3B4-4

Flickr 上の画像の WordNet への自動マッピング

Automatically Mapping Flickr Images to WordNet

馬場 雪乃[*1]　本位田 真一[*2]
Yukino Baba　Honiden Shinichi

[*1]東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

[*2]東京大学大学院情報理工学系研究科, 国立情報学研究所
Graduate School of Information Science and Technology, The University of Tokyo, National Institute of Informatics

In this research, I intend to obtain the image corpus from Web for understanding the semantics of images. The tagged images on Flickr are automatically mapped to the corresponding WordNet synset and scored by the suitability for a training data of generic object recognition. I will propose a method to utilize the built image corpus for object recognition of the untagged and tagged images. The benefit of this research is building an automatic updating image corpus. I believe that the image corpus possibly improve the accuracy of the search and classification drastically.

1. はじめに

大量の画像が Web 上で入手可能となってきたことに伴ない，必要な画像を効率良く検索する手法が求められている．一般的な画像検索手法では，画像周辺の文字列や画像に付与されたタグなどのテキスト情報を利用して，検索クエリと関連度が高い画像を選び出している．しかし，テキスト情報と画像とが直接結びつかないことも多いため，テキスト情報のみを利用して目的の画像を探し出すことは難しい課題である．

テキスト情報を用いずに，画像の意味を理解することで意味内容に基づく検索・分類を達成するための研究が行われている．画像の意味理解という問題は一般物体認識と呼ばれ，写真画像中のシーンに関する意味カテゴリ（例えば「空」「海」）や，画像中のオブジェクトの意味カテゴリ（例えば「リンゴ」「猫」）を認識するタスクとなる．

一般物体認識では，画像に対して意味カテゴリがラベルづけされた，大規模な学習データセットが必要となる．現在は人手で構築されたデータセットが主に用いられており，一般的に用いられているものとしてカリフォルニア工科大学によって作成された Caltech-101, 256[Fei-Fei 04, Griffin 07] がある．Caltech-101 は 101 種類（例．face, zebra）の画像から成り，主に Google Image Search を用いて人手で集めた 9,144 枚の画像から構成されている（1 カテゴリ辺りの画像数は 31 から 800 枚）．Caltech-256 は 256 種類，30,607 枚から成る．

このような人手で作られたデータセットには以下の 3 つの問題点がある．

- 限られたカテゴリ数
- 限られた画像数
- 少数のデータセット作成者の意図だけが反映されている

これらの問題点は全て，少数の人間の手によってデータセットが作成されていることに起因する．

大量の人間を利用することでこの問題に対応したのが ImageNet[Deng 09][*1]である．ImageNet は英語の概念辞書である WordNet[Fellbaum 98] 上の名詞概念に人手で画像を割り当てようという試みである．WordNet 上には約 80,000 の概念が存在するため，現状のデータセットよりも多くのカテゴリに対応することができる．また，著者らは Amazon Mechanical Turk (AMT) という，オンラインでユーザにタスクを割り振り金銭の支払を行うサービスを利用して，大量の画像へのカテゴリラベル付けを行っている．1 つの概念への割り当て画像数は 500 から 1000 個になるのが目標とされている．また，カテゴリ分類は複数人の投票によって行われているため，少数の人間の意図のみが反映されている可能性が減少している．なお，現状では 3.2million の画像が 5,247 個のカテゴリに分類されている．

ImageNet は，今までの少数の人間によって作成されたデータセットの問題点を解決しているが，そのために大量の人的リソースを消費している．人手を使わずに，コンピュータで自動的に ImageNet と同様のものを構築することはできないのだろうか？これが，本研究で解きたい問題である．さらに，自動構築をすることで，次々と Web 上に追加される新しい画像をデータセットに追加することも容易となり，自動アップデートされる画像データセットの構築が可能となる．

WordNet の概念に画像を割り当てているため，ImageNet には他の人手で作成されたコーパスと比べて別の利点も持つ．例えば，ESP dataset[Von Ahn 04] というコーパスは，人間がゲームをしながら画像と関連するラベルを入力することで構築されているが，ESP dataset における問題点を ImageNet のプロジェクトメンバーは指摘している．ESP dataset では，ラベルづけを行うプレイヤーが”bank” というラベルを入力した場合に，それが”a river bank（川岸）” を指すのか”financial institution（銀行）” か，不明確となってしまう．しかし ImageNet では，多義語の概念があらかじめ分けられている WordNet を利用しているため，このような問題は起こらない．

同様の利点が本研究にもある．キーワードと関連する画像を自動的に収集する手法がいくつか提案されているが [Fergus 05, Wang 06, Yanai 05]，これらの既存手法で用いるキーワードが複数の意味を持つ場合，上述の ESP dataset と同じ問題が生じてしまう．本研究では，WordNet に画像を自動的に割り当てることで，多義語に対してその意味ごとに画像を割り当て

連絡先: ybaba@nii.ac.jp

*1 http://www.image-net.org

ることを目指す.

ImageNet のような画像コーパスを作成するため，本研究では Flickr で画像に付与されているタグに着目する．Flickr では画像に対して，画像と関連する語が人手によりタグとして選ばれ，画像に付与されている．これらのタグは，画像中のシーン，オブジェクトを表していることが多いと予想される．そこで，タグを活用することで，WordNet 上の概念への画像割り当てを可能とする手法についての研究を行う.

特に本稿では，タグを利用して，多義語のタグを持つ画像を，各概念に割り当てる手法について述べる．以下では，まずシステムの全体像について述べた後，タグを用いての画像割り当て手法について述べ，実験結果を紹介する.

2. システム構成

本節では，まず ImageNet が用いているコーパス構築手順について述べ，次に提案システムでの構築手順の概略を述べる.

2.1 ImageNet

ImageNet は，*Collecting candidate images* と *Cleaning candidate images* の 2 ステップから成る.

2.1.1 Collecting candidate images

ImageNet は，複数の画像検索エンジンを用いて候補画像を収集している．収集に用いるクエリは，(1) 各概念に対応する synset[*2], (2) 上位概念の synset のうち，対象概念の定義文中に現れる語と，対象概念の語の組み合わせ．例えば，“whippet” という概念に”small slender dog of greyhound type developed in England” という定義文が与えられていたとする．このとき，“whippet dog” と”whippet greyhound” というクエリを用いる．さらに，これらのクエリを他の言語に翻訳することで，より多くの画像を収集している.

2.1.2 Cleaning candidate images

Cleaning ステップは，収集した画像が確かに概念を表現しているかどうかを判定する．ImageNet は，AMT を利用して集められたユーザを用いて，このステップを人手で処理している．各概念について，収集した画像と概念の定義文をユーザに提示する.「各画像が，概念によって表されるオブジェクトを含んでいるか？」という質問をユーザに投げ，ユーザはその質問に Yes か No で回答する．人手によるミスを避けるため，同じ画像，概念に対する評価は複数のユーザによって行われる．最終的に，画像が確かに概念を表現しているかどうかは，多数決で決定される.

2.2 提案システム

2.2.1 Collecting candidate images

ImageNet においては，WordNet から得られる情報のみを用いてクエリを選択していた．しかし，Flickr 上で得られる情報を用いてクエリをさらに拡張することが考えられる．例えば，WordNet 上では New York City という概念に対して，“New York”, “Greater New York” という synset が与えられている．しかし，New York City を”nyc” と略すこともあり，Flickr 上では”new york” タグと共に”nyc” タグが画像に与えられることがある．このような情報をタグの共起関係などを用いて抽出しクエリ拡張に用いることで，ImageNet より多くの候補画像を収集できるだろう．本ステップの具体的な手法については本稿では議論せず，今後の取り組みとする.

*2 WordNet では，各概念に概して synset と呼ばれる同義語のグループが与えられている

2.2.2 Cleaning candidate images

ImageNet においては人手で,「収集した画像それぞれが確かに概念を表現しているかどうか」の判定を行っていた．このステップを自動で行うことが，本研究の主要な取り組みである．Flickr 上においては，画像に関連するタグが人手で与えられている．このタグ情報を利用して，画像が概念を表現しているかどうかを推定していく．具体的な手法は次節で述べる.

3. Cleaning candidate images

本稿では，*Cleaning candidate images* のステップを，タグを利用して自動的に行う手法を提案する.

このステップがどのようなものであるかを，例を用いて説明する．WordNet 上で与えられる「恒星」という概念に画像を割り当てるとする.「恒星」概念には”star” という synset が割り当てられているため，“star” というクエリでまず画像を収集したとする．このとき，“star” というタグを持つ画像には図 1 のようなものがある.「恒星」を表す画像以外に，星マークや,「スター」である歌手,「スターウォーズ」に関連する画像などが含まれている．この中から，タグだけを用いて「恒星」を表す画像を選び出すというのが，このステップで行う処理である.

本節においては，このステップがどのようなタスクであるかを述べた後，提案手法と実験結果について述べる.

3.1 問題定義

提案システムにおける *Cleaning candidate images* のステップにおいては，収集した画像それぞれが確かに概念を表現しているかどうかを，画像に付与されたタグを用いて判定する．各画像に対して，入力と出力は以下となる.

INPUT

T_s: 画像 i_s に付与されたタグ集合

OUTPUT

$a(s,c) = \{True, False\}$: 画像 i_s が概念 c を表現しているか否か

3.2 提案手法

画像に与えられたタグ集合から，どのようにして画像が概念を表現しているか否かを判定すれば良いだろうか？提案手法では，以下の手順でこの問題を解く.

1. タグ $t_j \in T_s$ について，対象の概念 c との関連度 $r(j,c)$ を求める
2. 画像 i_s と概念 c との関連度 $r(s,c)$ を次式で求める.

$$r(s,c) = \frac{\sum_{t_j \in T_s} r(j,c)}{|T_s|}$$

3. 次式で $a(s,c)$ の値を定める.

$$a(s,c) = \begin{cases} True, & \text{where } r(s,c) > threshold \\ False, & \text{otherwise} \end{cases}$$

従って，必要となるのはタグと概念の関連度 $r(j,c)$ である．この値を求める簡単なやり方としては，タグを WordNet 上の概念に割り当て，WordNet の構造を利用して 2 つの概念の類似度を計算することである（例えば，概念間のパス数などが考えられる）．しかし，ユーザが自由にタグとなる語を選べる Flickr 上においては，WordNet 上の概念に割り当てるこ

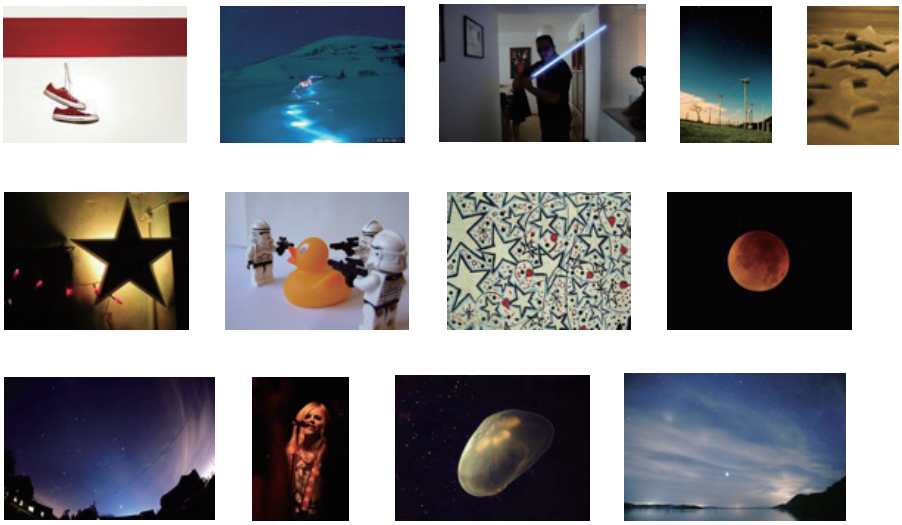

図 1: ”star” というタグを持つ画像の例

とができないタグが多数存在する．そこで本手法では，Flickr 上で求められるタグ類似度グラフに PageRank アルゴリズム [Brin 98] を適用し，WordNet 上に存在するタグと対象概念の類似度（関連度）を，WordNet 上に存在しないタグに対して伝搬させる．

具体的には，次式を用いてタグと概念の関連度計算を行う．

$$r_k(j,c) = \alpha \sum_i r_{k-1}(i,c)p_{ij} + (1-\alpha)v(j,c)$$

ここで，$r_k(j,c)$ は PageRank のステップ k における値，p_{ij} はタグ類似度グラフによって与えられる t_i と t_j の類似度，$v(j,c)$ は WordNet で与えられる概念 c とタグ j_c の類似度，α はタグ類似度と WordNet 類似度の重みパラメータである．

3.3 実験

提案手法の妥当性を確認するため，初期実験として小規模なデータセットに対して手法を適用した．具体的には，“star” というタグを持つ画像を，WordNet 上で”star” を synset として含むいくつかの概念に対して割り当てることを試みた．

3.3.1 データセット，設定

データセットとして，MIRFLICKR-25000[Huiskes 08] を利用し，“star” タグが含まれる画像 94 枚を選び出した．画像の例を図 1 に示す．WordNet 上で”star” で表される概念 8 個について，各画像がその概念を含むかどうかを人手で判定し，正解データとした．8 個の概念のうち，4 個の概念に関する画像は存在しなかったため，以降の実験は残りの 4 個についてのみ行った．実験で用いた概念を表 1 に示す．

タグ類似度グラフは，MIRFLICKR-25000 上のタグの共起数を利用して構築した．ノードはタグとし，6 回以上共起しているタグペアは，共起回数を重みとするエッジを持つ．WordNet 上の類似度を計算する手法は，Lin Similarity[Lin 98] を用いた．

また，ベースラインとして，「94 個の画像全てが，4 個の概念それぞれについて True である」という結果を用意した．

3.3.2 評価

表 2,3,4,5 に実験結果を示す．$\alpha = 0.0, 0.5, 1.0$ が提案手法である．いずれの概念においても，$\alpha = 0.0$ の時が最も F 値が良くなっている．$\alpha = 0.0$ はWordNet 上で得られる類似度のみを用いて判定を行ったことと同じであるため，共起関係から得られるタグの類似度を用いるよりも，WordNet 上の類似度のみを用いてタグと画像の関連度を計った方が，より良い結果が得られている．

また，ベースラインと比較すると，提案手法はわずかではあるが F 値を改善できている．このことから，提案手法によって誤った画像が割り当てられることを幾分回避できていることがわかる．

	ベースライン	$\alpha = 0.0$	$\alpha = 0.5$	$\alpha = 1.0$
precision	0.3936	0.4583	0.3978	0.3978
recall	1.000	0.8919	1.000	1.000
F-measure	0.5649	0.6056	0.5692	0.5692

表 2: 概念 1 の割り当て評価

	ベースライン	$\alpha = 0.0$	$\alpha = 0.5$	$\alpha = 1.0$
precision	0.3404	0.3889	0.3441	0.3441
recall	1.0000	0.8750	1.0000	1.0000
F-measure	0.5079	0.5385	0.5120	0.5120

表 3: 概念 2 の割り当て評価

	ベースライン	$\alpha = 0.0$	$\alpha = 0.5$	$\alpha = 1.0$
precision	0.3191	0.3962	0.3226	0.3226
recall	1.0000	0.7000	1.0000	1.0000
F-measure	0.4839	0.5060	0.4878	0.4878

表 4: 概念 3 の割り当て評価

	ベースライン	$\alpha = 0.0$	$\alpha = 0.5$	$\alpha = 1.0$
precision	0.0212	0.0256	0.0215	0.0256
recall	1.0000	1.0000	1.0000	0.5000
F-measure	0.0417	0.0500	0.0421	0.0488

表 5: 概念 4 の割り当て評価

	synset	定義
1	star	(astronomy) a celestial body of hot gases that radiates energy derived from thermonuclear reactions in the interior
2	star	any celestial body visible (as a point of light) from the Earth at night
3	star	a plane figure with 5 or more points; often used as an emblem
4	headliner, star	a performer who receives prominent billing

表 1: ”star” で表される概念例

4. むすび

本研究では，ImageNet のような画像コーパスを自動構築するために Flickr 上で画像に対して与えられているタグを用いる手法を提案した．特に，収集した画像それぞれが確かに概念を表現しているかどうかを，画像に付与されたタグを用いて判定する手法を提案した．

実験により，WordNet 上で得られる類似度のみを用いた方が，Flickr 上で得られる，共起関係に基づくタグ類似度を用いるよりも良いという結果が得られた．共起関係は，タグの類似度としては最も単純な指標である．今後，他の類似度指標も用いてさらに手法の改善を行っていきたい．また，画像と概念の関連度は，画像に与えられた各タグと概念の平均値で求めたが，タグごとに重みを与えるなどの手法も試していきたい．

参考文献

[Brin 98] Brin, S. and Page, L.: The anatomy of a large-scale hypertextual Web search engine* 1, *Computer networks and ISDN systems* (1998)

[Deng 09] Deng, J., Dong, W., Socher, R., Li, L., Li, K., and Fei-Fei, L.: ImageNet: a large-scale hierarchical image database, in *Proc. CVPR* (2009)

[Fei-Fei 04] Fei-Fei, L., Fergus, R., and Perona, P.: Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories, in *Proc. Workshop on Generative-Model Based Vision* (2004)

[Fellbaum 98] Fellbaum, C., et al.: *WordNet: An electronic lexical database* (1998)

[Fergus 05] Fergus, R., Fei-Fei, L., Perona, P., and Zisserman, A.: Learning object categories from google' s image search, in *Proc. ICCV* (2005)

[Griffin 07] Griffin, G., Holub, A., and Perona, P.: Caltech-256 Object Category Dataset, Technical report (2007)

[Huiskes 08] Huiskes, M. and Lew, M.: The MIR flickr retrieval evaluation, in *Proc. MIR* (2008)

[Lin 98] Lin, D.: An information-theoretic definition of similarity, in *Proceedings of the 15th International Conference on Machine Learning* (1998)

[Von Ahn 04] Von Ahn, L. and Dabbish, L.: Labeling images with a computer game, in *Proc. SIGCHI* (2004)

[Wang 06] Wang, X., Zhang, L., Jing, F., and Ma, W.: Annosearch: Image auto-annotation by search, in *Proc. CVPR* (2006)

[Yanai 05] Yanai, K. and Barnard, K.: Probabilistic web image gathering, in *Proc. Workshop on Multimedia information retrieval* (2005)