Web 上の大量画像を用いた 名詞と形容詞の関係分析

小原 侑也[1]　柳井 啓司[1]

概要：近年では，Flickr のように画像にタグを付与して投稿できるサービスが出現し，Web 上に多くのタグ付き画像が存在している．しかし，付与されるタグのほとんどが 1 語単位で付加されており，タグ間の関連性はあまり考慮されていない．そのため，タグで AND 検索を行うと本来想定しない画像も検索にかかることがある．このようなノイズとなる画像を除去するには，タグ間の関係性を考慮し，名詞と形容詞の同時認識を行うなど，より画像の内容に着目する必要があると考える．本論文では，名詞と形容詞の同時認識に繋げるための前処理として，名詞と形容詞の関係に着目し，視覚的関連性のある名詞と形容詞の組み合わせの発見を行う．視覚的関連性は，エントロピーの計算による方法を用いて画像分布を数値化し，各名詞と形容詞の組み合わせのクラスごとに比較し，考察を行っている．エントロピーの計算結果により，色や時間帯を示す形容詞と名詞の組み合わせで，エントロピーの減少するクラスが多いという結果を得ることができた．また，その他にも，winter+beach クラスや sun+beautiful クラスといった特定のクラスで視覚的関連性が高くなるという事象を確認した．

キーワード：一般物体認識 エントロピー 形容詞

1. まえがき

近年，インターネットおよび，デジタルカメラやカメラ機能付き携帯端末の普及によって，Web 上の画像量は爆発的に増加している．また写真投稿サイトには，flickr のように投稿者が画像に関連したタグを付加するサービスを提供しているものもあり，タグ情報が付加された画像も Web 上に多数存在している．これらのタグ情報は，画像検索や画像収集のキーワードとして，また，画像認識の際のメタデータとして使用されることもある．

しかしながら，タグを付ける際には 1 語単位で付与する場合がほとんどであり，タグ間の関連性が考慮されることは少ない．そのため，タグ情報で AND 検索を行った場合，本来検索対象として想定していない画像が検索にかかることも多々ある．例えば赤い車の画像を検索する際に，red と car で検索を行っても，赤い車の画像以外の画像も検索結果として表示される．図 1 は，赤い車だけでなく，黒い車と赤い空が写った画像も検索結果として提示されるという一例である．このような意図しない画像を自動で排除することができるならば，より直感的な検索が行えるだけでなく，ノイズの少ないデータセットを簡便に作成できるのではないかと期待できる．より精度の高い画像検索や画像収集を行うためには，タグ間の関係性を考慮し，より画像の内容に着目する必要があると考える．

図 1　car および red による検索結果の例

そこで本論文では，画像特徴量を用いたアプローチでタグ間の視覚的関連性の分析を行う．ここで分析の対象とするタグの組は，修飾・被修飾の関係になりやすい名詞と形容詞の組としている．２組の単語の組み合わせにおいて，画像の分布がより狭くなるような組み合わせは，分布に変化が見られない組み合わせより，視覚的な関連性が高いであろうと考えられる．よって，視覚的関連性の高低をエントロピーに基づいて数値化し，比較および考察を行うことが，本論文の目標である．

また今回分析した結果は，“画像内のある領域が赤色である” や，“車である” といった単体での認識でなく，“画像には車が写っており，その車の色は赤色である” のように，ある物体を名詞で認識し，さらに，その物体の状態を形容詞で認識するような同時認識に利用できると期待している．例えば，ある 1 つの形容詞と 2 つの名詞が 1 枚の画像から認識された場合，ある名詞と形容詞の視覚的関連性が他方の名詞との間の視覚的関連性より高ければ，形容詞が前者の名詞を修飾している可能性が高くなると考えられ

[1] 電気通信大学大学院 情報理工学研究科 総合情報学専攻

るため，今回の分析結果をメタデータの 1 つとして用い，同時認識の精度向上を目指したいと考えている．

本論文の構成は次のようになっている．2 章では関連研究と先行研究，3 章では実験の手順，4 章では実験で用いた手法について記載している．また，5 章では本論文で行った実験と実験結果についてを記載し，6 章では実験結果に対する考察を行い，7 章では全体のまとめ，および，今後の課題についてを記載している．

2. 関連研究

近年の画像認識分野における研究では，画像内に写っているオブジェクトの物体名のみでなく，色やテクスチャ，また，“赤い”，“美しい” のような形容詞など，属性を取り上げた研究も増加してきている．ここではそのような，属性に着目した研究を紹介する．

T. L. Berg らの研究 [1] では，色，形，テクスチャという属性に着目している．この研究では，ショッピングサイト内に記載されたテキストから属性に関連した単語を抜き出し，そのテキスト記述と対応した画像中から，最も属性を表している局所領域にラベル付けを行っている．また，局所領域は，75 × 75 ピクセルのブロックで表現されている．

D. Parikh らの研究 [2] では，“nameable” という観点から属性に着目している．“nameable” とは，人間が理解可能でかつ，言語表現が可能であるか否かを示すものである．この論文では，Amazon Mechanical Turk を用いた対話型アプローチによって，“nameable” な属性を発見している

A. Farhadi らの研究 [3] では，属性を用いた画像の描写を行っている．画像から “犬” のみを認識するのではなく，“まだら模様の犬” と認識を行う．そのために意味的な属性に着目し，物体そのものを認識できない場合でも，“まだら模様” といった描写を可能としている．また，この描写を用いることで，区別的な属性にも言及することが可能となっている．すなわち，犬は持っているが羊は持たない属性を取り上げるならば，この属性は犬と羊を区別する属性といえる．さらに，区別的な属性を求めたことにより，本来そのクラスが持つであろう属性の非存在や，本来は存在すべきでない属性の存在にも言及できている．そのため，ドアの写っていない車という認識も可能となっている．またこの論文では，属性が存在する領域の推定も行っている．

S. Dhar らの研究 [4] では，美しさ，興味深さという，特定の属性に着目している．美しさと興味深さの指針として，画像に写った物体の構成に関する属性と内容に関する属性という 2 種類の属性を取り上げている．構成に関する属性には，物体の顕著生や存在位置，色などがあり，内容に関する属性には，画像に主として写っている物体の種類や，撮影場所・場面がある．この研究では，より人間の主観に依存しやすい属性に着目している．

論文 [1]，[2] では主に属性単体の認識を行っており，論文 [3] では，車に対するドアのように存在すべき物体を属性として扱っているが，本論文では，属性を形容詞に限定し，さらに名詞との組み合わせによる関係性に着目している．また，論文 [4] では，形容詞の中でも，美しさや興味深さという特定の形容詞について分析しているのに対し，本論文では，より一般的な形容詞を取り扱っている．

次に，本研究の先行研究となる論文の紹介を行う．ここで挙げるのは，柳井らの研究 [5]，秋間らの研究 [6] および，川久保らの研究 [7] である．

柳井らの研究では，単語概念における視覚性を定量化する方法としてエントロピーを提案し，150 個の形容詞に着目して視覚的関連性について言及している．本論文内で使用するエントロピーによる単語の視覚性を数値化する方法，また，エントロピーの計算方法は，柳井らの研究によって提案された手法を用いる．

秋間らの研究では，概念間の距離関係や上下関係から概念間の階層構造を持ったデータベースを構築している．上下関係を求める際にエントロピーを使用し，画像の分布を計算している．また，画像に付与されているタグ情報も使用している．

川久保らの研究では，単語概念における視覚性と地理的分布を求めている．この研究では，視覚性を求めるために，領域分割やエントロピー計算を用いて，名詞や形容詞といった概念クラスの画像分布の計算を行っている．

本論文で使用するエントロピーの手法は，論文 [5] で提案されたアプローチを用いる．本論文と先行研究の差異は，クラスを 1 単語ではなく 2 単語の組み合わせで分類し，その組み合わせである名詞と形容詞の視覚的な関連度を求めている点である．

3. 提案手法の概要

3.1 実験概要

本研究では，名詞・形容詞間の視覚的関連性を，その名詞・形容詞の組み合わせに対応した画像の分布の広さで表現し，分布が狭いほど視覚的関連性が高いと判断する．分布の広さはエントロピーの概念を用いて数値化し，エントロピーの計算には画像から得られる局所特徴を用いる．

本研究での実験における実行手順を以下に示す．

実行手順

(1) Web 上からタグ検索による画像収集
(2) 画像の領域分割
(3) 部分領域ごとに特徴量抽出および BoF の作成
(4) 正領域判定
(5) 正領域ごとに特徴量分布を pLSA で計算
(6) エントロピーの計算

ステップ 1 では，本論文における研究対象であるタグ付きの画像を Web 上から取得する．

収集した画像には，検索に使用したタグ情報と関連のない物体や背景も写っている．これらは画像分布を調べる際にノイズとなるため，除去する必要がある．タグ情報と関連のある領域を正領域とし，この正領域をステップ 2 からステップ 4 で推定する，

ステップ 2 では，ノイズ除去の前処理として，画像の領域分割を行い，ステップ 3 では，分割された各部分領域ごとに画像特徴量を抽出する．抽出した特徴量は BoF を用いることでベクトル化し，部分領域を数値で表す．このベクトルを用い，ステップ 4 では SVM を利用して正領域の判定を行う．

ステップ 5 では，pLSA を用いて正領域とされた部分領域の特徴ベクトルをクラスタリングする．ここでは，エントロピーの計算に必要な，画像内での隠れトピックの同時確率 $P(z|d)$ を求める．

ステップ 6 では，名詞と形容詞を組み合わせたクラス概念ごとの画像分布を，エントロピーを計算することによって求める．エントロピーは画像分布が狭まる，すなわち画像特徴の分布が狭くなると減少し，分布が広まると増大する．よって，このステップで求めたエントロピーが小さくなるほど，名詞と形容詞の間には高い視覚的関連性があると判断できる．

3.2 タグ共起による関連度

本論文では，エントロピーによる視覚的関連度との比較のため，画像に付加されたタグの共起による類似度を，the Normalized Google Distance（NGD）によって求めた．これを用いて，視覚的な類似度と非視覚的な類似度による，タグ間の関係性の違いについての考察も行った．

4. 提案手法の説明

この章では，実験で用いた手法の説明を行う．本実験では，名詞と形容詞の間の視覚的関連性に言及するため，エントロピーの計算を行った．また，比較のためにタグ間の共起による関連度の計算も行った．

4.1 画像収集

Flickr API を用いて，名詞と形容詞の組み合わせ 1 組ごとに，200 枚のポジティブ画像を収集した．収集の際には，名詞を 1 つ目のタグ，形容詞を 2 つ目のタグとして AND 検索を行い，両方のタグを持っている画像を収集した．また，ネガティブ画像はどちらのタグも持たない画像とし，こちらは 1 組ごとに 800 枚用意した．

また，収集した画像にほぼ同一な画像が多く含まれていると特徴が偏るため，同一投稿者からは 1 枚のみ画像を得ることとした．

Flickr には，画像投稿者が自由に画像へタグを付与する機能がある．このタグ情報には多少のノイズがあることを考慮する必要があるものの，基本的には画像に写っている物体やシーン，また，形容詞のような状態の名前を示しているとみなせる．よって本実験での検索には，このタグ情報を利用している．

4.1.1 Flickr API

Flickr API は写真共有サービスである Flickr が提供する Web API である．Flickr API の利用には，Yahoo.com の ID を取得し，Flickr にユーザ登録する事で得られるようになる API Key が必要となる．

Flickr API を利用して画像の検索を行うには，flickr.photos.search メソッドを用いる．flickr.photos.search メソッドを利用する際には，パラメータを変更することで，画像検索の条件や取得データ内容を変更することができる．

flickr.photos.search メソッドの API リクエスト例

```
http://www.flickr.com/services/method=
flickr.photos.search&api_key=*****
&tags=car,red&tag_mode=all&content_type=1
&sort=interestingness-desc&per_page=500
&page=1&extras=date_taken
```

4.2 領域分割

収集した画像からノイズを減らすために，領域分割を行った．領域分割には JSEG [8] を用いた．分割する際には，最大分割領域数は 10 とし，サイズの小さい領域は統合する後処理を加えた．

4.3 特徴量抽出

本研究では色を考慮した SIFT[9] である Color-SIFT 特徴を抽出し，Bag-of-Keypoints 表現を用いてベクトル化することとした．特徴抽出では，領域分割をした画像の部分領域ごとに抽出を行い，抽出した特徴量は k=1000 の k-means でクラスタリングしたコードブックを基に BoF[10] を作成し，部分領域をベクトルで表現した．なお，特徴抽出における特徴点の決定には Grid を用いた．Grid による特徴点決定では，縦横共に 30 ピクセルごとの点を特徴点とした．そして，特徴点ごとにスケールを 12，15，18，23，30 として，SIFT 特徴の抽出を行った．

4.3.1 Color-SIFT 特徴

Color-SIFT は RGB 色空間の R，G，B それぞれの空間に対して SIFT 特徴を抽出し，それらを同一の特徴点ごとに連結させて新たな特徴ベクトルとする方法である．そのため，この特徴ベクトルのサイズは 128×3 次元で表現される．

4.4 正領域判定

画像からノイズとなる領域を除去するため，前節で説明した BoF を用いて正領域を判定した．正領域判定には Support Vector Machine（SVM）を使用した．

SVM による正領域判定手順を以下に示す．

正領域判定手順

(1) ポジティブ画像の全領域を正例およびテストデータ，ネガティブ画像の全領域を負例とする
(2) 正例および負例を用いて SVM の学習を行う
(3) テストデータの評価を行う
(4) 評価の高さ順にテストデータをソートして，上位のデータのみを新たな正例およびテストデータとし，下位のデータはテストデータから削除して負例に追加する
(5) 2 から 4 を繰り返す

正領域判定により正領域と判断された領域例を図 2 に示す．図 2 では赤線で領域が区切られており，正領域と判断された領域にはラベルが付けられている．

また，本論文では，SVM-light[11] というプログラムを使用して，分類を行っている．

4.5 特徴量分布の計算

特徴量分布を計算するために，pLSA を利用して特徴ベクトルのクラスタリングを行った．また計算には，fold-in heuristic[12] の手法を用いた．

4.5.1 pLSA

pLSA（probabilistic Latent Semantic Analysis）は，T. Hofmann によって提案された，テキストコーパス内のトピック検出を行う，統計的言語処理のためのモデルである [12].

pLSA の計算は次のように行う．まず，$d_i(i = 1, 2, \ldots, I)$ を文書，$w_j(j = 1, 2, \ldots, J)$ を単語，$z_k(k = 1, 2, \ldots, K)$ を隠れトピックとする．隠れトピックとは，文書内における単語の生成に関するトピック変数である．次に，文書 d_i と単語 w_j の生起は独立であると考え，文書 d_i と単語 w_j の同時確率 $P(d_i, w_j)$ を以下のように表す．

$$P(d_i, w_j) = \sum_{k=1}^{K} P(d_i|z_k)P(w_j|z_k)P(z_k) \tag{1}$$

また，文書 d_i 内において，単語 w_j が生成される確率 $P(w_j|d_i)$ は，隠れトピック z_k を用いて以下のように表すことができる．

$$P(w_j|z_k) = \sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i) \tag{2}$$

そして，文書 d_i 内での単語 w_j の出現回数を $n(d_i, w_j)$ と表すと，データの対数尤度は次のように表すことができる．

$$L = \sum_{i=1}^{I}\sum_{j=1}^{J} n(d_i, w_j) \log(d_j, w_j) \tag{3}$$

この対数尤度を最大にするような $P(z_k), P(d_i|z_k), P(w_j|z_k)$ を EM アルゴリズムを用いた最尤推定によって求める．

pLSA は統計的言語処理のモデルであるが，文書を画像，単語を局所特徴ベクトルに置き換えて考えることによって，画像認識においても利用されている．

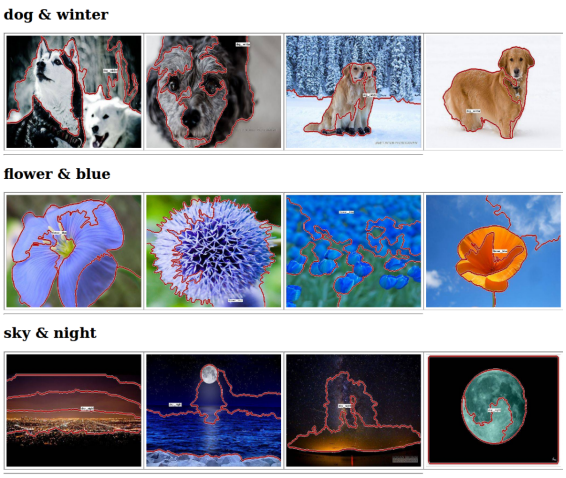

図 2 正領域の一例

4.6 エントロピーの計算

エントロピーは pLSA によって求められた確率を用いて計算を行った．エントロピーは正領域を表す特徴ベクトルの分布が広くなると増大し，狭くなると減少する．そのため，エントロピーの大小が，名詞と形容詞の組み合わせであるクラス概念に属する画像分布の広さを表す．すなわち，エントロピーを計算することが，組み合わせごとの視覚的関連性を求める事に繋がる．

4.6.1 エントロピー

pLSA を用いて求めた $P(z_k|d_i)$ を用いて計算を行う．各隠れトピック z_k に対して，

$$P(z_k|w_j) = \frac{\sum_{i=1}^{I} P(z_k|d_i)}{|I|} \tag{4}$$

を求める．次に，先ほど求めた $P(z_k|w_j)$ を用いて，各画像 d_i に対して，

$$H(P) = -\sum_{k=1}^{K} P(z_k|w_j) \log(P(z_k|w_j)) \tag{5}$$

を求める．この計算によって求まる $H(P)$ がエントロピーである．

4.7 共起による関連度

共起による非関連度は，the Normalized Google Distance（NGD）と呼ばれる以下の式を用いて計算した．

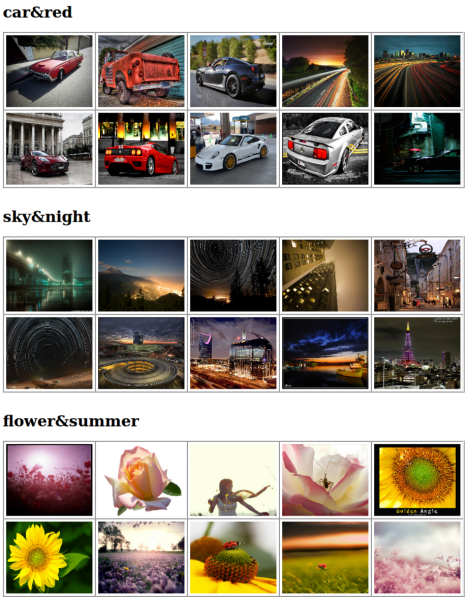

図 3 データセットの一例

$$NGD = \frac{\max\{\log f(x), \log f(y)\} - \log f(x,y)}{\log N - \min\{\log f(x), \log f(y)\}} \quad (6)$$

ここでは x を名詞，y を形容詞とし，$f(x)$，$f(y)$ はそれぞれ，flickr において名詞および形容詞でタグ検索を行った際の画像枚数であり，$f(x,y)$ は名詞および形容詞で AND 検索を行った際の画像枚数である．また，N は flickr の全画像枚数であるが，正確な枚数を把握することは困難なため，50 億枚と仮定した．次に，NGD を用いて，$1 - NGD$ をタグ共起の関連度として定義した．この関連度は，画像特徴を用いない，テキスト情報に基づく関連度である．

5. 実験

5.1 データセット

画像データは Flickr より API を利用して取得した．取得の際，同じ投稿者からの画像は 1 枚のみ取得するという制限をし，ポジティブ画像は 200 枚，ネガティブ画像は 800 枚を取得した．また，ポジティブ画像の取得では，Flickr の検索においてランキングが上位のものから順に取得した．ネガティブ画像は Flickr よりランダムに取得した画像の内，注目するクラスの名詞および形容詞のどちらのタグも付いていない画像を選択した．

今回の実験において，表 1 の 20 単語の名詞と，表 2 の 15 単語の形容詞を選択した．それにより，各名詞と各形容詞の組み合わせである 20 × 15 個のクラスおよび，形容詞による制限を加えない名詞のみの 20 クラスの計 320 クラスに着目し，エントロピーを計算した．

ここで，car+red クラス，sky+night クラス，flower+summer クラス，に含まれる画像の一部を図 3 に示す．

表 1 使用した 20 単語の名詞

beach	bird	boat	bridge	car	cat	cloud
cup	dog	flower	fruit	house	people	sea
sky	snow	sun	tower	train	tree	

表 2 使用した 15 単語の形容詞

red	blue	green	black	white
circle	square	morning	night	winter
summer	new	old	beautiful	cool

5.2 実験方法

前節のデータセットを用いて，以下のように実験を行った．

5.2.1 領域分割

領域分割には JSEG を使用した．分割を行う際の最大領域数は 10 領域である．ただし，この数字は最大の分割数であるため，領域数が 10 より小さくなる画像も多数存在する．また，小さい領域を統合する後処理も加えている．この際，画像を 1 とした時の相対サイズが 0.075 を超える領域は統合しないよう，パラメータ調整を行った．

5.2.2 BoF の作成

BoF を求めるために，まず，コードブックを作成した．コードブックの作成には，タグによる制限を加えずランダムに選択した 10000 枚の画像を使用した．この 10000 枚の画像に対して Grid による特徴点決定を行い，Color-SIFT 特徴量を抽出し，全特徴量からランダムに 100 万個の特徴量を選択した．この 100 万個の特徴量に対して，k-means 法を用いて k=1000 のクラスタリングを行った．そのため，本実験にて使用したコードブックのコードブックサイズは 1000 となっている．次に，データセットとして用意したポジティブ画像とネガティブ画像に対しても，同様に Color-SIFT 特徴を抽出した．そして先ほどのコードブックを用い，領域分割によって得た各部分領域ごとに，1000 次元の BoF を作成した．

5.2.3 正領域判定

正領域判定は，SVM を用いて行った．本実験では，SVM による学習とテストを 5 回繰り返すことで，正領域を推定した．まずポジティブ画像の全領域の BoF をポジティブデータ，ネガティブ画像の全領域の BoF をネガティブデータとして SVM の学習を行い，それを用いて全ポジティブデータをテストした．このテストで評価の高かった上位 600 個の BoF を次のポジティブデータおよびテストデータとし，残りの BoF はネガティブデータとして，次の学習を行った．そして，2 回目のテストでは上位 500 個，3 回目のテストでは上位 400 個と 100 個ずつ減らした数の BoF をポジティブデータと見なすことにより，最終的に残った 200 個の領域が正領域であると判断した．

5.2.4 pLSA による特徴量分布の計算

本実験では，fold-in heuristics を用いて各クラスで pLSA の計算を行うために，まず，pLSA のベース分布を求めた．ベース分布を求める際に使用したのは，ポジティブ画像に

名詞/形容詞	–	red	blue	green	black	white	circle	square	morning	night	winter	summer	new	old	beautiful	cool
beach	5.383	0.198	0.099	−0.009	0.027	−0.059	0.018	0.181	**0.338**	0.305	0.101	−0.058	0.037	−0.045	0.075	0.011
bird	5.478	0.147	0.193	0.182	0.029	−0.045	−0.009	0.115	**0.321**	0.034	0.103	0.212	−0.023	−0.012	0.063	0.082
boat	5.398	0.193	0.123	−0.065	0.110	−0.034	−0.045	0.122	**0.440**	**0.297**	0.095	0.020	0.065	−0.050	0.197	−0.053
bridge	5.466	0.071	**0.354**	0.161	0.232	0.078	−0.018	0.151	**0.336**	0.143	0.003	0.042	0.085	−0.028	0.016	−0.022
car	5.486	0.139	0.105	0.003	0.130	0.118	0.131	0.035	0.101	0.129	0.049	0.044	0.150	−0.003	0.018	0.039
cat	5.521	0.003	0.061	0.046	0.145	0.117	0.061	0.092	0.032	0.083	0.069	0.064	0.046	0.044	0.070	0.048
cloud	5.334	0.078	0.066	−0.020	0.154	−0.024	0.030	0.217	0.220	0.135	0.063	−0.064	0.069	−0.005	0.086	0.014
cup	5.431	0.105	0.137	0.100	0.121	0.150	0.073	0.096	0.169	0.103	0.132	0.013	−0.027	−0.060	−0.015	−0.005
dog	5.522	0.027	0.024	0.069	0.120	0.124	0.144	0.086	0.137	0.211	0.069	0.048	0.050	0.038	0.066	0.008
flower	5.357	0.096	0.185	0.145	0.082	0.055	−0.040	0.175	0.153	0.088	0.011	0.077	0.106	−0.128	0.018	0.030
fruit	5.474	0.112	0.113	0.157	0.242	0.085	0.042	0.113	0.006	0.050	0.117	0.149	0.007	−0.046	0.061	0.018
house	5.536	0.114	0.170	0.163	0.161	0.060	0.040	0.091	0.224	0.078	0.129	0.033	−0.011	0.093	−0.003	−0.001
people	5.519	0.084	0.047	0.024	0.114	0.078	0.035	0.013	0.164	0.153	0.093	0.020	0.134	0.058	0.090	0.040
sea	5.368	0.211	−0.022	−0.038	0.198	−0.030	−0.032	0.108	**0.439**	0.237	0.198	−0.066	0.056	−0.021	0.077	−0.006
sky	5.387	0.188	0.108	0.016	0.146	0.030	−0.026	0.036	**0.287**	0.237	0.011	0.048	0.053	−0.022	0.006	−0.002
snow	5.490	0.036	**0.261**	0.038	0.084	0.044	−0.014	0.167	**0.279**	0.159	0.054	−0.009	−0.013	0.047	0.084	0.077
sun	5.380	**0.278**	0.044	0.027	0.069	−0.008	0.176	0.042	**0.237**	0.248	0.069	0.007	−0.016	−0.067	0.179	0.013
tower	5.473	0.113	0.234	0.051	0.151	0.046	0.022	0.063	**0.443**	0.101	0.044	0.012	0.043	0.037	0.015	0.015
train	5.535	0.056	0.133	0.054	0.242	0.128	0.149	0.071	0.036	0.145	0.028	0.016	0.040	0.045	0.050	0.023
tree	5.437	0.014	0.137	0.072	0.183	0.058	−0.022	0.173	**0.376**	0.186	0.164	0.056	0.046	0.056	−0.003	0.011

図 4　エントロピーの計算結果（名詞のみのエントロピーからの減少量，
赤字:名詞のみより増加，青字:名詞のみより減少，下線:減少量が大きいクラス）

表 3　エントロピー上位 10 組

順位	クラス	エントロピー	順位	クラス	エントロピー
1	sea+morning	4.929	6	beach+night	5.078
2	boat+morning	4.958	7	sky+morning	5.100
3	tower+morning	5.030	8	boat+night	5.101
4	beach+morning	5.045	9	sun+red	5.102
5	tree+morning	5.061	10	bridge+blue	5.112

表 4　エントロピー下位 10 組

順位	クラス	エントロピー	順位	クラス	エントロピー
10	train+winter	5.507	5	fruit+old	5.520
9	train+cool	5.512	4	bird+white	5.523
8	dog+cool	5.5143	3	house+cool	5.537
7	cat+red	5.518	2	house+beautiful	5.539
6	train+summer	5.519	1	house+new	5.547

おける全領域の BoF からランダムに選択した 20000 個の BoF である．この BoF に対して pLSA を用いてクラスタリングを行い，特徴量分布を求めた．この際のクラスタリングにおけるクラスタ数は 300 とした．また各クラスの分布を求める際は，クラスタ数の異なる 3 種類のベース分布による fold-in heuristics の手法を使用した．

5.2.5　エントロピーの計算

エントロピーは pLSA を用いて計算した同時確率 $P(z|d)$ を用いて計算を行った．計算結果は，次の節にて掲載する．

5.3　実験結果

各クラスについてエントロピーを計算した．pLSA によるクラスタリングのクラスタ数を 300 とした際の計算結果を図 4 に示す．形容詞と組み合わせたクラスでは，名詞のみのクラスのエントロピーからの減少量を掲載している．そして，エントロピーの小さい上位 10 組と大きい下位 10 組を表 3，4 に示す．また，共起を用いた類似度の計算結果の一部を図 5，6 に示す．

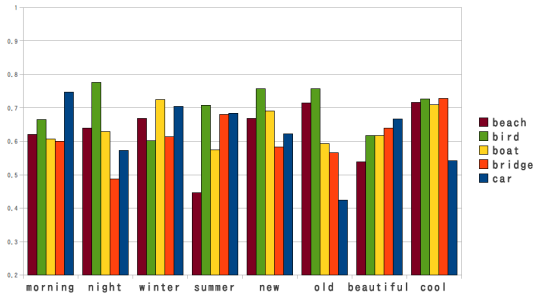

図 5　共起による類似度（beach，bird，boat，bridge，car）

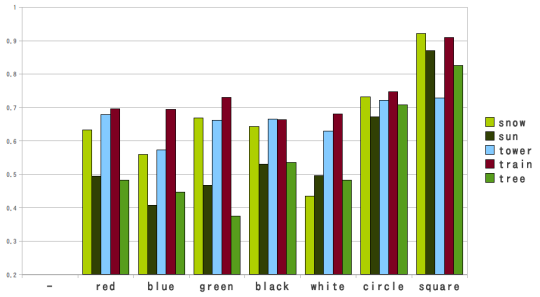

図 6　共起による類似度（snow，sun，tower，train，tree）

6.　考察

実験結果より，各クラスにおけるエントロピー，また，エントロピーの変化について比較を行う．エントロピーは名詞と形容詞を組み合わせた各クラスにおける画像分布が広がると増大し，狭まると減少する．したがって，エントロピーが小さいクラス概念では，名詞と形容詞の間に高い視覚的関連性があると判断する．

6.1　名詞と形容詞の視覚的関係性

色に関する形容詞を付与することで大きくエントロピー

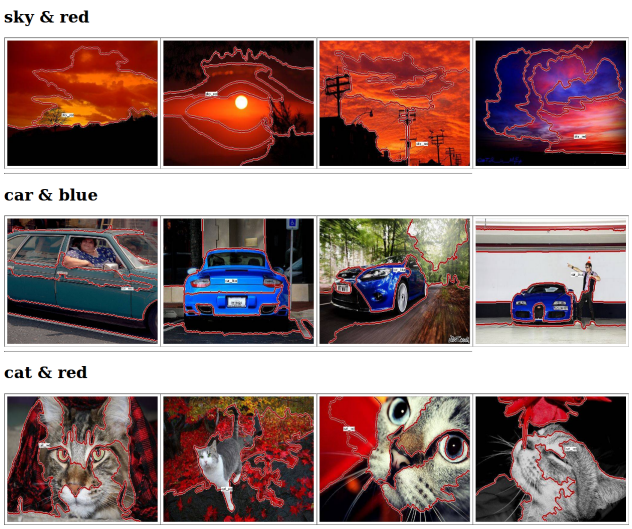

図 7 色に関する形容詞の正領域の一例

が減少しているクラスは，beach+red，sea+red，sky+red，sky+blue，flower+blue クラスが挙げられる．これらのクラスの共通点としては，名詞の示す物体自体の色を形容詞が示しているという点である．beach，sea，sky と red を組み合わせたクラスにおいて正領域とされた領域を持つ画像には，日の出や日没によって空や海が赤くなっている画像が多く見受けられている（図 7 参照）．そのため，car や flower，bird のように，その名詞が示す物体にカラーバリエーションの多い名詞のクラスでは，色に関する形容詞が付与されたクラスでエントロピーが減少することが多い（図 7 参照）．一方，dog や people に red や blue といった色に関する形容詞が付与されても，変化するのは服や背景の壁の色であり，物体自体の領域では色の変化はあまり無い（図 7 参照）．この事から，色を示す形容詞に関しては，エントロピーが減少する名詞と形容詞の組み合わせで，視覚的関係性が高くなっていると考えられる．

次に，時間帯を示す形容詞に関するクラスについて考える．morning や night と組み合わせてエントロピーが減少した名詞には，sky，sun，cloud，sea，boat，tower が挙げられる．これに関して共通しているのは空であると考えられる．sky，sun，cloud については空と直接関係した名詞であるため，朝と夜の空の色変化による影響でエントロピーが減少していると考えられる（図 8 参照）．よって視覚的関係性が高いと言えるだろうと思われる．また，boat と morning や night を組み合わせたクラスの画像には，船上から写した空や海といった景色の写真である画像が多く含まれており，tower のタグが付いた画像には，tower の全体像を写すための遠景写真や tower を見上げている構図の写真である画像が多くあるため，空の領域が多く含まれている（図 8 参照）．そのために，エントロピーが減少したと考えられる．よって boat と tower のクラスについてはこれらのタグが付いた画像と時間帯を示す形容詞の間に視

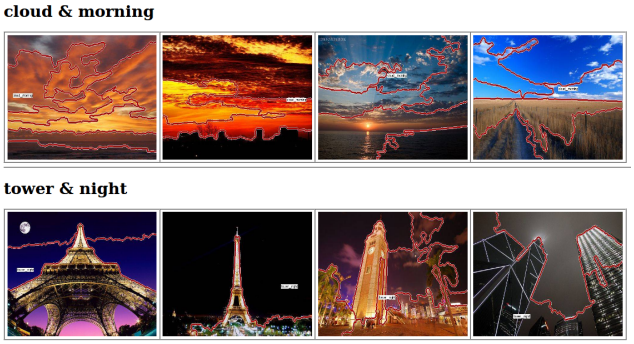

図 8 時間帯に関する形容詞の正領域の一例

覚的関係性が存在すると思われる．

新旧に関する形容詞について考える．old の付くクラスでは特に分布のばらつきが見受けられ，全 20 個の名詞のうち，名詞のみのクラスと比較してエントロピーが増加するクラスは 12 クラス存在する．これには，今回選択した名詞の種類による影響も考えられる．sun，sky，cloud といった新旧による変化のない名詞のクラスでは，new や old が名詞を修飾している事が少ないため，画像分布にばらつきが生じやすかったのではないかと思われる．ここでエントロピーが減少した残りの 8 クラスに着目すると，cat，dog，people といった生物に関係した名詞，house，tower，train といった人工物に関係した名詞が挙げられ，時間の経過による変化が見受けられる名詞に関しては new や old が名詞を修飾している画像が多い（図 9 参照）．この事から，エントロピーの減少幅は少ないものの，エントロピーが減少したクラスでは，名詞と形容詞の間に視覚的関係性が存在すると考えて良いと思われる．

beautiful や cool のような形容詞については，投稿者が自由にタグを画像に付加できるという Flickr のサービスのシステム上，投稿者の主観に影響を受けている事が考えられる．そのため，収集した画像全体の分布のばらつきが他の形容詞の付いたクラスよりも大きく，正領域判定において，ふさわしくない領域が選ばれていたり，正領域判定を行っても，ノイズを減少させる事ができなかったのではないかと考える．また，beautiful で特に減少が見られたのは，sun+beautiful，boat+beautiful，cloud+beautiful クラスである．これらのクラスでは，全体として朝焼けや夕焼けを写した画像が多く見受けられた（図 9 参照）．cool や beautiful の付くクラスでは他に大きくエントロピーが減少クラスは見られなかったため，sun，boat，cloud に関しては，美しいと考える景色に一定の共通概念があり，beautiful との視覚的関連性が高くなっていると考えられる．

summer や winter のような季節を示す形容詞では，beach+winter，sea+winter クラスでエントロピーが減少している．これらのクラスでは，海水浴の画像が減り，海や浜辺を主として写している画像が多くなっていることがエントロピーの減少に繋がっているように思われる．その

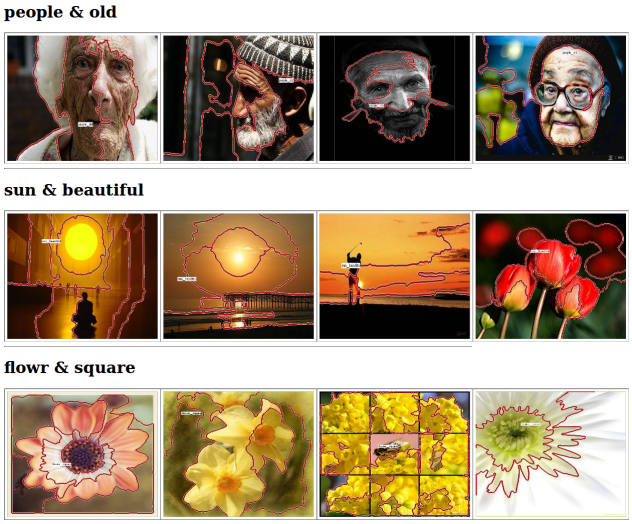

図 9 正領域の一例

ために画像分布が狭まったのだと考える．また，これらのクラスでは，tree+winter，flower+summer，bird+summer クラスでもエントロピーが減少しており，視覚的関連性があるようである．

6.2 タグ共起による関連度との比較

共起を用いた場合でも関連性を調べることは可能であるが，画像認識と組み合わせる場合には上手く機能しないことがある．その主な理由は，画像内の主となる物体に関してだけでなく，雑多なタグ付けが行われていることである．そのため共起の度合いが，必ずしも視覚的な関連性を表すことができていない．

例えば，beach+summer クラスでは共起による関連性があるものの，画像には浜辺ではなく人物や船といった被写体が写っている物も多い．そのため，視覚的な関連性は低いと言える．

その一方，極端に共起による関連度が高いクラスに着目すると，視覚的な類似度が高くなっている傾向があると考えられる．例えば，sun+red クラスや sky+blue クラスでは，形容詞が名詞を修飾しており，視覚的な類似度が高くなっている．

7. おわりに

7.1 まとめ

本論文では，Flickr から特定の名詞と形容詞の両方のタグが付いた画像を収集し，それらの画像から特徴量抽出を行い，画像分布をエントロピーという形で数値として計算した．また，名詞と形容詞の間にある視覚的関連性について，エントロピーの増減から名詞と形容詞を組み合わせたクラスごとに比較，考察を行った．

結果として，エントロピーが減少し，画像分布に偏りを持つ名詞と形容詞の組み合わせを発見することができた．分布に偏りがあり視覚的関連性の高い組み合わせには，色を示す形容詞や時間帯を示す形容詞が多く出現することを，この実験により得ることができた．また，beautiful のように人間の主観に依存しやすい形容詞に関しても，sun のように画像分布に偏りが生じる事例があることも分かった．

7.2 今後の課題

今回の実験ではデータセットが Flickr のみであることや，使用した特徴量が Color-SIFT のみであることから，他のデータセットや特徴量を使用した実験を行うことも改良点として挙げられるのではないかと考える．これについては，データセットや特徴量の組み合わせを数種類用意し，エントロピーの増減の変化が各組み合わせの間で見受けられるかどうかにも着目したいと思う．

また，今回求めた視覚的関連性の分析結果を利用し，今後，名詞と形容詞の同時認識に繋げていきたいと考える．

参考文献

[1] T. L. Berg, A. C. Berg, and A. J. Shih. Automatic attribute discovery and characterization from noisy web data. In *Proc. of European Conference on Computer Vision*, pp. 663–676, 2010.

[2] D. Parikh and K. Grauman. Interactively building a discriminative vocabulary of nameable attributes. In *Proc. of IEEE Computer Vision and Pattern Recognition*.

[3] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes.

[4] S. Dhar, V. Ordonez, and T.L. Berg. High level describable attributes for predicting aesthetics and interestingness. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 1657–1664. IEEE, 2011.

[5] 柳井啓司, Kobus Barnard. 一般物体認識のための単語概念の視覚性の分析. 情報処理学会論文誌: コンピュータビジョン・イメージメディア, Vol. 48, No. SIG10 (CVIM17), pp. 88–97, 2007.

[6] 秋間雄太, 川久保秀敏, 柳井啓司. Folksonomy を用いた画像特徴とタグ共起に基づく画像オントロジーの自動構築. 電子情報通信学会論文誌. D, 情報・システム, Vol. 94, No. 8, pp. 1248–1259, 2011.

[7] 川久保秀敏, 柳井啓司. 単語概念の視覚性と地理的分布の関係性の分析. 電子情報通信学会論文誌. D, 情報・システム, Vol. 93, No. 8, pp. 1417–1428, 2010.

[8] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 8, pp. 800–810, 2001.

[9] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, 2004.

[10] G. Csurka, C. Bray, C. Dance, and L. Fan. Visual categorization with bags of keypoints. In *Proc. of ECCV Workshop on Statistical Learning in Computer Vision*, pp. 59–74, 2004.

[11] T. Joachims. *SVM ^{light}: Support Vector Machine*. http://svmlight.joachims.org/.

[12] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, Vol. 43, pp. 177–196, 2001.