オントロジーを利用したWeb画像サイトからの階層的画像データセット生成

藤本 椋也[†,a]　山西 良典[††,b]
岩堀 祐之[†,c]　年岡 晃一[†,c]　福本　淳一[††,b]

† 中部大学大学院工学研究科情報工学専攻　†† 立命館大学情報理工学部メディア情報学科

a) *fujimoto@cvl.cs.chubu.ac.jp*　b) {*ryama*, *fukumoto*}*@media.ritsumei.ac.jp*

c) {*iwahori*, *toshioka*}*@cs.chubu.ac.jp*

概要　一般物体認識に用いられる学習用の画像データセットの作成には多大な人的および時間的コストがかかる．近年では，Flickr のような Web 画像サイトから生成することが行われているが，タグの持つ意味的な情報を利用したデータセットの生成や分類を行っているものは少ない．そこで本研究ではタグの持つ意味的な情報を利用し，階層構造上に分類された画像データセットを自動生成する手法を提案する．意味的な階層構造とは”犬”や”猫”の上位概念として”動物”が存在するような関係である．このような画像データセットを用いた一般物体認識の学習を行うことで，計算機が “動物”や “乗り物”といった広い範囲から徐々に詳しい認識を行うような認識メソッドを獲得することができる．実験では 47910 枚の画像が 183 のクラスに階層分類された画像データセットの自動生成に成功した．ラベルに対する画像の内容の妥当性を人手で評価したところ，自動的な画像データベース生成としては高い精度で，適切な画像が収集・構造化されていることを確認した．

キーワード　Web 画像マイニング; オントロジー; 階層構造データセット; 一般物体認識;

1　はじめに

一般物体認識とは実世界における制約のない画像が与えられたとき，計算機が画像上の物体を一般的な名称で認識することであり，画像認識の研究において最も困難な課題とされる [1, 2, 3].

一般物体認識の実現のためには，様々な物体を学習対象とすると共に，対象の姿勢や周囲の環境の変化といった膨大なアピアランスを学習できる画像データセットが必要である．この学習用の画像データセットをどのように構築するかが，一般物体認識の精度と有用性の向上において重要な課題となる．一般物体認識の研究では，実験用に人手で作成したされた画像データセットが用いられてきた [4, 5]．しかし様々なサンプルを含む画像データセットを人手で作成するには多大な人的および時間的コストがかかる．また，画像やラベル付けの内容が作成者に依存してしまい，学習された内容がそのまま一般の画像に対しても有効で無いといった問題がある．そこで近年では，Web から画像を自動的に収集することで，実世界に存在する一般的な画像を大量に得る Web 画像マイニング [6, 7] が注目されており，Web から画像データセットが生成されている．

Web 画像マイニングでは，画像の抽出源として大規模な画像投稿サイトである Flickr[1]が多く用いられる．Flickr から得た画像による一般物体認識の研究はいくつか報告されており [8, 9]，画像につけられたタグを直接教師ラベルとして用いることで学習及び認識を行っている．Flickr などから画像を得た際に問題となるのが，ラベルとなるタグに造語や無意味な文字列といった多くのノイズデータが含まれることである．これまでの研究ではタグが持つ意味については詳細な議論が行われておらず，ノイズデータを出現数や頻度，分散といった統計的な情報でしか排除することができなかった．

本研究では，意味的に階層構造化された画像データセットを Flickr から自動的に生成する手法を提案する．まず，Flickr におけるタグをオントロジー辞書に照合することで，画像データ間の意味的な関係性を得ると共にノイズタグを削除する．そして，各タグのつけられた画像をオントロジー辞書で示される概念構造上に配置することで，画像データセットの階層構造化を行う．意味的な階層構造とは，“犬”や “猫”の上位概念として “動物”が存在するように，一般的に人間が捉えている対象間の関係の集合である．このような構造を持つ画像データセットを用いて学習を行うことで，より人間の画像認識処理に近似した一般物体認識を実現することが期待される．

2　提案手法

本手法では，以下の手順で階層構造化された画像データセットの生成する．

1. 画像情報の取得
2. タグとオントロジーの照合

[1] http://flickr.com/

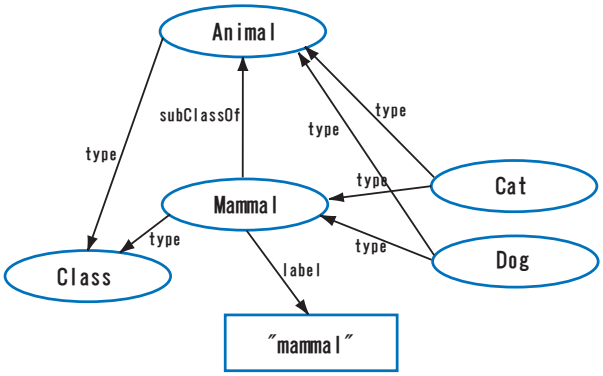

図 1 DBpedia の概念構造の一部

3. クラスとインスタンスの IS-A 関係を用いたタグの選別

4. データセットの階層構造化

まず，オントロジー内のクラスラベルを用いて Flickr に対しキーワード検索を行うことで，それぞれのクラスに関する画像情報やタグを収集する．次に，収集したタグをオントロジー内のインスタンスに照合することで，タグの持つ意味情報を得る．このとき特殊な造語や無意味な文字列のようなノイズとなるタグは排除される．次に，検索に使用したクラスと，インスタンスに対応付けられたタグとの IS-A 関係を用いて，階層構造化に適切なタグを選び出す．最後に，選び出されたタグをオントロジーに従って配置することで，画像データセットの階層構造化を行う．

2.1 オントロジー

本研究ではオントロジー辞書として DBpedia[10] を用いる．DBpedia とは Wikipedia からオントロジーを生成し，RDF と呼ばれる形式で公開しているサービスである．オントロジーの内容や概念同士の関係は，Wikipedia や人手によって定期的にアップデートされているため，オントロジーの更新性に優れる．また，概念同士の関係性が `http://www.w3.org/2002/07/owl#Thing` をトップノードとした明確なツリー構造となっているため，階層構造化に利用しやすい．オントロジーの検索にはクエリ言語 SPARQL を用いる．

RDF のデータは subject，predicate，object のトリプルで表現され，各概念は URI によって一意に表現される．DBpedia における概念構造の例を図 1 に示す．ただし図中では，簡略化のため URI の末尾のみを記述している．同図中に楕円で示されているものは各概念であり，矢印によって概念同士の IS-A や上下関係などが示される．同図中の *Mammal* や *Animal* のように *Class* と IS-A 関係にある概念をクラスと呼び，*Dog* や *Cat* をクラスのインスタンスと呼ぶ．図 1 では，*Mammal* や *Animal* と *Dog* や *Cat* は IS-A 関係（クラスとインスタンスの関係）にあり，*Mammal* と *Animal* は上位下位の関係であることが示されている．また *Mammal* に対して *label* の関係で結びつけられた文字列をクラスラベルと呼ぶ．提案手法では，画像につけられたタグを *Dog* や *Cat* のようなインスタンスに対応付け，概念同士の関係を得ることでタグの持つ意味情報を利用する．

図 2 [bazuca, bazooka, dog, animal, pet] のタグがつけられていた画像

2.2 画像情報の取得

画像データセットを生成するため Flickr から画像を得る．Flickr から画像を得るには何らかの検索ワードが必要となるが，ある特定のキーワードのみを用いると画像の内容に偏りが生じるといった問題がある．そこで，検索ワードとしてオントロジー内に存在するクラスのラベルを用いることで，各クラスに関連する画像を収集していく．このとき検索に用いたクラスを**検索クラス**と呼ぶ．この方法で画像を得ることで，提案手法の手順 4（2.4 節参照）において適切なタグを選び出すことが可能となる．

2.3 タグとオントロジーの照合

画像につけられたタグをオントロジー辞書の項目と照合する．タグをオントロジー内のインスタンスに対応付けることで，ノイズとなるタグの排除を行う．図 1 における *Dog* や *Cat* といった概念は，正確には `http://dbpedia.org/resource/Dog` や `http://dbpedia.org/resource/Cat` のように URI で記述される．同様に，タグとして付与された文字列を `http://dbpedia.org/resource/` の末尾に付与することで概念との対応付けを行う．このとき該当する概念が DBpedia 内に存在しなかった場合，オントロジー内に存在しない概念であると判断し，ノイズタグとして排除する．

表 1 図 4 の各タグが属するクラス

タグ	属するクラス
animal	*Animal*
cat	*Mammal*
fun	*MusicGenre*
grey	*Color*
italy	*Country*
nikon	*Company*

2.4 クラスとインスタンスの関係を用いたタグの選別

インスタンスとの対応が取れたタグの中で，検索クラスそのものを示すタグ，または，検索クラスの下位クラスと IS-A 関係にあるタグを選び出す．図 2 は，*Animal* クラスを検索クラスとした画像であるが，“bazooka”のようなタグがつけられていた．このような場合，*Animal* に関する画像が *Weapon* クラスなどに属する画像として処理されてしまい，学習時のノイズデータとなってしまう．そのため，この例の場合では，検索クラスである *Animal* と IS-A 関係にある “dog”や “animal”のみを階層構造化時に参照することで，画像データセットの適合率の向上を図る．

図 4 に, 具体的な処理の例を示す．検索クラス *Animal* によって Flickr から図 4 を得ると，一番左端のように 19 種類のタグが付与されていた．その中でオントロジーとの照合に成功したタグは中央の 6 種類であり，各タグと直接 IS-A 関係にあるクラスは表 1 の通りである．このオントロジーの項目と照合に成功した 6 種類のタグの中で，検索クラスに対して適切と判断されるのは，検索クラスそのものである *Animal* と，*Animal* の下位クラスである *cat* である．したがって，6 種類のタグの内 “animal”と “cat”が階層構造化時に利用される．

2.5 画像データセットの階層構造化

これまでの手順によって選び出されたタグを，オントロジー辞書に従って配置することで画像データセットの階層構造化を行う．階層構造化を行った画像データセットのイメージを図 3 に示す．

図 3 のように，*Animal* などのクラスにそれぞれタグが属し，タグには画像が紐付けられる．このとき，検索クラスとの IS-A 関係で結ばれた適切である可能性が高いタグのみを利用しているため，画像データの不適切な配置（例えば，Weapon クラスに犬の画像が含まれるなど）は可能な限り抑えられている．

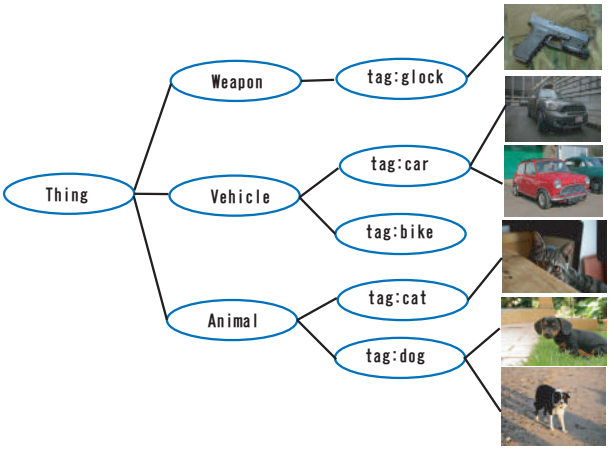

図 3 階層構造化された画像データセットのイメージ

3 階層的画像データセット生成

Flickr から画像を収集し，階層構造化された画像データセットの自動作成を行った．検索に用いた検索クラス数は 518 クラスであり，パラメータとして各検索クラス 1 つあたりの画像取得数を最大 4000 枚とする．

3.1 実験データ

各クラスをクエリとした検索の結果, Flickr から 1603542 枚の画像と 612727 種類のタグを得た．提案手法によって，これらの画像とタグを基に階層的画像データセットを作成した．表 2 に，提案手法の各手順におけるデータの遷移を示す．

各手順で，ノイズタグや不適切なタグを排除した結果，タグが残らなかった画像は画像データセットから除去された．そのため，手順が進む毎に画像枚数は減少する結果となった．表 2 より，最終的に画像データセットに採用された画像は 1603542 枚のうち 47910 枚であった．また，612727 種類のタグのうち階層構造化に用いられたタグは 4827 種類であった．

3.2 作成された画像データセット

本実験で作成された画像データセットについて，ラベルと画像の例を図 5～図 9 に示す．図 5～図 7 はタグ（インスタンス）を入力して得られる画像，図 8 と 9 はクラスを入力して得られる画像である．図 5～図 9 より，幅広い分野の画像が適切に集められていることが分かる．ただし図 5 のように，トラの画像に “cat”というタグが付けられるなど Flickr 上のタグ付けに依存した問題も見られた．このようなタグを誤りとして処理するには，今回使用された *Mammal* よりも狭義な概念によって構成されたオントロジーが必要となる．

表 3 に，タグ毎の画像枚数の上位 10 タグを示す．表 3 から，*Animal* や *Plant* などの上位語である *Species* に属するタグの画像が多いことが見て取れる．これは, Flickr

検索結果として得た状態

[ambra, animal, cat, colors, cute, eyes, freetime, fun, gatto, grey, house, italy, kitten, moment, nikon, pet, shadow, shot, wood]

概念との対応付けが取れたタグのみが残された状態

[animal, cat, fun, grey, italy, nikon]

検索クラスとの関係によって適切なタグを残した状態

[animal, cat]

図 4 各手順によるタグと画像の変化の例（検索クラスは *Animal*）

表 2 各手順におけるデータの変化

	画像枚数	タグの種類数	画像あたりの平均タグ数
Flickr からの検索結果	1603542 枚	612727 種類	15.01
オントロジーとの対応付けの結果	835598 枚	30250 種類	3.85
クラスとインスタンスの関係により適切なタグを残した結果	47910 枚	4872 種類	1.55

図 5 cat タグの画像例

に動物の画像が多く投稿されていることや，画像中の動物に対してタグが明確につけられやすいことが理由と考えられる．

クラスの中には図 10 の *Language* のように，画像を当て嵌めるのが困難と思われるクラスも存在していた．このようなクラスの場合，画像の内容には統一性がなく，一般物体認識に用いる学習用データとしては不適切なデータとなった．

本実験で作成された画像データセットの階層構造の様子を図 11 に示す．ここで，図 11 はツリーの一部のみを展開した状態である．左端のトップノードに位置するクラスが `http://www.w3.org/2002/07/owl#Thing` である．*Thing* をトップノードとしてタグを各ノードや末端に持つツリー構造を見て取ることができる．階層構造内のクラス数は，提案手法の処理過程を経て，検索クラス数の 518 クラスから減少して 183 クラスとなった．

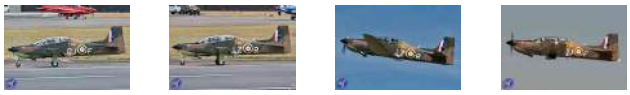

図 6 tucano タグの画像例

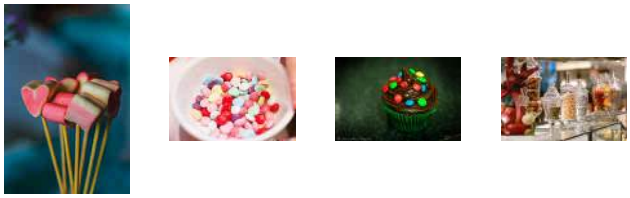

図 7 candy タグの画像例

4 評価実験

画像データセット中の画像に対するラベル付けの精度について評価実験を行った．

4.1 評価方法

まず，実験では，人間がラベルと画像の対応を判断できないと考えられる *Place*, *Time Period*, *Topical Concept*

図 8 *Animal* クラスの画像例

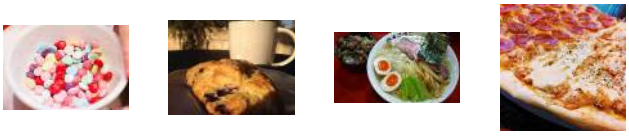

図 9 *Food* クラスの画像例

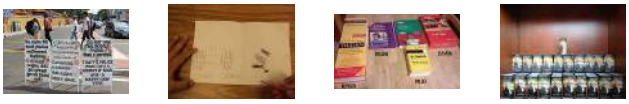

図 10 English タグ (*Language* クラス) の画像例

表 3 タグごとの画像枚数の上位 10 タグ

タグ	画像枚数
spider	3172
cycad	1507
moss	1445
animal	1410
crustacean	1408
fern	1311
person	1136
fungus	1065
snail	798
moon	758

に関するクラスやタグを対象から除外した．そして，画像が 100 枚以上収集できたラベルを用いた．この条件を満たす 121 種類のタグと 87 種類のクラスの中から，ランダムに 20 種類のタグと 10 種類のクラスを選び出した．そして，これら計 30 個のキーワード（20 種類のタグと 10 種類のクラス）を入力して得られる画像から，各 100 枚をランダムに抽出した．図 12 に評価テストに用いる Web ページを示す．図 12 のようにテストはブラウザ上で行い，ページの最上部にはこのページでテストするラベル名が表示される．ページには 100 枚の画像と画像の下にチェックボックスが配置され，被験者はラベル名に対して適切でないと感じた画像にチェックを入れる．各ラベルが何を示す単語であるかについて，被験者間での認識の統一化のため，タグやクラスについての Wikipedia の説明文を記載した．本稿では，各ラベルの画像 100 枚のうち，被験者 7 人のうち 6 人以上が適切な画像と判断した画像の割合（適合率）を評価値とする．

4.1.1 評価結果

20 種類のタグについて評価結果を表 4 に，10 種類のクラスについての評価結果を表 5 に示す．表 4 に示したタグについての評価実験結果では，“conidae”や “lizard” など動物を示すタグについて高い適合率が確認された．

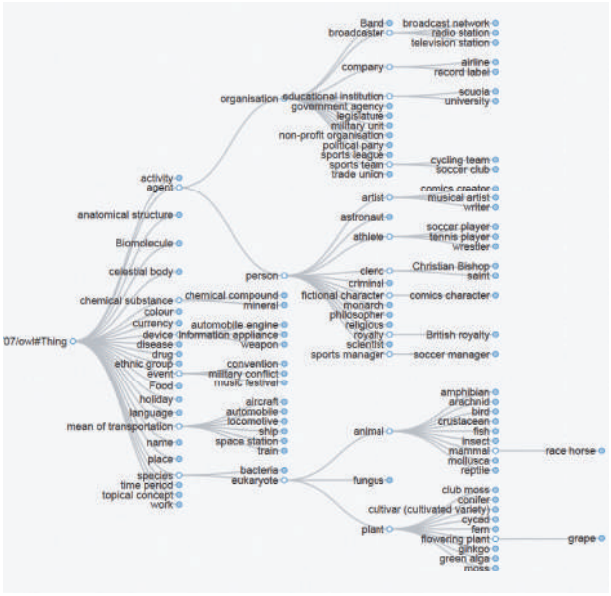

図 11 画像データセットの持つ階層構造

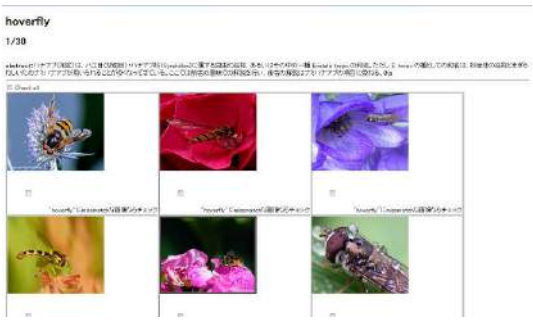

図 12 評価テスト用の Web ページ

“quartz”, “green”, “earth” といったタグでは低い適合率となったものの，実験に用いたタグの平均は 62.89%と比較的高い値が示された．“green”や “earth” は画像の内容を示すタグとしては曖昧な語であり，統一性の無い様々な画像が存在したことが，低い適合率が示された要因と考えられる．例えば，“earth” の場合，惑星としての地球やそれをデフォルメしたもの，単に自然の風景を撮影した画像や月の画像など様々であった．また “quartz” の場合，鉱石そのものの画像よりも採掘地の画像が多かった．“fencing” というタグについては，特定の物体を指すタグでないにも関わらず，高い適合率が確認された．画像の多くはフェンシングを行う競技者を撮影した画像であり，適切な画像が収集されたと言える．

表 5 に示したクラスについての評価実験結果では，タグの評価実験結果に比較して，平均の適合率は 15%ほど高い結果となった．これはクラスの扱う概念がタグに比べて広義であるためと考えられる．例えば，図 5 と図 8 におけるトラの画像は，タグ (“cat”) という解釈では不適切な画像となるが，クラス (*Animal*) と捉えれば適切

表 4 タグについての評価試験結果

タグ	適切な画像の割合 (%)
animal	82.0
cat	64.0
coffee	53.0
conidae	90.0
cypraeidae	78.0
dog	79.0
earth	31.0
elephant	59.0
fencing	69.0
fern	65.0
green	25.0
lion	70.0
lizard	89.0
moth	62.0
pine	79.0
plant	52.0
pulmonata	76.0
quartz	19.0
spider	53.0
AVG	62.9

表 5 クラスについての評価試験結果

クラス	適切な画像の割合 (%)
amphibian	98.0
animal	80.0
arachnid	66.0
flowering plant	85.0
fungus	90.0
insect	86.0
mammal	83.0
mollusca	66.0
plant	70.0
sport	48.0
AVG	77.2

な画像となる．

5 おわりに

本稿では，オントロジーを利用することで Web から取得した画像を基にした階層的画像データ生成手法を提案した．生成した画像データベース中の，画像とラベルの対応について評価実験を行ったところ，自動で生成した画像データベースとしては，比較的高い適合率が示され，一般物体認識のためのデータベースとしての実用性が示唆された．

今後は，階層構造を利用した学習アルゴリズムを構築し，提案手法によって生成した画像データベースを一般物体認識に応用していく．

謝辞

本研究の一部は科研費基金基盤研究 (C)(#26330210) 及び中部大学研究費による．ここに感謝申し上げる．

参考文献

[1] Robert Bergevin and Martin D. Levine. Generic object recognition: Building and matching coarse descriptions from line drawings. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 15, No. 1, pp. 19–36, 1993.

[2] Yann LeCun, Fu Jie Huang, and Leon Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, Vol. 2, pp. II–97. IEEE, 2004.

[3] Andreas Opelt, Axel Pinz, Michael Fussenegger, and Peter Auer. Generic object recognition with boosting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 28, No. 3, pp. 416–431, 2006.

[4] Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, Vol. 1, pp. 1–2, 2004.

[5] Gregory Griffin, Alex Holub, and Pietro Perona. Caltech-256 object category dataset. 2007.

[6] Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yantao Zheng. Nus-wide: a real-world web image database from national university of singapore. In *Proceedings of the ACM international conference on image and video retrieval*, p. 48. ACM, 2009.

[7] 柳井啓司. 一般画像自動分類の実現へ向けた world wide web からの画像知識の獲得. 人工知能学会論文誌= Transactions of the Japanese Society for Artificial Intelligence: AI, Vol. 19, pp. 429–439, 2004.

[8] 中山英樹, 原田達也, 國吉康夫. 大規模 web 画像のための画像アノテーション・リトリーバル手法. 画像の認識・理解シンポジウム (MIRU), pp. 103–110, 2009.

[9] Xirong Li, Cees GM Snoek, and Marcel Worring. Unsupervised multi-feature tag relevance learning for social image retrieval. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, pp. 10–17. ACM, 2010.

[10] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, et al. Dbpedia-a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal*, 2013.