

名詞比喩表現における解釈の検索モデル

岡 隆之介[†] 楠見 孝[†] 大島 裕明^{††}

[†] 京都大学大学院教育学研究科 〒606-8501 京都府京都市左京区吉田本町

^{††} 兵庫県立大学大学院応用情報科学研究科 〒650-0047 兵庫県神戸市中央区港島南町 7-1-28

E-mail: †{oka.ryunosuke.45s,kusumi.takashi.7u}@kyoto-u.ac.jp, ††ohshima@ai.u-hyogo.ac.jp

あらまし 本研究では名詞比喩表現における解釈の検索モデルの提案と評価を行う。名詞比喩表現（例：笑顔は花のようだ）に対して、人が生成するような自然な解釈（例：華やかだ）を検索するモデルを作ることは、自然言語処理分野において注目を集めている課題の1つである。この課題を解く際には、人がどのように名詞比喩表現を理解しているのかに関する知見を踏まえてモデルを検討する必要がある。本研究では、人の名詞比喩表現に対する解釈生成過程を考慮した様々な計算モデルの性能を評価する。これまで提案されてきた Kintsch の Predication モデルを中心として、そのバリエーションについての評価を行う。実験に用いるコーパスは現代日本語書き言葉均衡コーパスである。正解データは、著者らがこれまでに被験者実験によって得られたデータから、参加者から得られた解釈に含まれる形容詞を用いて作成する。計算モデルに与えられた比喩表現に対して、出力された形容詞のランキングを基にして、検索性能を得ることでモデルの評価を行う。

キーワード 比喩表現, 解釈, 検索モデル, Predication モデル

1. はじめに

比喩表現はあるものを別のものにたとえて伝える言語表現である。たとえば、「笑顔は花のようだ」は比喩表現である。このとき、「笑顔」は主題、「花」は喩辞と呼ばれる。一方、「笑顔は美しい」という表現は比喩表現ではなく、字義表現と呼ばれる。字義表現では、主題（笑顔）が持つ特徴の側面（美しさ）が明示されるのに対して、比喩表現ではそれが明示されない。そのため、人が比喩表現の意味を理解する際には、その比喩表現が主題のどのような特徴について説明しているのかを考える必要がある。たとえば、「笑顔は花のようだ」という比喩表現を理解する場合には、「笑顔は美しい」という解釈であると想起することが求められる。

本研究では、比喩表現の意味を理解することを、比喩表現の解釈を検索する問題であるととらえる。事実、人はこの問題を解くために、主題と喩辞の言葉の意味を記憶から検索し、何らかの計算を踏まえて解釈を検索していることが知られている [3] [2] [7] [4]。また、本研究では比喩表現の解釈は形容詞に絞る。ここで、形容詞とは「美しい」や「明るい」や「白い」といった語のことを指す。つまり、本研究の課題は、「笑顔は花のようだ」という比喩表現から、「美しい」や「明るい」といった形容詞を検索するモデルを評価することである。

計算機によって比喩表現の意味理解を実現する場合にも、人と同様の過程を踏まえて問題を解くことができると考えられる。すなわち比喩表現の意味を、人の記憶のような知識表現に対して、何らかの計算を行うことで、比喩表現の解釈を得ることができるだろう。人が比喩表現の解釈を検索する際に行っていると考えられる理解の過程を参考にして、計算機による比喩表現の意味理解を実現することは、人らしい比喩表現の解釈を検索する上で有用であると考えられる。本研究では、名

詞比喩表現の主題と喩辞を入力とし、名詞比喩表現の解釈を出力とする検索モデルの提案と評価を行う。より具体的には、はじめに主題と喩辞のそれぞれの語の意味を、日本語コーパスに基づいて生成される n 次元の特徴ベクトルとして求める。そして、この主題と喩辞のベクトルに対して、2種類の計算モデル（平均モデルと Predication モデル [6]）を適用し、それぞれの方法によって計算される比喩表現の特徴ベクトルを求める。この比喩表現ベクトルと、コーパス中の全ての形容詞のベクトルとのコサイン類似度を求め、この類似度のランキングを比喩表現の解釈とする。

モデルによって検索される比喩表現の解釈から、モデルの性能を評価するために、本研究では岡ら [12] で収集された、120個の直喩表現 [14] に対して参加者 ($N = 50$) が生成した解釈に含まれていた形容詞を用いる。モデルの検索性能の評価は、MRR と MAP と $nDCG@10$ を用いる。

本論文の構成は以下の通りである。2節では、比喩表現の解釈を検索することに関して、人の比喩表現の解釈の検索過程について明らかになっていることを整理した上で、Kintsch の Predication モデルについて説明する。3節では、本研究の提案手法について述べる。4節では、データセットと実験結果について述べる。5節では、まとめと今後の課題について述べる。

2. 関連研究

本節では、人の比喩表現の解釈の検索に関わる先行研究と、本研究で用いる Predication モデルを紹介する。

2.1 名詞比喩表現

まず、比喩表現の解釈についての研究を紹介する。ここでは名詞比喩表現（「〈名詞〉は〈名詞〉のようだ」という形式の比喩表現）を対象とした解釈の検索モデルに絞る。心理学において、比喩表現の解釈の検索モデルは、対象とする比喩の種類に

よって、その表現に対する人の理解の仕方は異なることが想定されており [13]、本研究で提案するモデルを異なる種類の比喩表現（たとえば、「時が流れる」のような動詞比喩表現）に当てはめても上手くいくとは限らない。名詞比喩表現以外の研究として、動詞比喩表現の研究については、Shutova ら [9] が詳しい。また、形容詞比喩表現（例：赤い声）の研究については研究が少ないが、Utsumi と Sakamoto [10] があげられる。

2.2 人の比喩表現の解釈の検索過程

Glucksberg と Keysar [3] の類包含モデル (class inclusion model) は、人が比喩表現の解釈を検索する際に、比喩表現を「主題が喩辞のアドホック（その場で作られた）なカテゴリに含まれている」ことを明示した文であるとみなして処理していると主張する。たとえば「笑顔は花のようだ」という比喩表現の解釈を検索する際には、笑顔は花の抽象的なカテゴリに含まれていると捉える。このとき、喩辞の「花」は具体的な花（例：チューリップ）それ自体ではなく、「花に代表されるようなカテゴリ（花のように、美しく、華やかな何か）」のことを指していると考えられる。人は比喩表現をこのように見なすことで、「笑顔は花のようだ」という表現から美しく、華やかであるといった解釈を検索できると主張する。

類包含モデルは、喩辞の単語を、喩辞に代表される抽象的なカテゴリとみなすことで、喩辞と喩辞と類似する意味をもつ単語の意味を適切に評価することができる。一方で、人が比喩表現を理解する際には、喩辞だけでなく、主題の意味についても考慮する必要がある。たとえば、「笑顔は花のようだ」という表現の解釈を検索する場合と、「扇風機は花のようだ」という表現の解釈を検索する場合では、どちらも同じ喩辞を共有しているものの、検索されるべき特徴は異なってくるだろう。すなわち、前者は「華やか」という解釈が適切であり、後者は「羽根がある」が適切であろう。いくつかの心理学的な実験は、人が比喩表現を理解する際には、喩辞の特徴のなかで主題と関連する意味のみが検索されている可能性を示唆している [2] [4]。

以上の先行研究は、人が比喩表現を理解する際に、(i) 喩辞を、喩辞の単語に代表されるカテゴリとしてみなす（すなわち、喩辞だけでなく、喩辞の意味と関連が深い周辺語についても考慮する必要がある）ことと、(ii) 喩辞の中でも主題と関連のある意味だけが検索されることを示唆している。Kintsch [7] の Predication モデルは、人の比喩表現の理解における上述の特徴を適切に反映したものとなっている。しかし、Predication モデルの要素をどのように組み合わせることが適切かや、最適なパラメータの組み合わせがどのようになるかは検証されてこなかった。そこで、本研究では Predication モデルを基にして様々なモデルとパラメータを検証することを目指す。

2.3 Predication モデル

Kintsch [7] の Predication モデルは文の述部（例：サメ）の意味を主部に応じて適切に特徴ベクトルとして表現するためのモデルである。たとえば、「弁護士は獐猛だ」という文の場合、主部は「弁護士」であり、述部は「獐猛だ」である。特に、本研究のように Predication モデルを名詞直喩表現に適用する場合には、主部は主題（例：弁護士）であり、述部には喩辞（例：

サメ）である。ここで、「弁護士はサメのようだ」という名詞直喩表現を解釈する場合には、喩辞は「弁護士はサメのように泳げる」のようなサメそのものの特徴というよりも、「弁護士はどう猛である」のような特徴を持っていると解釈するのが正しいだろう。Predication モデルでは、(i) 単語の意味の特徴ベクトル化と、(ii) Predication アルゴリズムの適用によって、文の意味を特徴ベクトルとして表現する。(i) 単語の意味の特徴ベクトル化は、大規模なテキストコーパスから単語-文書（あるいは段落）行列を生成し、この行列に対して潜在意味解析 [1] を適用することで実現する。すなわち、ある単語の意味を、文書（あるいは段落）から抽出された高次元の特徴ベクトルとして表現する。(ii) Predication アルゴリズムの適用では、述部（例：サメ）と意味的に関連の深い単語（例：魚、ヒレ、獐猛）の中で、主部（例：弁護士）とも意味的に関連の深い単語（例：獐猛）を検索する。そして、それらと主部と述部の意味を表す特徴ベクトルの重心を、文の意味を表す特徴ベクトルとする。Predication モデルは比喩表現のみならず、文に基づく因果推論や、同音異義語の意味の推定や、文の意味の類似性の判断を適切に表現できるとされている [7]。

3. 名詞比喩表現の計算モデル

本研究では、語や、語の組を特徴ベクトルで表現する。そこで、まず、コーパスに基づいてある語をどのように特徴ベクトルにするかについて説明する。続いて、本研究で実装した 2 種類の計算モデル（平均モデルと Predication モデル）のそれぞれで、比喩表現の意味がどのように特徴ベクトルとして表現されるのかについて手法の概要を説明し、手法の各ステップの詳細について述べる。

3.1 語の意味の特徴ベクトルによる表現

先行研究では、語の意味を特徴ベクトルとして表現する方法として、テキストコーパス中の単語と文書（あるいは段落）の共起関係を利用する手法が提案されている [1]。本研究では [1] にならない、大規模なテキストコーパスの単語数 (n_{word}) と文書の段落数 ($n_{paragraph}$) に基づいて作成される、単語-文書段落頻度行列 ($n_{word} \times n_{paragraph}$) を利用する。そして、この単語-文書段落頻度行列に対して潜在意味解析を適用することで、行列を文書段落の特異値の高い上位 D 次元に分割し、単語-文書段落頻度行列 ($n_{word} \times D$) と文書段落文脈行列 ($D \times D$) と文書段落文脈-文書段落頻度行列 ($D \times n_{paragraph}$) に分解する。この手続きによって生成される単語-文書段落頻度行列と文書段落文脈行列を掛け合わせたものを、単語の意味空間として表現する。この意味空間上において、ある語の意味は、 D 次元の特徴ベクトル w として表現される。

3.2 平均モデル

平均モデルは、主題の特徴ベクトルと喩辞の特徴ベクトルを平均した特徴ベクトルを比喩表現の意味とみなすモデルである。このモデルは、多くの心理学的研究で用いられる、文の意味を表現するための計算手法である [6] [5] [11]。Predication モデルを極度に単純化したモデルであると考えることが可能である。先行研究ではしばしば文の主部（本研究の主題と対応）と述部

(本研究の喩辞と対応)の特徴ベクトルの単純な相加平均のみを扱っている。本研究では主題の特徴ベクトルと喩辞の特徴ベクトルにそれぞれ重み付けした加重平均を用いる。平均モデルによって計算される特徴ベクトルは以下の式で表す。

$$M_{Average} = \alpha w_T + (1 - \alpha) w_V$$

ここで、 w_T は主題の特徴ベクトルであり、 w_V は喩辞の特徴ベクトルである。また、 α は主題 (あるいは喩辞) の特徴ベクトルに対する重み付けのための定数である。これらの特徴ベクトルの作成方法については 3.1 節で述べた通りである。

3.3 Predication モデル

Kintsch [7] の Predication モデルは、主題と喩辞の特徴ベクトルに加えて、喩辞と最も類似する m 語の中で主題と最も類似する k 語の特徴ベクトルを平均したものを、比喩表現の意味とみなすモデルである。Predication モデルによって計算される特徴ベクトルは以下の式で表す。

$$M_{Predication} = Average(w_T + w_V + \sum_{w_i \in W_m(w_V)} sortsim(w_T, w_i))$$

ここで $w_i \in W_m(w_V)$ は、喩辞と最も類似する m 語の特徴ベクトルの集合である。さらに $sortsim(w_T, w_i)$ は、主題と $W_m(w_V)$ に属する単語のすべての類似度を計算し、値が i 番目に大きい単語の特徴ベクトルを返す関数である。

Kintsch [7] は、Predication モデルによって検索される比喩表現の解釈が適切であることを、事例をもとに説明している。たとえば、Predication モデルによって計算される「My lawyer is a shark」の特徴ベクトルと「vicious」(特徴)の特徴ベクトルのコサイン類似度が、「lawyer」(主題)の特徴ベクトルと「vicious」(特徴)の特徴ベクトルのコサイン類似度よりも高いことを報告している。

このモデルについてより定量的な評価を行ったのが Kintsch と Bowles [8] である。彼らは、30 個の比喩表現に対して参加者 ($N = 24$) に解釈を書かせた。そして、参加者から得られた解釈を、出現頻度に基づいた意味を表す特徴ベクトルとして扱い、この特徴ベクトルと Predication モデルによって表現される特徴ベクトルのコサイン類似度が十分に高い (0.50~0.51) ことから、Predication モデルの妥当性が高いことを主張した。

これらの研究は、比喩表現の解釈の検索における Predication モデルの妥当性の一端を示していると考えられる。しかしながら、検索性能の評価という観点から、Kintsch らの 2 つの研究には問題点がある。大きな問題点の一つとして、Kintsch と Bowles [8] では 30 個の比喩表現のみを対象としていたため、タスク数が充分ではないという点があげられる。また、検索モデルの評価として、参加者の産出した解釈をもとにして算出した意味を表す特徴ベクトルを用いるという手法も一般的ではない。さらに、Predication モデル以外のモデルとの比較も行われていない。

こうした問題点を踏まえて、本研究ではより多くの比喩表現をタスクとし、検索モデルの代表的な評価指標を用いることで、Predication モデルの比喩表現の解釈の検索性能を評価する。また、ベースラインのモデル (平均モデル) と検索性能を比較する。

表 1 名詞直喩表現と解答例の具体例 (括弧内は回答頻度)

| 名詞直喩表現 | 第一解釈 | 第二解釈 | 第三解釈 |
|-----------|----------|----------|----------|
| 困難は壁のようだ | 高い (6) | 大きい (4) | 難しい (4) |
| 笑顔は花のようだ | 明るい (19) | 美しい (19) | かわいい (5) |
| 人生は航海のようだ | 激しい (18) | 熱い (8) | 恐ろしい (3) |
| 激怒は噴火のようだ | 広い (6) | 大きい (4) | 温かい (3) |

4. 実験

本節では、提案手法を用いて行った実験と結果およびその評価を述べる。

4.1 タスクと正解

本研究では、中本と楠見 [14] で収集された名詞直喩表現をタスクとし、岡ら [12] で参加者から収集された名詞直喩表現に対する解釈に含まれる形容詞を、モデルを評価する際の正解とする。

以下では、岡ら [12] における、名詞比喩表現の解釈の収集方法について説明する。この課題には大学生・大学院生 50 名が参加した。参加者は画面中央に提示される直喩を読み、その解釈を回答欄に最低 1 つ、最大 3 つまでキーボードで入力することが求められた。この実験では、中本と楠見 [14] で使用された 120 個の直喩表現が用いられた。

この課題で参加者が各直喩に対して産出したすべての解釈を対象に形態素解析を行った。そして、形態素解析結果において出現した形容詞とその頻度を正解データとして用いた。形容詞を正解データとして用いるのは、名詞比喩表現 (例: 困難は壁のようだ) が表す特徴の多くが、形容詞的に表現されうる (例: 固い) と考えたためである。本研究でタスクとして用いた名詞直喩表現と、回答頻度の多い上位第三解釈までの形容詞の具体例を表 1 にまとめた。

4.2 コーパス

コーパスは、「現代日本語書き言葉均衡コーパス (version 1.1)」中に含まれる出版サブコーパス内の書籍サンプルを使用した。本研究では意味空間を形成するために、Inohara ら [6] に従い、単語-段落行列を用いた。このとき、コーパス中に含まれる動詞、形容詞、形容動詞、名詞を対象とした。コーパス中に含まれる対象とした総単語数は 82,068 語であり、総段落数は 266,036 段落であった。

4.3 潜在意味解析

コーパスの単語-段落行列に対して、TF-IDF 法による重み付けを行ったのちに、潜在意味解析を適用した。そして、単語 (82,068 語) -段落文脈 (D 次元) 行列に特異値行列 (D 次元) をかけた行列を、単語の意味空間として用いた。

本研究では、平均モデルと Predication モデルのそれぞれで、この単語の意味空間上で比喩表現の特徴ベクトルを計算する。そして、計算された比喩表現の特徴ベクトルと、コーパス中に含まれる全ての形容詞 (659 語) との類似度のランキングを求める。

4.4 モデルの検索性能の評価

本研究では、4.3 節で説明した比喩表現の特徴ベクトルと形容

表 2 各モデルの検索性能の評価値（カッコ内は標準偏差）

| モデル | パラメータ | nDCG@10 | MAP | MRR |
|-------------|---------------------------|---------------------|---------------------|---------------------|
| 平均 | $\alpha = 0.5, D = 300$ | 0.125(0.183) | 0.118(0.097) | 0.356(0.376) |
| 平均 | $\alpha = 0.6, D = 200$ | 0.125(0.176) | 0.125(0.106) | 0.366(0.369) |
| 平均 | $\alpha = 0.7, D = 200$ | 0.122(0.171) | 0.122(0.104) | 0.376(0.374) |
| Predication | $m = 10, k = 1, D = 200$ | 0.118(0.178) | 0.110(0.094) | 0.334(0.371) |
| Predication | $m = 200, k = 1, D = 200$ | 0.109(0.164) | 0.121(0.099) | 0.354(0.371) |

詞の類似度ランキングを、4.1 節で説明した参加者から収集された形容詞とその出現頻度によって評価する。すなわち、各モデル（平均モデル、Predication モデル）の特徴ベクトルに基づいて計算される形容詞の類似度のランキングが、参加者から収集された形容詞の出現頻度に基づくランキングとてらした時に、どのくらい妥当であるかを評価する。この目的に照らして、モデルの検索性能を評価するための評価指標としては、MRR と MAP と nDCG@10 を用いることにした。ここで、nDCG@10 における関連度は、4.1 節で説明した参加者から収集された形容詞とその出現頻度を用いた。

続いて、モデルの検索性能の評価に用いるハイパーパラメータの設定は以下の通りである。はじめに、特異値分解に用いる特異値の数 D は、50, 100, 200, 300 とする。続いて、平均モデルにおける α は、0 から 1.0 の範囲で、0.1 刻みで変化させる。最後に、Predication モデルにおける m は 10, 30, 50, 70, 90, 150, 200 とした。そして、 k は 1, 2, 3, 4, 7, 10 とした。

4.5 実験結果と考察

はじめに、モデルの検索性能の評価値（MRR, MAP, nDCG@10）を、平均モデルと Predication モデルで比較する。続いて、平均モデルと Predication モデルのそれぞれについて、nDCG@10 が最も高いモデルのパラメータの安定性を確認する。さらに、それぞれのモデルの検索結果と中本ら [14] の心理的な印象評定との関係について、探索的な相関分析をもとに検討する。最後に、それぞれのモデルの検索結果について、具体例を元に考察する。

4.5.1 モデルの検索性能の比較

モデルの検索性能を評価する際には、タスクで使用する比喻表現の主題と喩辞が、コーパス中に含まれている必要がある。しかしながら、120 個中 21 個のタスクについては、コーパス中に主題か喩辞の少なくともどちらかの語について含まれていなかった。そこで、以降の解析では、残された 99 個のタスクに関する結果を報告する。

それぞれのモデルの検索性能の評価値を表 2 にまとめた。表 2 に示すように、平均モデルにおける nDCG@10 の評価値が最も高いモデル ($\alpha = 0.5, D = 300$) の nDCG@10 は 0.125 であった。この結果は、Predication モデルにおける nDCG@10 の評価値が最も高いモデル ($m = 10, k = 1, D = 200$) の nDCG@10 である 0.118 を上回っていた。これらの結果は、名詞直喩表現の解釈の検索モデルとして Predication モデルが平均モデルよりも優れている [8] という結果とは異なっていた。

4.5.2 モデルのパラメータの安定性

以下の節では、平均モデルと Predication モデルのそれぞれについて定性的な評価を行う。人は名詞直喩表現から様々な解釈を生成できる [15] と言われている。この知見を踏まえると、名詞直喩表現の解釈検索モデルを評価する場合には、参加者が産出した複数の解釈を回答頻度に基づいて重み付けて評価できる nDCG@10 が良い指標であると考えられる。そこで、以下に述べるモデルの定性的な評価では、nDCG@10 の評価値が最も高かった平均モデル ($\alpha = 0.5, D = 300$) と Predication モデル ($m = 10, k = 1, D = 200$) で得られた、モデルの検索結果をもとに議論する。

本節では、平均モデルと Predication モデルのそれぞれで、nDCG@10 が最も高いモデルのパラメータの安定性を評価する。平均モデルにおいて、 α を 0.5 に固定して D を 50, 100, 200, 300 で動かした時の nDCG@10 の値を図 1 にまとめた。また、 D を 300 に固定して α を 0 から 1 で動かした時の nDCG@10 の値を図 2 にまとめた。図 1 から、 α が 0.5 の状況で D は値が大きいほど検索性能が高く、 D の値をより大きくすることで、モデルの検索性能がより高くなることが示唆された。一方で、図 2 から明らかなように、 D が 300 の状況で α は 0.5 に近づくほど（すなわち、主題と喩辞の重み付けが同じほど）検索性能が高くなることが示された。この結果は、平均モデルにおいて、 α が安定していることを示している。

続いて、Predication モデルにおいて、 k を 1, D を 200 で固定して、 m を動かした時の nDCG@10 の値を図 3 にまとめた。また、 k を 1, m を 10 で固定して、 D を動かした時の nDCG@10 の値を図 4 にまとめた。さらに、 m を 10, D を 200 で固定して、 k を動かした時の nDCG@10 の値を図 5 にまとめた。図 3 から、 m の値が小さいほど検索性能が高く、 m の値をより小さくすることで、モデルの検索性能がより高くなることが示唆された。また、図 4 から明らかなように、 D の値は 200 に近づくほど検索性能が高かった。最後に、図 5 から明らかなように、 k の値が小さいほど検索性能が高かった。これらの結果は、Predication モデルにおいて、 D と k が安定していることを示している。

4.5.3 モデルの検索結果と心理的な印象評定との相関

中本ら [14] は本研究で用いた直喩表現について 248 名の参加者に対して、(a) その直喩が言い表そうとしている意味をどの程度明確に理解できるか（理解可能性）、(b) 主題と喩辞がどの程度意味的に似ているか（構成語類似性）、(c) その直喩がどの程度見聞きしたことのないオリジナリティのある文か（独創性）、そして (d) その直喩の意味内容に面白さや意外性が感じら

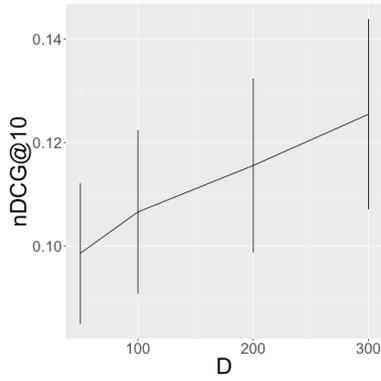


図1 平均モデル ($\alpha = 0.5$): エラーバーは標準誤差

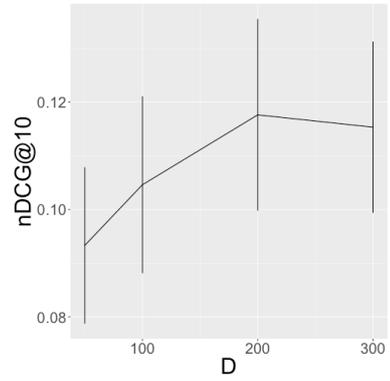


図4 Predication モデル ($k = 1, m = 10$): エラーバーは標準誤差

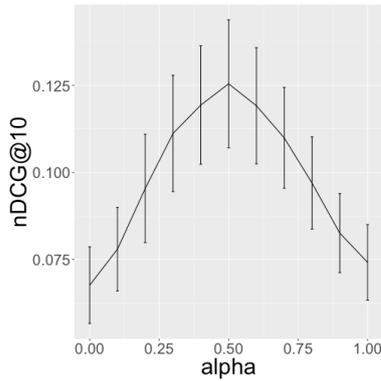


図2 平均モデル ($D = 300$): エラーバーは標準誤差

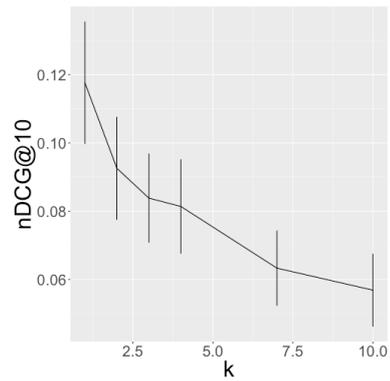


図5 Predication モデル ($m = 10, D = 200$): エラーバーは標準誤差

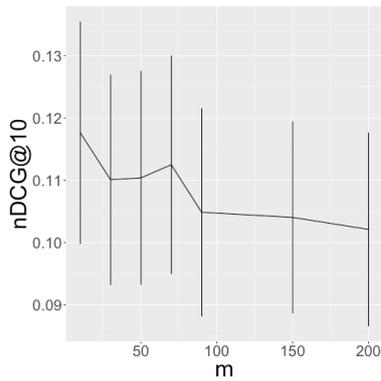


図3 Predication モデル ($k = 1, D = 200$): エラーバーは標準誤差

れるか(面白さ)をそれぞれ9件法で評価させた。そして、中本らは調査で得られた評定平均値について、論文中で報告していた。ここでは、中本らの結果と、本研究で得られた結果の関連について探索的に検討するために、2つのモデル(平均モデル, Predication モデル)のnDCG@10と、上述の4つの指標間を用いて相関分析を行った(表3)。

表3から明らかのように、nDCG@10の値と、指標間に有意な相関関係は見られなかった。本研究の結果は、複数解釈に基づくモデルの検索性能の評価値は、理解可能性や構成語類似性のような心理的な印象評定値とは関連がみられない可能性を示唆した。

4.5.4 モデルの検索結果の具体例

続いて、平均モデルとPredicationモデルの定性的な評価を

表3 それぞれのモデルのnDCG@10と中本・楠見(2004)の指標間の相関

| 指標 | 1 | 2 | 3 | 4 | 5 | 6 |
|-------------------------|---------|------|----------|----------|------|---|
| 1. nDCG@10(平均) | - | | | | | |
| 2. nDCG@10(Predication) | .84 *** | - | | | | |
| 3. 理解可能性 ^a | -.01 | .11 | - | | | |
| 4. 構成語類似性 ^a | .00 | .11 | .92 *** | - | | |
| 5. 独創性 ^a | .04 | -.01 | -.48 *** | -.56 *** | - | |
| 6. 面白さ ^a | .00 | .16 | .74 *** | .61 *** | -.05 | - |

注:^a中本・楠見(2004)に基づく。*** $p < .001$

行う。nDCG@10の成績でPredicationモデルが平均モデルよりも優れていたタスクの結果の具体例を表4から表7にまとめた。

Predicationモデルは、「笑顔は花のようだ」や「粉雪は羽毛のようだ」といった直喩表現から、それぞれについて人が多く産出する特徴である明るい(表4)や白い(表5)といった重要な特徴を、高い順位で出力している。

Predicationモデルが比喩の理解においてより効果的に機能している様子は、「心は沼のようだ(表6)」や「失望は樹液のようだ(表7)」などの名詞直喩表現の出力に見られる。こうした名詞直喩表現では、人が多く産出している特徴が平均モデルでは上位に現れない特徴(たとえば、「心は沼のようだ」という表現における深いという解釈や、「失望は樹液のようだ」という表現における「甘い」という解釈)を、Predicationモデルが適切に反映していることが明らかになる。

表4 「笑顔は花のようだ」のモデルの出力結果と正解

| 順位 | 平均 | Predication | 正解 | 頻度 |
|---------|-------|-------------|--------|----|
| 1 | 美しい | 美しい | 明るい | 19 |
| 2 | 白い | 明るい | 美しい | 17 |
| 3 | 可愛らしい | 青い | かわいい | 5 |
| 4 | 赤い | 濃い | かわいらしい | 4 |
| 5 | 黄色い | 淡い | ない | 3 |
| 6 | 愛らしい | 白い | 温かい | 2 |
| 7 | 青い | 温かい | 良い | 2 |
| 8 | 明るい | 赤い | いい | 2 |
| 9 | 温かい | 黄色い | 楽しい | 2 |
| 10 | 濃い | 薄い | よい | 1 |
| nDCG@10 | 0.621 | 0.780 | | |

表5 「粉雪は羽毛のようだ」のモデルの出力結果と正解

| 順位 | 平均 | Predication | 正解 | 頻度 |
|---------|-------|-------------|-------|----|
| 1 | 黄色い | 白い | 白い | 25 |
| 2 | 赤い | 黒い | 軽い | 18 |
| 3 | 黒い | 赤い | 柔らかい | 10 |
| 4 | 白い | 黄色い | やわらかい | 7 |
| 5 | 青白い | 淡い | 美しい | 4 |
| 6 | 細い | 青い | よい | 3 |
| 7 | 青い | 細い | 優しい | 3 |
| 8 | 薄暗い | 薄い | かるい | 2 |
| 9 | 太い | 青白い | ない | 2 |
| 10 | 細長い | 眩い | こちよい | 1 |
| nDCG@10 | 0.217 | 0.505 | | |

表6 「心は沼のようだ」のモデルの出力結果と正解

| 順位 | 平均 | Predication | 正解 | 頻度 |
|---------|-------|-------------|-----|----|
| 1 | 空しい | 遠い | 深い | 20 |
| 2 | 奥深い | 小高い | 暗い | 13 |
| 3 | 遠い | 程近い | ない | 7 |
| 4 | 清い | 奥深い | 汚い | 3 |
| 5 | 程近い | 美しい | 悪い | 2 |
| 6 | 近い | 青い | 難しい | 2 |
| 7 | 温かい | 近い | だるい | 1 |
| 8 | 頼り無い | 深い | 危ない | 1 |
| 9 | 浅い | 細長い | 多い | 1 |
| 10 | 清々しい | 暗い | わるい | 1 |
| nDCG@10 | 0.000 | 0.282 | | |

5. まとめと今後の課題

本研究では、名詞比喩表現における解釈の検索モデルの提案と評価を行った。本研究では、人の名詞比喩表現に対する解釈生成過程を考慮した2つの計算モデル（平均モデルと Predication モデル）の検索性能を評価した。実験では現代日本語書き言葉均衡コーパスを用いて、このコーパスの単語-段落行列に潜在意味解析を適用し、単語の意味空間を獲得した。そして、岡ら [12] の研究で参加者から得られた解釈に含まれる形容詞とその出現頻度を正解データとして使用した。最後に、計算モデルに与えられた比喩表現に対して、出力された形容詞のランキングを基にして、検索性能を得ることでモデルの評価を行った。

表7 「失望は樹液のようだ」のモデルの出力結果と正解

| 順位 | 平均 | Predication | 正解 | 頻度 |
|---------|-------|-------------|------|----|
| 1 | 恐ろしい | 苦い | 甘い | 9 |
| 2 | 容易い | 程良い | ない | 4 |
| 3 | 激しい | 重い | 悪い | 2 |
| 4 | 重い | 恐ろしい | にくい | 2 |
| 5 | 固い | 生白い | 賢い | 1 |
| 6 | 弱い | 冷たい | 重い | 1 |
| 7 | 重たい | 酸っぱい | 重々しい | 1 |
| 8 | 脆い | 固い | 苦い | 1 |
| 9 | 力強い | 激しい | 酸い | 1 |
| 10 | 素晴らしい | 甘い | しつこい | 1 |
| nDCG@10 | 0.028 | 0.267 | | |

nDCG@10に基づく検索性能の結果から、平均モデルは Predication モデルよりも優れていることが示唆された。Predication モデルの検索性能の優位性を主張した先行研究 [8] [7] は、nDCG を用いては評価していなかったため、本研究との直接的な結果の比較はできないが、本研究では平均モデルも Predication モデルと同程度かそれ以上の検索性能を持っていることを示唆した。また、nDCG@10 と中本ら [14] の心理的印象評定の間には有意な相関関係がみられなかった。特に、理解可能性と nDCG@10 に相関関係が見られなかったことは、理解可能性が高い比喩と低い比喩でモデルの出力と、参加者が生成した解釈に基づく特徴ベクトルの類似度が異なることを報告した Kintsch ら [8] らの結果と異った。ただし、「心は沼のようだ」や「失望は樹液のようだ」のように、Predication モデルによって適切な解釈が検索結果の上位にみられる結果も得られた。

今後の課題としては以下の四つが考えられる。第一に、モデルの評価に用いる解釈を整理することである。本研究では、岡ら [12] の研究で参加者から得られた解釈に含まれる形容詞とその出現頻度を正解データとして用いた。この研究では参加者に回答させる解釈は自由であったため、直喩表現によってはほとんど形容詞が含まれていないものもあった。例えば、「呼吸はポンプのようだ」という直喩表現では、頻度が1よりも大きい形容詞は「規則正しい」しかなかった。今後はこうした問題を解消するために、形容詞のみを産出させた正解データが必要になる。

第二に、モデルの評価に用いる直喩表現を整理することも求められる。本研究では、中本ら [14] で報告された直喩表現を用いていたが、この中には様々な比喩のバリエーションが存在していた。たとえば、「笑顔は花のようだ」という直喩表現は、抽象的な意味での特徴の類似（例：笑顔は華やかな雰囲気になり、花は華やかである）が解釈となる。一方で、「粉雪は羽毛のようだ」という直喩表現は、双方の具体的な特徴の類似（例：どちらも柔らかい）が解釈となる。さらに、「激怒は噴火のようだ」という直喩表現のように、抽象的な意味での特徴の類似が動作（例：燃え上がる）として解釈される場合もある。このように、直喩表現の中でも、その解釈を検索する際の基準（抽象的か具体的か、形容詞か動詞か）が異なる場合がある。今後はこうした直喩表現のバリエーションについても事前に特定した上で、

適切な直喩表現と解釈を用意する必要がある。

第三に、単語の意味空間の生成について様々なバリエーションを検討する必要がある。本研究では、Kintschら[8][7]になら、潜在意味解析に基づいて単語の意味空間を形成した。しかしながら、近年ではトピックモデルや word2vec などの、より人の直感にあった単語の意味空間の形成方法も提案されている。今後はこうしたモデルを用いて、名詞直喩表現の解釈の検索モデルを検討していく必要がある。

第四に、単語の Predication モデルを改良する必要がある。本研究では、Kintsch[7]になら、主題と喩辞、そして喩辞と最も類似する単語の特徴ベクトルの中で、主題と最も類似する特徴ベクトル k 個の語の特徴ベクトルを平均したものを、比喩表現の意味とみなした。しかしながら、上記のモデルは下記のモデルのように、一般化して書き直すことが可能である。

$$M_{Predication} = \alpha w_T + \beta w_V \\ + (1 - \alpha - \beta) \sum_{w_i \in W_m(w_V)}^k \text{sortsim}(w_T, w_i)$$

上記のモデルは平均モデルをネストしたモデルであることに加えて、 $(1 - \alpha - \beta) \sum_{w_i \in W_m(w_V)}^k \text{sortsim}(w_T, w_i)$ の部分が Predication モデルの独自の成分であることが明示されているという点から、今回実装した Predication モデルよりも優れたモデルであることがわかる。今後はこの改良版の Predication モデルを用いて結果を評価していく必要がある。

謝 辞

本研究の一部は、京都大学大学院デザイン学連携プログラムによる研究助成を受けました。また、JSPS 科学研究費助成事業 (JP15H01718, JP16H02906) による研究助成も受けました。ここに記して謝意を表します。

文 献

- [1] SC Deerwester, ST Dumais, GW Furnas, TK Landauer, and R Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.
- [2] Morton. Ann Gernsbacher, Boaz Keysar, Rachel. R.W. Robertson, and Necia. K. Werner. The role of suppression and enhancement in understanding metaphors. *Journal of Memory and Language*, 45(3):433–450, 2001.
- [3] Sam Glucksberg and Boaz Keysar. Understanding metaphorical comparisons : Beyond similarity. *Psychological Review*, 97(1):3–18, 1990.
- [4] Sam Glucksberg, Mary R Newsome, and Yevgeniya Goldvarg. Inhibition of the literal: Filtering metaphor-irrelevant information during metaphor comprehension. *Metaphor & Symbol*, 16(3/4):277–298, 2001.
- [5] Arthur C. Graesser and Danielle S. McNamara. Computational Analyses of Multilevel Discourse Comprehension. *Topics in Cognitive Science*, 3(2):371–398, apr 2011.
- [6] Keisuke Inohara, Ryoko Honma, Takayuki Goto, Takashi Kusumi, and Akira Utsumi. The relationship between reading literary novels and predictive inference generation: A corpus-based approach employing a corpus of japanese novels. *Scientific Study of Literature*, 4(1):46–67, 2014.
- [7] W Kintsch. Metaphor comprehension: a computational theory. *Psychonomic bulletin & review*, 7(2):257–66, 2000.
- [8] Walter Kintsch and Anita R. Bowles. Metaphor compre-

hension: What makes a metaphor difficult to understand? *Metaphor & Symbol*, 17(4):249–262, 2002.

- [9] Ekaterina Shutova, Simone Teufel, and Anna Korhonen. Statistical Metaphor Processing. *Computational Linguistics*, 39(2):301–353, 2013.
- [10] Akira Utsumi and M. Sakamoto. Computational evidence for two-stage categorization as a process of adjective metaphor comprehension. In *Proceedings of the Second European Cognitive Science Conference (EuroCogSci2007)*, pages 77–82, 2007.
- [11] Michael B W Wolfe. Memory for narrative and expository text: Independent influences of semantic associations and text organization. *Journal of Experimental Psychology: Learning Memory and Cognition*, 31(2):359–364, 2005.
- [12] 楠見 孝 岡 隆之介, 大島 裕明. 隠喩-直喩選好性に与える喩辞慣習性, 解釈多様性, 構成語類似性の影響. 心理学研究, 投稿中.
- [13] 内海 彰. 比喩理解への計算論的アプローチ-言語認知研究における計算モデルの役割-. 認知科学, 20(2):249–266, 2013.
- [14] 楠見 孝 中本 敬子. 比喩材料文の心理的特性と分類-基準表作成の試み-. 読書科学, 48(1):1–10, 2004.
- [15] 楠見 孝 平知宏, 中本 敬子. 比喩理解における親しみやすさと解釈の多様性. 認知科学, 14(3):322–338, 2007.