

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 5 日現在

機関番号：14401

研究種目：挑戦的萌芽研究

研究期間：2012～2013

課題番号：24650123

研究課題名(和文) 画像による言語的意味のグランディングの可能性の探求

研究課題名(英文) Exploring the grouping of linguistic concepts in images on the Web

研究代表者

林 良彦 (Hayashi, Yoshihiko)

大阪大学・言語文化研究科(研究院)・教授

研究者番号：80379156

交付決定額(研究期間全体)：(直接経費) 2,900,000円、(間接経費) 870,000円

研究成果の概要(和文)：本研究は、言語的な概念を画像により表現する可能性を探求することを目的とする。本研究期間においては、ある具体的概念が有する意味特徴を適切に表すための画像をWebから収集するためのクエリパターンについて検討するとともに、どのような概念と意味特徴の組み合わせ(複合概念)に対して適切なWeb画像が収集できるかの調査を行った。その結果、クエリにおいて動詞の進行形を利用することが有効であること、動物の摂食行動や、物理的動作に対するWeb画像が比較的高い適合度を示すことが分かった。また、適合画像に対して独立に付与した言語注釈と複合概念表現の間の類似度と画像適合度の間に中程度の相関関係があることを確認した。

研究成果の概要(英文)：The recent research direction toward multimodal semantic representation would be further advanced, if we could have a machinery to collect adequate images from the Web, given a target concept. With this motivation, this research particularly investigated into the Web imageabilities of the behavioral features of a basic-level concept. The primary contributions made in this research are twofold: (1) "beaver building dams"-type queries can better yield relevant Web images, suggesting that the present participle form ("-ing" form) of a verb, as a query component, is more effective than the base form; (2) the behaviors taken by animate beings are likely to be more depicted on the Web, particularly if the behaviors are, in a sense, inherent to animate beings (e.g., motion, consumption). We further analyzed linguistic annotations that were independently given to some of the images, and discusses an aspect of the semantic gap between image and language.

研究分野：総合領域

科研費の分科・細目：図書館情報学・人文社会情報学

キーワード：画像が伝える意味 Web画像 グランディング 言語注釈

1. 研究開始当初の背景

単語が指示するアトミックな概念、さらには、語句により想起される複合的な概念が画像や記号により視覚的に提示できれば、インタラクティブな情報アクセスにおいて、ユーザの意図の明確化を支援する手段として利用できると考えられる。特に、画像や記号が(ある程度までは)言語独立であることを考えれば、言語横断情報検索のような「言語の壁」を越える必要があるタスクにおいて、有用な手段となると考えられる。

一方、言語の意味を画像などの言語以外のメディア情報と対応付けたり、画像から抽出した高位の属性を意味表現に適用するといったマルチモーダルな意味処理に関する研究が活発化しつつある。ある状況において特定の意味を適切に表現すると解釈される画像が、別の状況においても同様の意味を想起させるものとして解釈されるとは限らないので、このような内容(content)と解釈(interpretation)のギャップがどのような意味概念、あるいは、画像において起こりうるかを調べておくことは、言語的意味の画像によるグランディング、言語と画像の対応付けの向上のために重要である。

2. 研究の目的

本研究は、言語的な概念を画像により表現する可能性を探求することを大きな目的とし、本研究期間においては、特に以下の2点についての見通しを得ることを目的とする。

- a. **Web 画像による概念意味特徴の表示:** ある具体的概念が有する個々の意味特徴を適切に表すための画像を Web から収集するためのクエリパターンを明らかにするとともに、どのような概念と意味特徴の組み合わせ(複合概念)に対して適切な Web 画像が収集できるかについての見通しを得る。
- b. **概念意味特徴の画像に対する解釈の分析:** もとの概念意味特徴に対して適合度の高い画像が独立に提示されたときにどのような言語的意味を想起するかを調査し、言語と画像間の意味ギャップを探求するための手がかりを得る。

3. 研究の方法

上記の a, b それぞれの研究項目に分けて説明する。

a. Web 画像による概念意味特徴の表示

本研究項目の全体構成を図 1 に示す。

(1) McRae らによって、英語を対象に、被験者実験により心理言語学的な立場から作成された意味属性規範 (semantic feature norm) を収集したデータベース (541 の基本概念と 2,576 の意味属性により 7,526 の概念意味特徴を収録) から、生物の動作、無生物の振る舞いを表す概念意味特徴 535 件(概念名詞は 218 件)を抽出した。

(2) これらから既存の画像検索エンジン

(Google Images を使用) を利用して Web 画像を収集するためのクエリ(計 2,126 件)を以下のパターンにより生成した。

- q0:概念名詞のみ、
- q1:概念意味特徴そのまま、
- q2:q1 における動詞を進行形(-ing 形)に変更したもの、
- q3:q1 における概念名詞を後置した名詞句に変換したもの

(3) 収集された Web 画像 (27,970 点)に対して、1 名の評定者により適合度(0:適合なし~4:最もよく適合)を付与した。

(4) 得られた評定値に対して、情報検索の分野で使われる指標 (Average Precision (AP) と Normalized Discounted Cumulative Gain (NDCG))を計算し、ある概念意味特徴に対して収集された Web 画像の適合度 (Web 画像度)を計算した。AP, NDCG とともに値が大きいほど、その概念意味特徴に対して適切な Web 画像が得られる傾向にあることを示す。

(5) 上記で計算した指標と、クエリ生成パターン間の関係、概念名詞、意味特徴に使用した動詞、の意味分類的特徴との間の関係を分析した。

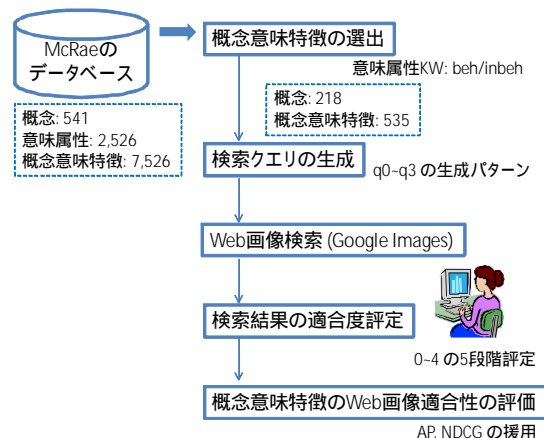


図 1: Web 画像による概念意味特徴の表示に関する検討

b. 概念意味特徴の画像に対する解釈の分析

本研究項目の全体構成を図 2 に示す。

(1) 上記の研究項目で得られたデータの中から、NDCG による Web 画像適合度が比較的高かった(0.4 以上) 293 件の概念意味特徴に対する 3,653 点の画像を分析対象として抽出した。

(2) 日本語、英語を母語とするそれぞれ 2 名の評定者(上記の適合度評定には関わっていない)により、画像の解釈に関する言語注釈を付与してもらった。ここで、語彙を制限することはしなかったが、文型はオリジナル概念意味特徴に従うよう指示した。

(3) オリジナルの概念意味特徴と評定者による言語的注釈の間の意味的類似度を計算し、これらと Web 画像適合度の間の相関関係を調べた。なお、意味的類似度は、概念名詞、動詞、目的語名詞、それぞれの間の類似度を

求め、これらを重み付き和により統合することにより計算する．重回帰分析(Lasso 法)，および，サポートベクトル回帰 (SVR) を適用することにより，それぞれの要素の重みを最適化するとともに，それらの影響の度合いを調べた．

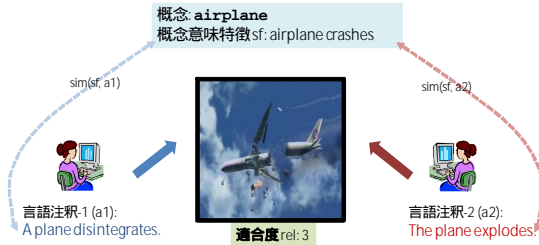


図 2: 概念意味特徴に対して得られた Web 画像に対する他者の解釈の分析

4. 研究成果

a. Web 画像による概念意味特徴の表示

- 適合性評定値の分布: 表 1 に示す．半分近く (44.5%) の画像が何らかの適合性を有していたことが分かった．以下，AP を用いた分析では，適合度が 1 以上の画像を適合画像とみなす．
- クエリの生成パターン: 図 3 に AP，NDCG の結果を示す．この結果から，q2 パターンによるクエリ(動詞-ing 形)がもっとも性能が良いことが分かる．画像はある状況のスナップショットであると考えられ，これに対する言語記述としては進行形を用いた表現が多く使われていることが想像される．
- 概念意味特徴のタイプ: 図 4 に結果を示す．McRae のデータベースにより視覚的であるとされた意味特徴はやはり高い傾向にあり，さらに，生物の動作は無生物の振る舞いよりも高い傾向にある．
- 概念名詞の意味分類: 図 5 に結果を示す．動物(animal)が両指標で高い値を示しているのに対し，人工物(artifact)は低い傾向にある．
- 動詞の意味分類: 図 6 に結果を示す．動作(motion)や摂食行動(consumption)は高い傾向にあり，創造行動(creation)は低い傾向にある．動作の主体は動物であることが多く，この結果は概念名詞に関する結果と符号する．
- 概念名詞・動詞の組み合わせ: 図 7 に結果を示す．上記までの結果から推測されるように，動物の動作や摂食行動の適合度は比較的高い傾向にある．

適合度	0	1	2	3	4
画像件数	6,705	8,734	4,025	1,910	6,596
総計: 27,970	15,439 (55.2%)	12,531 (44.8%)			

表 1: 画像適合度の分布

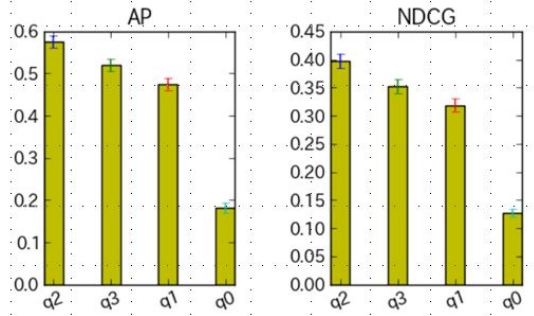


図 3: クエリ生成パターンと適合度

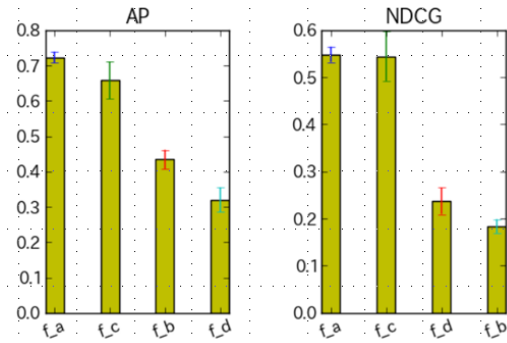


図 4: 概念意味特徴タイプと適合度

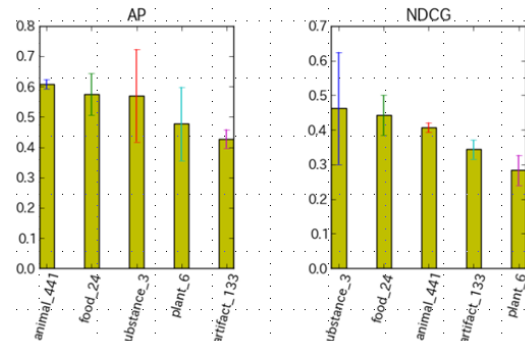


図 5: 概念名詞の意味分類と適合度

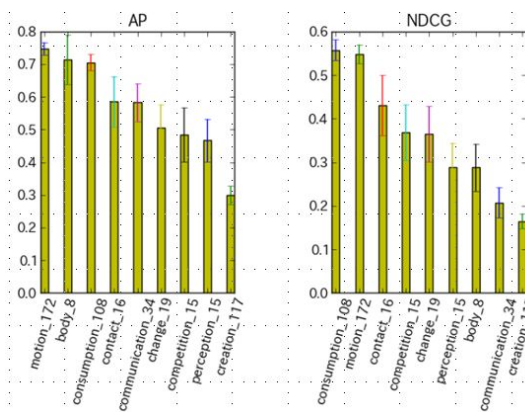


図 6: 動詞の意味分類と適合度

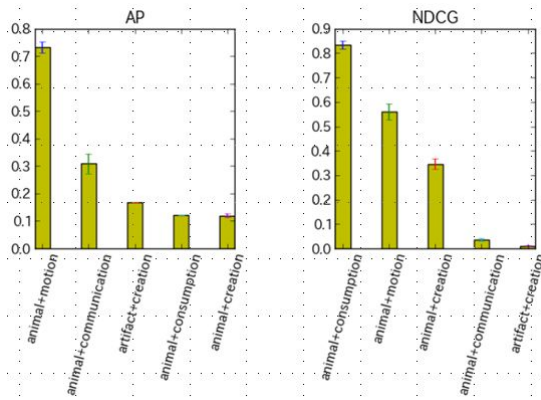


図 7: 名詞・動詞の組み合わせと適合度

b. 概念意味特徴の画像に対する解釈の分析

相関関係の分析結果を表 2 に示す。ここで、走行条件として、A~D の 4 つの場合を考慮した。

走行条件	Lasso 法		SVR		
	rel	NDCG	rel	NDCG	rel [†]
A: 評定者:a+b, 文字列類似度:あり	0.474	-0.549	0.574	-0.511	0.528
B: 評定者:a+b, 文字列類似度:なし	0.472	-0.549	0.556	-0.497	0.513
C: 評定者:a, 文字列類似度:なし	0.437	-0.556	0.481	-0.535	0.471
D: 評定者:b, 文字列類似度:なし	0.399	-0.561	0.425	-0.619	0.403
E: a, b 間類似度のみ	0.341	-0.552	0.371	-0.602	0.354

表 2: 相関関係の分析結果

- A: 2 名の評定者の評定を使用。類似度において文字列類似度も併用
- B: 2 名の評定者。文字列類似度は非使用
- C: 評定者 a の評定のみを使用。類似度において文字列類似度は非使用
- D: 評定者 b の評定のみを使用。類似度において文字列類似度は非使用
- E: (参考) 評定者 a, b の評定間の類似度のみを使用。

表 2 において、rel とは 1 点ごとの画像を単位とし、適合度の評定値系列と類似度に基づく評定値の予測値系列の相関係数を示す。rel に対する結果から以下のことが言える。

- 言語表現の類似度と画像評定値には中程度の正の相関($p < 0.001$)がある。
- 表現の文字列類似度の影響はさほど大きくない
- 評定表現の冗長性(評定者 a+b) は相関を高めるのに有効
- 評定者間の言語表現類似度だけから(走行条件 E) ,画像適合度を予測するのは難しい

一方、表 2 において NDCG とは、各概念意味特徴を単位とした時の NDCG 値と画像群に対する予測値の 2 乗平均誤差の間の相関係数を表す。NDCG に対する結果から以下のことが言える。

- NDCG 値と平均予測 2 乗誤差の間には、中程度の負の相関 ($p < 0.001$) がある。
- すなわち、概念意味特徴に対する Web 画像適合性が低い、すなわち、適合画像の Web からの取得が「難しい」ほど、その画像に対する第三者の言語注釈はオリジナルの概念意味特徴と類似していない傾向にある。

一方、各要素(概念名詞, 動詞, 目的語名詞)の重みに対する結果を表 3 に示す。

	a:sim1	a:sim2	a:sim3	b:sim1	b:sim2	b:sim3
Lasso	-0.273	1.011	0.060	0.193	0.685	0.0
SVR	-3.446	0.258	-0.346	3.360	1.638	-3.387

表 3: 各要素の重み

ここで、a, b は評定者の別, sim1:概念名詞, sim2:動詞, sim3:目的語名詞を表す。この表からは、以下の傾向が観察される。

- 評定者 a, b ともに: 動詞要素 (sim2) の重みが大きい。これは、同一、または、意味的類似度の高い動詞の選択が相関関係に大きく影響することを示している。
- 評定者 a, b ともに: 目的語要素 (sim3) の重みは小さい。これは、概念意味特徴において、自動詞構文が多いことが影響している。
- 主語となる概念名詞(sim1) の重みが評定者 a, b で異なる。評定者 a はオリジナルと同じ名詞を用いる傾向があった。

c. 考察

概念意味特徴(すなわち検索クエリ) に対して得られた画像集合から計算される NDCG 値と予測値の平均二乗誤差を乗じた値が大きいほど、検索された画像集合の概念意味特徴に対する適合度が高いにもかかわらず、第三者は別の見方をしうる可能性があることを示す。

今回の結果に関して、このような概念意味特徴を調査してみたが、一定の明確な傾向が得られるまでには至っていない。そこで、以下ではいくつかの例を示し、そこから示唆される傾向について考察する。

図 8(a) は、"cheetah hunts" という概念意味特徴に関して、適合度:4 と判定された画像であるが、"cheetah chases prey", および、"cheetah runs" という言語注釈が与えられている。hunt をするにはその対象(獲物)が存在するので、前者の評定者はそのような状況になるべく述べようとしたと推測できる。一方、後者の評定者は行為の対象よりもターゲットの概念(cheetah)の動作(run) に焦点を当てたと思われる。

図 8(b) は、"faucet leaks" という概念意味特徴に関して、適合度:4 と判定された画像であるが、"water drips", および、"The faucet drips" という言語注釈が与えられている。まず、何をターゲット概念(すなわち主語) と捉えるかが両者で異なっている。また、動詞はいずれも"drip"が用いられているが、これは、「液体が滴る」という一般的な物理現象を描写するのに対し、そもそもの概念意味特徴における"leak"は「蛇口・水道などから水漏れが生じている」という事態を描写しており、それらの意味レベルには差がある。

以上のように、評定者は画像に描写されている状況をなるべく忠実に描写しようとする

るが、注釈の表現がバリエーションを持つことは避けられないと言える。

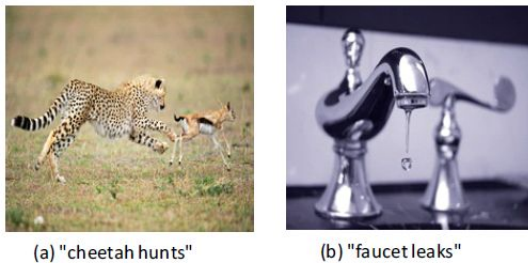


図 8 : 適合画像の例

d. まとめ

まず、a の研究項目に関しては、動物の摂食行動や、主に物理的な動き・移動に関した基本行動が Web 画像により表現されやすいことを確認した。また、英語による概念意味特徴において、より良い画像検索結果を得るためのクエリ生成手法について評価を行い、動詞の-ing 形を用いることが有用であることを確認した。しかしながら、情報アクセスシステムに組み込める形の支援機能を実現するためには、ある概念意味特徴に基づいた検索クエリによって検索された画像が、その概念意味特徴をどの程度適切に表しているかを予測する手法の開発が必要であり、Web コンテンツの言語的な解析と画像解析を統合した手法の開発が必要である。

一方、b の研究項目に関しては、画像にどのような解釈を与えるかは個人による差が大きく、対応する言語注釈も豊富なバリエーションを有すること、このようなバリエーションには、ターゲットとなる概念の特性、画像に描写される行い・振る舞いの典型性、あるいは、顕在性、ターゲットと同時に描写されている事物の特性などのほか、画像の画像的な特徴（構図や動きの表現など）も大きく関わっていることが示唆された。

今後は、図像的 (iconographic) な画像に対する言語注釈が複雑な言語構造を持ちうるという知見も考慮しながら、言語と画像の意味ギャップ、言語的意味の画像によるグランディングに対する研究を深めていきたい。ここでは、頑健な画像パラメータによる画像的特徴の分析や、動作・振る舞いに関する言語・世界知識を統合していく必要があると考えられる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔学会発表〕(計 4 件)

(1) Yoshihiko Hayashi. Web-imageability of the Behavioral Features of Basic-level Concepts. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pp. 3609-3614. 2014 年 5 月 30 日, レイキャビク市・アイスランド. (査読有).

http://www.lrec-conf.org/proceedings/lrec2014/pdf/627_Paper.pdf

(2) 林 良彦. 概念意味特徴の画像による表示とその解釈に関する分析. 言語処理学会第 20 回年次大会, B1-4, 2014 年 3 月 18 日, 北海道大学 (札幌市).

(3) 林 良彦. Web 画像による語義・概念の視覚的な提示に関する検討. 人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会 (第 5 回), 2013 年 10 月 25 日, 慶応義塾大学 (横浜市).

(4) Yoshihiko Hayashi. Migrating Psycholinguistic Feature Norms into Linked Data in Linguistics. *Proceedings of 2nd Workshop on Linked Data in Linguistics (LDL-2013)*, pp.70-75. 2013 年 9 月 23 日, ピサ市・イタリア. (査読有).

<http://www.aclweb.org/anthology/W13-5511.pdf>

6. 研究組織

(1) 研究代表者

林 良彦 (HAYASHI, Yoshihiko)

大阪大学・大学院言語文化研究科・教授

研究者番号：80379156