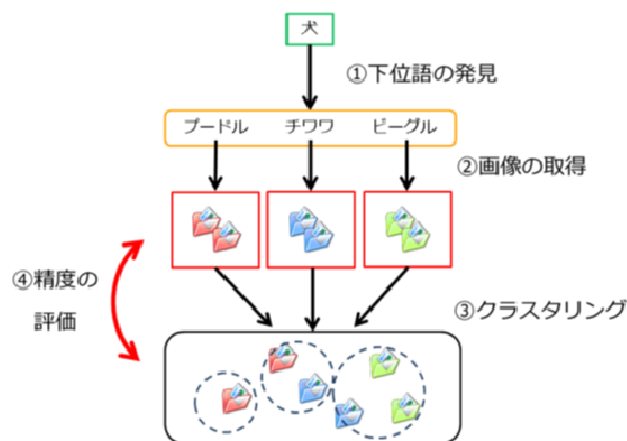


図2 視覚的多様度の算出のフロー

4. 詳細手順

提案手法の各手順の詳細について次節に述べる.

4.1. Web 検索エンジンを用いた下位語抽出

本節では, Web 検索エンジンが保有する索引情報を利用して高速に下位語を発見する手順について説明する. 下位語とはある語に対する下位概念を表す語である. 例えば, 「犬」に対して「ビーグル」や「小型犬」といった語は, 「犬」をより具体的に示すものであり, 下位語であると言える. しかし, 下位語の中でも「ビーグル」と「小型犬」では概念の深さが異なる. 「ビーグル」と並列に考えられる語として「チワワ」や「プードル」があるが, 「小型犬」はこれらの語よりも抽象的であり, 上位概念にあると言える. このように, 語の概念の深さを考えた

¹ <http://nlpwww.nict.go.jp/wn-ja/>

場合に、並列に並べられるものを同じ概念階層に存在すると呼ぶことにする。

視覚的多様性を数値化する際には、同じ概念階層に存在する下位語を発見することが重要である。「犬」の下位語として、「ビーグル」「チワワ」「プードル」「小型犬」が得られたとする。これら4語についてそれぞれ画像を取得し、クラスタリングを行う場合、一般的には「チワワ」は「小型犬」に包含される関係にあると考えられるため、「小型犬」をクエリとして検索された画像の中で、真に「チワワ」が写っている画像をいずれに分類するのが正しいか、クラスタリングの基準に曖昧性が生まれる。本手順では、基準が曖昧とならないよう、同じ概念階層に存在する下位語集合を発見することを特徴とする。その実現のため、「などの」という言語パターンに注目する。また、発見した下位語の重要度を算出し、対象となる語を表す画像を取得する上で欠かせない下位語から順に発見することを可能にした。

入力として語 t を与えることで、同じ概念階層に存在する下位語集合族 $\{H_i\}$ が出力として得られる。 $\{H_i\}$ を取得する手順は以下の通りである。

- (1) クエリ「“などの AND t ”」で Web のテキスト検索を行う。ダブルクォーテーションで囲っているのはフレーズ検索であることを示している。
- (2) 検索結果のタイトルとスニペットから「などの t 」というパターンを含んだ文を抽出する。
- (3) 「などの t 」の前に現れる一つまたは複数の下位語候補を抽出し、まとめて下位語候補の集合とする。
- (4) 共通する下位語候補を含んだ集合同士を統合する。最終的に得られた集合 H_i の集まりを $\{H_i\}$ とする。
- (5) 各下位語候補に重要度を表すスコアを与える。

語 t が「犬」である場合について図3を使って説明する。手順(1)と(2)では、「などの犬」という言語パターンを含む表現を取得する。もし Web 検索エンジンが保有する索引情報に「チワワやビーグルなどの犬」という表現が得られた場合、「チワワ」と「ビーグル」は「犬」の下位語であり、かつ、この2語を同じ概念階層に存在する下位語である可能性が高い。そこで手順(3)では、「チワワ」と「ビーグル」を「犬」の下位語候補として抽出する。

このように、「などの」という言語パターンの前に並列に現れる語同士は同じ概念階層に存在する下位語候補である場合が多い。このように、言語パターンを用いて下位語を取得する手法は、Hearst による提案[4]などが挙げられるが、Web 検索を用いることによってオンデマンド

に可能になった。また、Web 上のリソースを用いた下位語取得として、HTML の構造を用いる手法[5]や、共起頻度を用いる手法[6]などが提案されている。本手法は言語パターンのみを用いる簡単なものであるが、ある程度の精度を保ち、かつ、たかだか100件程度の検索結果を用いるだけで十分な量の同一概念階層に存在する下位語候補の集合を発見できる手法である。

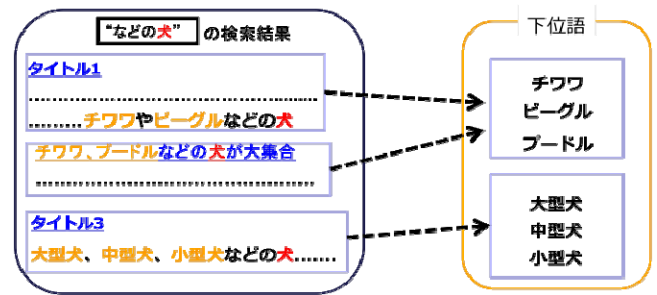


図3 Web 検索エンジンを用いた下位語発見

ここで、与えられた表現から下位語候補の区切りを推定する手順について詳細を説明する。例えば、「チワワやビーグルなどの犬」と「チワワやプードル、ビーグルなどの犬」という表現が与えられたとする。この場合、「や」「,」は下位語候補同士を区切るために使われており、抽出された表現を右から左へ探索することで、これらの区切り文字から「プードル」と「ビーグル」は正しく抽出することができる。次に、「チワワ」の直前には区切り文字がないため、表現中に現れる文字の頻度に基づいて推定を行う。正しい下位語候補は「チワワ」であるため、「チワワ」の直前に出現する文字は一般的にスパースである。そこで、表現を右から左へ探索し、Web 検索エンジンが保有する索引情報の中で、文字の出現頻度の差が最も大きい部分を下位語候補の区切りであると推定する。このように区切りを判定することで、形態素解析の精度に影響を受けずに下位語候補を発見することができる。

抽出した下位語候補の集合のうち、 $\{\text{チワワ, ビーグル}\}$ と $\{\text{チワワ, プードル}\}$ には「チワワ」という下位語候補が共通して含まれている。そこで、手順(4)では $\{\text{チワワ, ビーグル}\}$ と $\{\text{チワワ, プードル}\}$ を統合し、 $\{\text{チワワ, ビーグル, プードル}\}$ という集合に置き換える。これは、ある集合に含まれている下位語候補は相互に同じ概念階層に存在することから、集合間で共通の下位語候補が含まれていれば、統合された集合に含まれる下位語候補もまた全て同一概念階層に存在すると考えられるためである。このようにして統合を収束するまで順次繰り返し、最終的に生成された集合 H_i の集まりが出力の下位

語集合族 $\{H_i\}$ である。

最後に、各下位語候補に重要度に応じたスコアの付与を行うが、下位語候補の重要度には様々な定義が考えられる。手順(5)では、下位語候補 h の位語としての確からしさを表す信頼度 $Certainty(h)$ を次式で定義する。

$$Certainty(h) = freq(h) \cdot \sum_{s \in S(h)} cooccur(h, s) \quad (1)$$

ただし、 h と同じ概念階層に存在する下位語の中で、 h 以外の他の下位語を集めた集合を $S(h)$ とする。 $freq(h)$ は、「などの」というパターンを含む表現に h が出現した回数を表す。 $cooccur(h, s)$ は「などの」というパターンを含む表現に h と $S(h)$ に含まれる下位語候補 s が共起した回数を表す。例えば、「チワワやビーグルなどの犬」と「チワワやプードル、ビーグルなどの犬」という表現が与えられた場合、 $freq(\text{チワワ})$ は2であり、 $\sum_s cooccur(\text{チワワ}, s)$ は3となる。 $freq(h)$ が高いことはWeb検索エンジンが保有する索引情報の中で h が下位語候補として頻繁に現れていることを意味する。また、 $\sum_s cooccur(h, s)$ が高いことは、 $S(h)$ に含まれる他の下位語候補から h が同一概念階層に存在することを支持される度合いが高いことを意味する。よって、 $Certainty(h)$ が大きければ、 h は下位語として確からしいと言える。最終的には、 $Certainty(h)$ が最大の下位語候補が含まれる集合 H_i を特定し、同 H_i 内に含まれる下位語候補を全て下位語として出力する。

本手法は、下位語を含む表現の取得にWeb検索を1度しか行わないため、極めて高速に下位語を検索することができる。また、下位語の抽出に形態素解析を使っていないため、MeCab²や茶釜³などの形態素解析器を導入せず、簡単にシステムを使うことができる。

4.2. 画像特徴量

画像は、本来3次元空間中の被写体を2次元の平面で切出したものであり、被写体の姿勢の変化、隠蔽等により、見た目は容易に変動する。このような見た目の変動による影響を避けるため、画像認識では特徴量変換により、見た目の変動に影響を受けず、被写体が固有に持つ不変の特徴量を画像から抽出し、特徴空間上で処理するのが一般的である。一般物体認識では、特に被写体の姿勢の変化、隠蔽等に強いと考えられている、局所特徴量の分布パターンを使ったbag-of-featuresのフレームワークが近年高い認識精度を示している。Loweらは回転およびスケール変化に不変な局所特徴量SIFTを提案した[7]。

SivicおよびCsurkaらはSIFT特徴ベクトルをベクトル量子化し、画像を量子化代表ベクトルの集合であるvisual wordsの出現頻度で表現した[8, 9]。Nisterらは、k-meansを繰り返し適用し、図4に示すようなvisual wordsに階層的な構造を持たせたvocabulary treeを提案した。同時に、tf-idfの考えに基づいた局所特徴量の出現頻度に基づく重みを各ノードに設定し、対応するクラスターが持つ空間の粒度を近似的に再現している[10]。

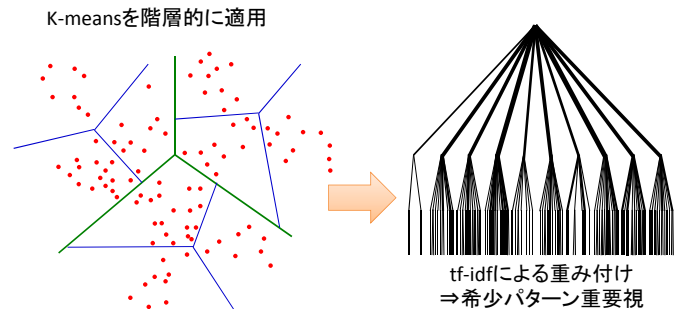


図4 階層的な構造を持たせた Vocabulary Tree

提案手法でも、bag-of-featuresのアプローチで画像を特徴量に変換することによって、単語概念が持つ視覚的多様性を数値化する際に妨げとなる、被写体の姿勢の変化、隠蔽等による見た目の変動を抑制し、被写体が固有に持つ不変の特徴量を抽出することを目指した。具体的には、局所特徴量としてSIFTを選択し、グレースケール画像からDoG処理に基づいて特徴点を抽出した上で、128次元の局所特徴量を抽出する。また、特徴空間を k 個に分割するk-meansを階層的に p 回適用し、 $(k^p-1)/(k-1)-1$ 個の量子化代表ベクトルを含むVocabulary Treeにより、画像から抽出された局所特徴量をベクトル量子化して、その分布パターンを特徴量化する。

4.3. クラスタリング

画像集合を画像特徴が似ているまとまりに分割するため、特徴空間上でクラスタリングを行う。提案手法では階層的クラスタリングを導入し、画像集合に含まれる二画像間の距離を総当たりで計測して、距離の近い画像から順に同じクラスターへと統合を繰り返す。この操作により、図5に例示したデンドログラムと呼ばれる樹状の分類構造を構成する。図は1～15の15枚の画像が階層クラスタリングされた例である。デンドログラム内で二つのクラスターが統合される点の横軸における高さは、両クラスター間の距離に相当する。良好なクラスタリング結果を得るためには、クラスターが統合される度に毎回クラスター

² <http://mecab.sourceforge.net/>

³ <http://chasen.naist.jp/hiki/ChaSen/>

間の距離を総当たりで再計測する必要がある。統合後のクラスタ間の距離の再定義には、最短距離法、最長距離法、メディアン法、重心法、群平均法、可変法、ウォード法と様々な手法が提案されている。また、距離に関するしきい値を設定し、デンドログラムをしきい値に相当する高さの垂直断面で切ることにより、いくつかのクラスタが得られる。分割数は断面に交わる線の数に相当する。例えば、下図にてしきい値 10 を設定した場合、{1, 6, 7, 2, 3}, {8, 11, 12, 13}, {4, 5}, {9, 10, 15, 14} の 4 つのクラスタが得られる。

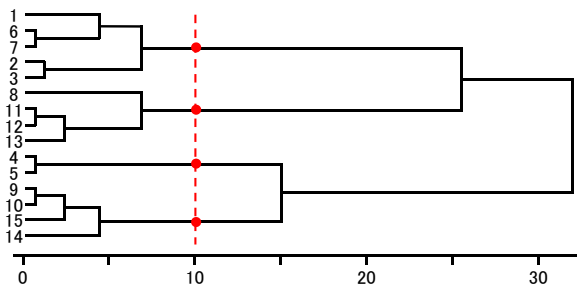


図5 階層的クラスタリングにより作成されたデンドログラム

4.4. クラスタリング結果の評価

正解集合が用意されているか、クラスタの数が与えられているか等、条件によって評価方法がいくつか考えられる。今回は、下位語による分類を正解集合とみなせるため、正解集合が存在し、クラスタ数が既知である場合の評価値の一つである、エントロピーにより評価を行う。クラスタリングの結果を $C=\{C_1, C_2, \dots, C_k\}$ 、正解を $A=\{A_1, A_2, \dots, A_k\}$ 、 X_{ij} を $|C_i \cap A_j|$ 、すなわち C_i と A_j に共通して含まれる画像の個数とする。なお、クラスタリング結果の分割数 k は正解集合と同じ数となるようにしきい値を調整する。あるクラスタ C_i に対するエントロピー E_i を次式で定義する。

$$E_i = -\sum_{j=1}^k P(A_j | C_i) \log_k P(A_j | C_i) = -\sum_{j=1}^k \frac{X_{ij}}{|C_i|} \log_k \frac{X_{ij}}{|C_i|} \quad (2)$$

クラスタに含まれる画像の数により、重み付き平均を取り、全体のエントロピー E を次式で定義する。

$$E = \sum_{i=1}^k \frac{|C_i|}{N} E_i \quad (3)$$

ただし、 N は全画像数である。 E の値は 0 から 1 の値をとり、値が低いほどクラスタリング結果が正解集合に近いことを示す。

5. 実験

提案手法の有効性を評価するため実験を行った。

5.1. 実験条件

本節では、一般的な名称からその単語概念が持つ視覚的多様性を数値化する手法の評価を行う。多くの一般物体認識では、比較的上位の階層にあるカテゴリ名称を扱うものが多い。例えば、画像認識のコンテストである PASCAL VOC 2010⁴では「人」、「自動車」、「飛行機」等の 20 種類の物体を対象として認識精度が競われている。そこで、視覚的多様性の大きいと考えられる「犬」と「熱帯魚」、視覚的多様性の小さいと考えられる「カエル」に対し、提案手法によって数値を算出し、どのような結果が得られるのかを確かめた。

提案手法は Web に格納されたテキストおよび画像を利用する。下位語を含む表現の取得には Yahoo! Search BOSS API⁵を使用し、検索結果の上位 50 件のタイトルとスニペットを対象とした。Web 画像検索には Google 画像検索⁶ならびに Yahoo! Japan によって提供されている API⁷を使用し、検索結果の上位から 10 枚の画像を取得した。加藤らは、Google 画像検索において実在する物体の名称をクエリとした場合の、検索結果上位 20 件および 100 件に含まれる画像の適合率を調査した[11]。表 1 に加藤らの調査結果を示す。上位 100 件では急速に低下しているが、上位 20 件までであれば、80~100%程度の高い確率で適合する画像が含まれる。本実験ではさらに高い適合率が期待し、上位 10 件の画像を実験に用いた。

表 1 クエリが具体物である場合の Google 画像検索の適合率

クエリ	上位 20 件	上位 100 件
moon	1.00	0.82
bird	0.95	0.68
rose	1.00	0.55
butterfly	0.80	0.70
bear	0.95	0.75
cat	1.00	0.86
dog	1.00	0.76
pigeon	1.00	0.77
iPod	1.00	0.79
headphone	1.00	0.77

収集した画像集合から抽出された全 SIFT 特徴量を $k=10$ の k -means を適用して 3 階層のクラスタリングを行い、1,110 個のノードを持つ Vocabulary Tree を構築した。

⁴ <http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2010/>

⁵ <http://developer.yahoo.com/search/boss/>

⁶ <http://www.google.co.jp/>

⁷ <http://developer.yahoo.co.jp/webapi/search/>

各ノードに対応する量子化代表ベクトルを用いて、画像を 1,110 次元の頻度分布へと変換した。

階層的クラスタリングでは、距離の尺度にユークリッド平方距離を用いて、クラスタ間距離の再定義にワード法を適用した。階層的クラスタリングの結果と正解画像集合間のエントロピーを用いて評価を行う。エントロピーが大きいほど、正解集合とは違ったクラスタリングが行われていることを意味するため、視覚的多様性が低く、下位語による分類ほど詳細に区別して分ける必要は無い、と考えられる。本実験では、この仮説が正しいのかを検証する。

5.2. 下位語発見の精度

視覚的多様性の算出手法を検証する前に、どの程度の精度で下位語を発見できているのかを示す。表 2 は「熱帯魚」、「カエル」、「犬」に対して、下位語発見手法を用いて得られた結果である。「熱帯魚」の中に含まれる「エイ」や「海水魚」などは誤りであるものの、全体的に見て概ね正しい下位語が抽出されている。Web における表記ゆれや誤った記述により、本来実在しない下位語が抽出される恐れがある。実際に、「カエル」の中に含まれる「ヤマオカガエル」や「モリオアオガエル」は実在しない可能性が高い。このような下位語としてふさわしくない語は、発見した下位語をクエリとして再度 Web 検索を行い、ヒットしたページ数が極端に少ないか否かを判定することによって、取り除くことができると考えられる。

表 2 下位語発見の結果

クエリ	下位語
熱帯魚	ディスカス, フラワーホーン, アロワナ, ×エイ, ×海水魚, エンドリケリー, エンゼルフィッシュ, グッピー, テトラ, プレコ
カエル	アマガエル, ニホンアカガエル, ×ヤマオカガエル, シュレーゲルアオガエル, ヒキガエル, トノサマガエル, ×モリオアオガエル, トウキョウダルマガエル, ヤマアカガエル, アズマヒキガエル
犬	ビーグル, ラブラドルレトリバー, 土佐犬, ボーダーコリー, テリア, ダックスフンド, ブルドック, ダルメシアン

5.3. 視覚的多様性の評価

視覚的多様性の評価は抽出された下位語を基に画像集合を取得し、これを言語概念上の分類による正解画像集合と考える。収集した全ての画像をクラスタリングした

結果と正解画像集合間のエントロピーを比較することで行う。まずは下位語発見によって正しく得られた語を対象に、Google 画像検索を利用して画像を収集することとする。つまり、「熱帯魚」における「エイ」と「海水魚」、また、「カエル」における「ヤマオカガエル」と「モリオアオガエル」は除外した。この結果、「熱帯魚」、「カエル」、「犬」共に 8 つの下位語を有することになる。表 3 に各クエリのエントロピーを示す。この表より、エントロピーの最小値が最も小さいのは「熱帯魚」であり、最も大きかったのは「カエル」であることがわかる。この結果は我々の仮説に反していない。「カエル」という単語概念は、画像特徴量によるクラスタリング結果と言語概念上の分類とが異なり、言語概念で分類されている程の多様性を視覚的には持っていないと言える。また、誤って抽出した下位語を含めた場合のエントロピーも合わせて示した。誤った下位語は視覚的多様性の数値化に大きく影響を及ぼすことが確認できる。具体的な解決策とその有効性検証は今後の課題である。

表 3 各クエリのエントロピー

クエリ	下位語数	エントロピー
犬	8	0.8128
熱帯魚	8	0.7045
カエル	8	0.8129
熱帯魚 (誤り含む)	10	0.6835
カエル (誤り含む)	10	0.7346

その他の一般的な名称について、提案手法により視覚的多様度を解析した結果を、表 4 に示す。Google 画像検索により収集した画像を使用し、エントロピーの大きい順に並べた。実際に姿形の多様性が高いと考えられる「楽器」のエントロピーが比較的小さな値を示すことが確認できる。また、僅かではあるが「トヨタ車」は「ホンダ車」よりエントロピーが小さい。これは、今回下位語として抽出された代表的な「トヨタ車」のラインナップは、視覚的により多様であったと想像できる。同様に、「ぶどう」は「梨」よりエントロピーが小さく、視覚的な多様性が高いと考えられる。

表 4 その他の名称のエントロピー

クエリ	下位語数	エントロピー
ホンダ車	6	0.8486
トヨタ車	6	0.8655
梨	7	0.7836
ぶどう	11	0.7207
楽器	13	0.6875

しかしながら，提案手法により発見された下位語の数とエントロピーの値には負の相関が見られる．すなわち，表 4 では下位語の数が多いほどエントロピーが小さい傾向を示している．下位語の数に依存せず，各語のエントロピーを同一の条件で比較する必要があると考えられる．そこで，発見された下位語から無作為に 6 件を選択し，一定の下位語数に制限した上で，エントロピーを再測定した．結果を表 5 に示す．主観にもよるが，下位語の数を一定に制限することにより，一般的な感覚に近い結果を示すことが確認できる．すなわち，多様性が低いと考えられる「梨」，「カエル」のエントロピーは大きな値を示し，逆に視覚的多様性が高いと考えられる「熱帯魚」のエントロピーは小さな値を示すようになった．なお，表 5 は Yahoo API により収集した画像であり，提案手法をシステムに組み込み，与えられた語からオンデマンドで自動的に算出したエントロピーの値である．

表 5 下位語を一定数にした場合の各名称のエントロピー

クエリ	下位語数	エントロピー
ホンダ車	6 (6 語中)	0.8658
梨	6 (7 語中)	0.8429
カエル	6 (10 語中)	0.8191
犬	6 (8 語中)	0.7899
楽器	6 (13 語中)	0.7417
ぶどう	6 (11 語中)	0.7368
トヨタ車	6 (6 語中)	0.7303
熱帯魚	6 (10 語中)	0.7165

6. まとめ

本研究では，ある名称からその単語概念が持つ視覚的多様性を数値化する手法を提案した．ある名称の下位語を取得し，各下位語による Web 画像検索の結果を言語概念上の正解分類として，画像特徴に基づくクラスタリングの結果と比較することによって，その一意度合いから視覚的多様度を算出した．実験により提案手法の有効性を確認した．一例として，エントロピーが大きい値を示した「カエル」は，画像特徴量によるクラスタリング結果が言語概念上の分類に一致せず，視覚的には言語概念で分類されている程の多様性は無く，事前の想定通り多様性が低いと判定できることを確認した．これらの一連の処理は，Web に格納されたテキストおよび画像だけを利用して，完結していることが特徴である．

視覚的多様性の数値化により，一般物体認識の認識精度を向上させる学習が期待できる．ある名称の視覚的多様性が低いと判定されれば，下位語を別々に学習さるの

ではなく，元の名称で学習させる．逆に，視覚的多様性が高いと判定されれば，各下位語に固有の特徴を学習させる．算出された視覚的多様度に基づいた学習手順の詳細化およびその有効性検証は，今後の課題である．

文 献

- [1] 柳井啓司, K. Barnard, “一般物体認識のための単語概念の視覚性の分析,” 情処学論, vol. 48, no. SIG10(CVIM17), pp. 88-97, 2007.
- [2] K. Yanai and K. Barnard, “Image region entropy: A measure of “visualness” of web images associated with one concept,” Proc. of ACM International Conference Multimedia, pp. 420-423, 2005.
- [3] 川久保秀敏, 柳井啓司, “単語概念の視覚性と地理分布の関係性の分析,” 信学論 D, vol. J93-D, no. 8, pp. 1417-1428, 2010.
- [4] M. A. Hearst, “Automatic Acquisition of Hyponyms from Large Text Corpora,” Proc. of the 14th International Conference on Computational Linguistics, pp. 539-545, 1992.
- [5] 新里圭司, 鳥澤健太郎, “HTML 文書からの単語間の上位下位関係の自動獲得,” 自然言語処理, vol. 12, no. 1, pp. 125-150, 2005.
- [6] 大石康智, 伊藤克亘, 武田一哉, 藤井敦, “単語の共起関係と構文情報を利用した単語階層関係の統計的自動識別,” 情報処理学会研究報告 音声言語情報処理, vol. 2006, no. 40, pp. 25-30, 2006.
- [7] D. G. Lowe, “Object recognition from local scale invariant features,” Proc. of IEEE International Conference on Computer Vision, pp. 1150-1157, 1999.
- [8] J. Sivic, and A. Zisserman, “Video Google: A Text Retrieval Approach to Object Matching in Videos,” Proc. of IEEE International Conference on Computer Vision, pp. 1470-1477, 2003.
- [9] G. Csurka, C. Bray, C. Dance, and L. Fan, “Visual categorization with bags of keypoints,” in Proc. of European Conference on Computer Vision Workshop on Statistical Learning in Computer Vision, pp. 59-74, 2004.
- [10] D. Nister and H. Stewenius, “Scalable Recognition with a Vocabulary Tree,” Proc. of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2161-2168, 2006.
- [11] 加藤誠, 大島裕明, 小山聡, 田中克己, “Web 画像の「らしさ」検索：語の典型的特徴を表す語集合のソーシャルタギング情報からの取得による Web 画

像検索,” 第 19 回データ工学ワークショップ
(DEWS2008), B9-6, 2008.