

2 段階からなる比喩理解の計算モデルの構築

寺井あすか¹、中川正宣²

(¹東京工業大学グローバルエッジ研究院、²東京工業大学大学院社会理工学研究科)

本研究では、比喩理解過程はカテゴリ化過程とダイナミックインタラクション過程の 2 段階からなると仮定し、言語統計解析に基づく比喩理解の計算モデルを構築した。はじめに、言語統計解析結果に基づき概念を表す意味ベクトルを推定する。次に、カテゴリ化過程モデルとして、喩辞を典型例とするアドホックなカテゴリのメンバーとして被喩辞の意味を推定する。さらに、ダイナミックインタラクション過程において、カテゴリ化過程モデルによって推定された、喩辞を典型例とするアドホックなカテゴリのメンバーである被喩辞の意味ベクトルに対し、特徴間の相互作用が影響を及ぼし比喩の意味が推定される。最後に、構築した比喩理解モデルの妥当性の検証を行うため心理実験を行った。その結果、カテゴリ化過程のみのモデルと比較し 2 段階過程モデルがより妥当であることが示された。

1. はじめに

本研究の目的は、言語コーパスを用いて、「A のような B」という形式で表現される比喩（直喩）の理解過程を表現する計算モデルを構築することである。直喩の理解においては、喩辞（A）と被喩辞（B）がどのような特徴を持つのかという、知識構造の違いが比喩理解に影響を及ぼすことが楠見（1995）によって明らかにされている。しかし、喩辞（A）と被喩辞（B）に関する知識構造について、人間が扱う比喩を網羅するのに十分な量を心理実験のみによって抽出することは、現実的には非常に困難である。そこで、本研究では言語統計解析により得られた意味空間に基づき、喩辞（A）と被喩辞（B）を含む各概念を、意味空間内のベクトルを用いて表現することで、比喩理解の計算モデルを構築する。

言語コーパスを用いた比喩理解モデルがすでにいくつか提案されている（Terai & Nakagawa 2006, 2007a, Utsumi 2006, Kintsch 2000）。Kintsch, Utsumi らの比喩理解モデルは、クラス包含理論（Glucksberg & Kayser 1990）に基づき、構築されている。クラス包含理論は、喩辞を典型例とするアドホックなカテゴリに被喩辞が含まれるとみなすことで、比喩が理解されると説明するものである。Kintsch, Utsumi らのモデルでは各概念の意味を Latent Semantic Analysis (LSA) (Deerwester et al. 1990)を用いて推定しているため、比喩を表現するベクトルが持つ意味は、他のベクトルとの余弦に基づき推定される。そのため、比喩ベクトルが持つ意味の解釈が困難であるという問題点がある。また、Terai, Nakagawa のモデルでは、言語統計解析 (Kameya, Sato 2005)を用いて、各概念を P(形容詞 | 概念) という条件付確率で表現することで、

LSAを用いた推定において発生する問題点を解消している。このモデルは顕現性落差理論 (Ortony 1979)に基づき、構築されている。顕現性落差理論では、喩辞において顕現性の高い特徴と被喩辞において顕現性が中程度の特徴との類似性を発見することで比喩理解が成立されると説明されている。しかし、喩辞と被喩辞の類似性に着目した理論では喩辞と被喩辞の役割の違いがうまく説明できないという問題点がある。そのため、顕現性落差理論を用いて構築されたモデルでは、喩辞と被喩辞の比喩理解における役割の違いがうまく説明されていないという問題がある。

一方で、比喩理解過程において、喩辞、被喩辞との関連が弱いにも関わらず、比喩理解において顕現する特徴（創発特徴）も比喩理解に大きな影響を及ぼすことが明らかになっている (Nueckles, Janetzko 1997, Gineste et al. 2000)。内海 (2000) では、喩辞と関連の強い特徴と情緒（「明暗」「強弱」等）的に類似し、かつ被喩辞を説明しうる特徴が創発特徴となると説明している。すなわち、比喩理解過程において、特徴間での影響関係により創発特徴が顕現すると考えられる。Terai & Nakagawa (2006, 2007)のモデルでは、これらの創発特徴を比喩理解過程における特徴間のダイナミックなインタラクションによって顕現すると説明している。

本研究では、先行研究の問題点を克服し、さらに創発特徴を表現することが可能なモデルの構築を目指し、比喩理解過程はクラス包含理論に基づくカテゴリ化過程とダイナミックインタラクション過程の 2 段階からなると仮定し、言語統計解析 (Kameya, Sato 2005)を用いて比喩理解の計算モデルを構築する。

2. 言語統計解析を用いた概念の表現

毎日新聞 10 年分 (1993 年～2002 年) から、CaboCha (工藤, 松本 2002) を用いて形容詞と名詞の係り受け頻度データを抽出した。抽出された形容詞と名詞の係り受け頻度データに対し、Pereira のモデル (Pereira et al. 1993) や PLSI (Hofmann 1999) と同型のモデル (Kameya, Sato 2005) を用いて言語統計解析を行い、潜在クラス c が与えられたときの名詞 n_i または形容詞 a_j の条件付確率 ($P(n_i|c_k)$ または $P(a_j|c_k)$) と、潜在クラスの出現確率 ($P(C_k)$) を推定する。

本研究では、言語統計解析結果に基づき名詞 (概念) が与えられたときの形容詞 (特徴) の条件付確率を求め、それらの値を用いて概念と特徴の関連強度の推定を行う。

$$P(a_j|n_i) = \frac{\sum_k P(a_j|c_k)P(n_i|c_k)P(c_k)}{\sum_k P(n_i|c_k)P(c_k)} \quad \dots(1)$$

この条件付確率は、潜在クラスを介して推定される形容詞と名詞の関連強度であることから、データスパースネス問題がより深刻である場合でも関連強度の推定を行うことが可能である。

さらに、新聞データには、“冷たい氷” などといった定義的な特徴が表現されないため、国語辞典データとして、学研国語大辞典 (金田一、池田、1988) を用いて、条件付確率の修正を行う。見出し語 (名詞 n_i) に対し、説明文中に 2

度以上出現している形容詞 (a_j) (そのような形容詞の集合を $\text{Dic}(n_i)$ と示す) は、その見出し語 (名詞) の重要な特徴を表わしていると考えられる。そのような名詞と形容詞の関連強度は、新聞データによって計算される名詞と形容詞との関連強度より大きな値をとると仮定し、条件付確率の修正を行う。したがって、各概念 (名詞) を表すベクトル (概念ベクトル) の j 番目の要素 $V_j(n_i)$ は式 (2) によって計算される。

$$V_j(n_i) = \begin{cases} P(a_j|n_i) + \max_{i,j} P(a_j|n_i) & \text{if } a_j \in \text{Dic}(n_i) \\ P(a_j|n_i) & \text{else} \end{cases} \quad \dots(2)$$

“山のような仕事”、“鬼のような先生” という比喩で用いられている “山”、“仕事”、“鬼”、“先生” という概念に関して推定された意味ベクトルの結果として、各意味ベクトルから各概念と強く関連していると推定された特徴を表 1 に示す。

3. 比喩理解モデル

本研究では、比喩理解過程はカテゴリ化過程とダイナミックインタラクション過程の 2 段階からなると仮定し、モデルの構築を行う。

3. 1. カテゴリ化過程モデル

言語統計解析結果を用いて計算した概念ベクトルを用い、喩辞を典型例とするアドホックなカテゴリのメンバーとしてみなされた被喩辞

表 1: 言語統計解析結果に基づき推定された意味ベクトル

	山のような仕事				鬼のような先生			
	山		仕事		鬼		先生	
	特徴	$V_j(\text{山})$	特徴	$V_j(\text{仕事})$	特徴	$V_j(\text{鬼})$	特徴	$V_j(\text{先生})$
1	高い	1.0714	重要な	0.0659	恐ろしい	0.8814	若い	0.0968
2	深い	0.0403	新しい	0.0628	不思議な	0.0465	良い	0.0318
3	低い	0.0400	主な	0.0373	白い	0.0400	好きな	0.0292
4	広い	0.0337	良い	0.0282	高い	0.0255	よい	0.0270
5	美しい	0.0330	よい	0.0222	巨大な	0.0252	厳しい	0.0269
6	近い	0.0305	悪い	0.0214	黒い	0.0212	弱い	0.0204
7	明るい	0.0212	厳しい	0.0202	赤い	0.0209	悪い	0.0202
8	狭い	0.0209	好きな	0.0194	奇妙な	0.0194	不自由な	0.0141
9	いい	0.0177	可能な	0.0190	美しい	0.0186	楽しい	0.0132
10	白い	0.0137	さまざまな	0.0186	青い	0.0157	有名な	0.0128

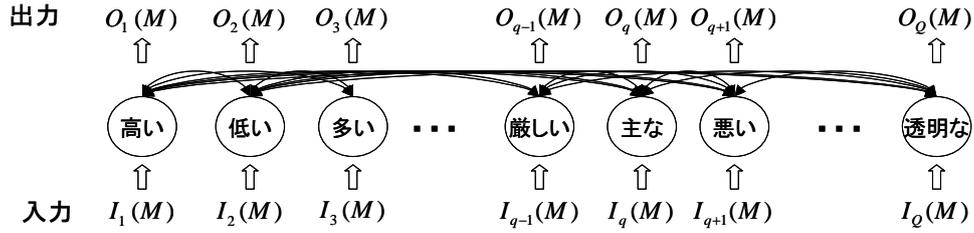


図1：ダイナミックインタラクション過程モデルのアーキテクチャ

の意味を計算する。

はじめに、概念間の類似度は、式(3)によって計算される。

$$sim(n_i, n_j) = \frac{V(n_i) \cdot V(n_j)}{\|V(n_i)\| \|V(n_j)\|} \quad \dots(3)$$

この類似度に基づき、喩辞ベクトルと類似している s 個の概念ベクトルからなる、喩辞ベクトルの近傍 ($N^s(\text{vehicle})$) を計算する。

次に、喩辞の近傍 $N^s(\text{vehicle})$ から、被喩辞に類似した L 個の概念 (n_i) を抽出する。これらの L 個の概念と喩辞からなるカテゴリを、喩辞を典型例とするアドホックなカテゴリとみなし、被喩辞、喩辞、 L 個の抽出された概念のセントロイド (重心) を求めることで、喩辞を典型例とするアドホックなカテゴリのメンバーとして被喩辞の意味ベクトル ($V(M)$) を計算する (式(4)参照)。

$$V(M) = \frac{\sum_i V(n_i) + V(\text{target}) + V(\text{vehicle})}{L+2} \quad \dots(4)$$

このアルゴリズムは、Kintsch(2000), Utsumi(2006)のカテゴリ化理論に基づく比喩理解モデルと同様のものである。

3. 2. ダイナミックインタラクション過程モデル

はじめに、前述のカテゴリ化過程モデルで計算した $V(M)$ の j 番目の要素である $V_j(M)$ の値が閾値 θ を越えた場合、形容詞 a_j を比喩理解に関わる特徴として抽出する。次に、抽出された特徴を用いてリカレントニューラルネットワークモデルを構築する(図1)。各ノードは各特徴を表現する。これらのノードは入力ノードでありかつ出力ノードである。 $x_q(t)$ により q 番目ノードの時間 t の活性値を表すとき、各ノードの振舞いは、式(5)に示した連立微分方程式によって決定する。

$$\frac{dx_q(t)}{dt} = -x_q(t) + f\left(\beta \sum_r w_{qr} x_r(t) + I_q(M)\right) \quad \dots(5)$$

f は値域が $[-1, 1]$ のシグモイド関数を表す。各ノードは $dx_q/dt \approx 0$ が成り立つとき、出力値

$O_q(M) = x_q(t)$ を出力する。 $O(M)$ が比喩の意味を表現する。 $I_q(M)$ は q 番目のノードへの入力値を表し、 $I_q(M)$ は式(6)により計算される。

$$I_q(M) = \frac{V_{j(q)}(M) - \overline{V'(M)}}{SD'(M)}$$

$$\overline{V'(M)} = \frac{\sum_q V_{j(q)}(M)}{Q}$$

$$SD'(M) = \sqrt{\frac{\sum_q (V_{j(q)}(M) - \overline{V'(M)})^2}{Q-1}} \quad \dots(6)$$

$j(q)$ は q 番目のノードが表す特徴と特徴 (形容詞) a_j が一致する j を示す。また、 Q は抽出された特徴の数を示す。 β は比喩理解過程における特徴間のインタラクションが及ぼす影響を表し、 w_{qr} は r 番目のノードから q 番目のノードへの結合加重値を示し、 r 番目のノードと q 番目のノードが示す特徴と被喩辞、喩辞の兄弟概念の関連強度に関する相関係数によって推定する。本研究では、被喩辞、喩辞の兄弟概念として、被喩辞、喩辞の近傍 $N^{s_I}(\text{target})$, $N^{s_V}(\text{vehicle})$ に含まれる概念、各 s_I 個を用いた。これらのノード間の結合加重値 (w_{qr}) により特徴間のインタラクションを表現する。また、反対の意味を持つ属性値間の関連を高く推定するという言語統計解析の問題点を、反対語対照語辞典 (北原、東郷 2004) を用いて修正を行い、反対語対照語辞典において、反対語であると定義された特徴間の結合加重値は -1 とすることで、より自然なパラメータ推定を行った。

3. 3 モデルシミュレーション

本研究では、 $s=250$, $L=5$, $s_I=100$, $\theta=0.0029$ ($=10/\text{形容詞数}$)、 $\beta=0.5$ を用いて、“山のよう

な仕事”、“鬼のような先生”という比喻に関しモデルシミュレーションを行った。各シミュレーション結果として、カテゴリ化過程モデルの結果($V(M)$)と2段階モデルの結果($O_q(M)$)と強く関連している特徴を表2に示す。2段階モデルの結果は、値を[0,1]の範囲に変換した結果を示す。

どちらの結果がより妥当であるかを検証するため、次に述べる心理実験を行った。

4. 心理実験によるモデルの妥当性の検証

モデルの妥当性を検証するため、心理実験を行った。被験者：31名、課題：“山のような仕事”、“鬼のような先生”。被験者に被喩辞、喩辞、比喻文を提示し、各特徴を形容詞または形容動詞を用いて回答してもらう。2名以上の被験者

表2：モデルシミュレーション結果（“山のような仕事”、“鬼のような先生”）
（太字斜体下線で示した特徴は、心理実験において2名以上の被験者によって比喻の特徴として回答されたもの）

	山のような仕事				鬼のような先生			
	カテゴリ化モデル		2段階過程モデル		カテゴリ化モデル		2段階過程モデル	
	特徴	出力値	特徴	出力値	特徴	出力値	特徴	出力値
1	高い	0.3502	高い	0.8932	恐ろしい	0.1310	恐ろしい	0.9174
2	低い	0.0459	多い	<u>0.5702</u>	不思議な	0.0382	若い	0.8844
3	多い	<u>0.0250</u>	厳しい	<u>0.5252</u>	若い	0.0239	怖い	<u>0.8608</u>
4	近い	0.0248	近い	0.5143	奇妙な	0.0158	自然な	0.8598
5	厳しい	<u>0.0199</u>	素晴らしい	0.5005	良い	0.0154	良い	0.8140
6	新しい	0.0182	新しい	0.4967	高い	0.0143	強力な	0.7562
7	良い	0.0174	大事な	0.4906	怖い	<u>0.0099</u>	特別な	0.7222
8	重要な	0.0154	可能な	0.4903	好きな	0.0095	強い	<u>0.7205</u>
9	悪い	0.0108	さまざまな	0.4897	自然な	0.0095	美しい	0.7169
10	少ない	0.0102	大変な	<u>0.4866</u>	悪い	0.0092	黒い	0.6990

表3：心理実験結果（“山のような仕事”、“鬼のような先生”）（数字：回答者数）
（太字斜体下線で示した特徴は、2段階過程モデルによってのみ推定された特徴）

山のような仕事			鬼のような先生		
山	仕事	山のような仕事	鬼	先生	鬼のような先生
17：大きい	9：楽しい	20：多い	19：怖い	10：優しい	20：怖い
13：高い	8：つらい	8：忙しい	12：強い	8：怖い	16：厳しい
4：美しい	7：忙しい	8：辛い	9：大きい	5：厳しい	<u>7：強い</u>
3：寒い	5：めんどくさい	<u>7：大変な</u>	6：恐ろしい	3：うるさい	5：冷たい
2：広い	3：難しい	3：疲れる	4：赤い	2：偉い	3：怒っぽい
	2：固い	3：難しい	2：たくましい	2：賢い	2：暴力的な
	2：辛い	3：やりがいのある	2：青い	2：きれい	2：大きい
	2：大切な	2：終わらない		2：すごい	
	2：きつい	2：めんどくさい			
		2：きびしい			

が被喩辞、喩辞または比喩文の特徴として回答した特徴を表3に示す。

モデルシミュレーション結果(表2)と心理実験結果(表3)を比較する。“山のような仕事”の特徴として2名以上の被験者が回答している、“多い”、“大変な”、“厳しい”という特徴を、2段階過程モデルは比喩と強く関連している上位10個に含まれる特徴として推定しているのに対し、カテゴリ化過程モデルは“大変な”という特徴を比喩と強く関連している上位10個に含まれる特徴として推定できていない。また、実験結果から“多い”、“大変な”、“やりがいのある”、“終わらない”、“めんどくさい”、“厳しい”という特徴が創発特徴であると考えられる。2段階過程モデルはカテゴリ化モデルと比較し、より上位の特徴として“多い”、“大変な”という特徴を推定している。

同様に、“鬼のような先生”の特徴として2名以上の被験者が回答している、“怖い”、“強い”という特徴を2段階過程モデルは比喩と強く関連している上位10個に含まれる特徴として推定しているのに対し、カテゴリ化過程モデルは“強い”という特徴を比喩と強く関連している上位10個に含まれる特徴として推定できていない。

したがって、実験結果はカテゴリ化モデルと比較して2段階過程モデルがより妥当であることを示している。

5. 結論

本研究では、比喩理解過程はカテゴリ化過程とダイナミックインタラクション過程の2段階からなると仮定し、言語統計解析(Kameya, Sato 2005)を用いてモデル化を行った。

始めに、言語統計解析結果に基づき概念の意味ベクトルを計算した。次に、これらの意味ベクトルを用いて、喩辞を典型例とするアドホックなカテゴリのメンバーとみなされた被喩辞の意味を推定した(カテゴリ化過程)。さらに、アドホックなカテゴリのメンバーとみなされた被喩辞の意味に対し、特徴間のインタラクションによって影響を及ぼされた比喩の意味を推定した(ダイナミックインタラクション過程)。最後に心理実験を行うことで、カテゴリ化過程モデルと比較し、2段階過程モデルがより妥当であることを検証した。

今回の心理実験では、被験者に被喩辞、喩辞、比喩文を提示し、各特徴を形容詞または形容動

詞を用いて自由に回答してもらった。そのため、それらの特徴として適当であると考えられるにも関わらず、回答されていない特徴が存在する可能性がある。そこで、より詳細にモデルの妥当性の検証を行うため、モデルにより予想された特徴を提示し、被喩辞、喩辞、比喩文との関連強度を測定する必要があると考えられる。

また、本研究では形容詞全体の空間の中のベクトルとして意味を表現している。形容詞には、「良い」「いい」のようにほぼ同じ意味を表わすもの、「新しい」「新奇的」など意味的に近いものが存在することから、意味ベクトルの距離を問題にする場合には、形容詞自体の構成する意味空間構造を考慮する必要があると考えられる。そこで、今後は言語統計解析結果により推定された潜在クラスの空間の中のベクトルとして意味を表現することを検討したいと考えている。

謝辞

本研究は東京工業大学21世紀COE「大規模知識資源の体系化と活用基盤構築」と文部科学省における平成18年度科学技術振興調整費による委託事業「若手研究者の自立的な研究環境整備促進事業」の援助を受けて行われた。

参考文献

- Deerwester, S., Dumains, S., Furnas, G., Landauer, T., Harshman, R. (1990) Indexing by latent semantic analysis. *Journal of the Society for Information Science*. 41(6). 391-407
- Gineste, M., Indurkha, B., Scart, V. (2000) Emergence of features in metaphor comprehension. *Metaphor and Symbol*. 15(3), 117-135
- Glucksberg, S., Keysar, B. (1990) Understanding Metaphorical Comparisons: Beyond Similarity. *Psychological Review*. 97(1), 3-18
- Hofmann, T. (1999) Probabilistic latent semantic indexing, *Proc. of the 22nd International Conference on Research and Development in Information Retrieval (SIGIR '99)*. pp50-57
- Kameya, Y., Sato, T. (2005) Computation of probabilistic relationship between concepts and their attributes using a statistical analysis of Japanese corpora.

- Proc. of Symposium on Large-scale Knowledge Resources: LKR2005*. 65-68
- Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, **7**(2), 257-266
- Nueckles, M., Janetzko, D. (1997) The role of semantic similarity in the comprehension of metaphor. *Proc. of the 19th Annual Meeting of the Cognitive Science Society*. 578-583
- Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, **86**, 161-280
- Pereira F., Tishby N., Lee L. (1993) Distributional clustering of English words. *Proc. of 31th Meeting of the Association for Computational Linguistics*. pp.183-190
- Terai, A., Nakagawa, M. (2006). A Neural Network Model of Metaphor Understanding with Dynamic Interaction Based on a Statistical Language Analysis. *S. Kollias et al. (Eds.): ICANN 2006, Part 1, LNCS 4131*. Springer-Verlag Berlin Heidelberg. pp495-504
- Terai, A., Nakagawa, M. (2007a). A Neural Network Model of Metaphor Understanding with Dynamic Interaction Based on a Statistical Language Analysis: Targeting a Human-like Model. *International Journal of Neural Systems*. 17(4). 265-274
- Terai, A., Nakagawa, M. (2007b). A Computational Model of Metaphor Understanding Consisting of Two Processes. *J. Marques de Sa et al. (Eds.): ICANN 2007, Part 2, LNCS 4669*. Springer-Verlag Berlin Heidelberg. pp963-972
- Utsumi, A. (2006). Computational Exploration of Metaphor Comprehension Processes. *Proc. Of the 28th Annual Meeting of the Cognitive Science Society*. 2281-2286
- 内海彰 (2000). 比喩の認知/計算モデル. *Computer Today*. 96(3). 34-39
- 北原保雄、東郷吉男編 (2004) 反対語対照語辞典第9版. 東京堂出版
- 金田一春彦、池田弥三郎 (1988) 学研国語大辞典第2版. 学研
- 楠見 孝 (1995) 比喩の処理過程と意味構造 風間書房
- 工藤 拓、松本 裕治 (2002) チャンキングの段階適用による日本語係り受け解析、情報処理学会論文誌. 43(6). 1834-1842