

# 新奇的な比喩表現の生成手法 (分散表現・概念辞書・係り受け解析の統合的アプローチ)

小柳津 久嗣                      橋本 翔                      柳澤 秀吉  
 東京大学 工学部                  東京大学 大学院工学系研究科

{oyaizu6342,hashimoto229,hide}@mail.design.t.u-tokyo.ac.jp

## 1 背景と目的

製品開発において、人間の感性に評価を依存する品質(以下、感性品質)が要件となる場合が少なくない [1]. 感性品質の設計においては、アイデアの発想、評価および分析を繰り返すデザインプロセスが辿られる [2]. この過程では、日本語・英語間の翻訳や専門用語の言い換え等のように様々な言葉が必要とされる. 感性品質の言語化に応用可能な技術としては、Multilingual Semantic Networks(MLSN) が提案されており、翻訳妥当性の評価や曖昧な語の意味解析を可能としている [3] [4].

また、このデザインプロセスで用いられる語句には、相手にこちらの意味していることを効果的に伝える力である訴求力が重要である. 訴求力の高い表現としては、比喩が挙げられる [5]. しかし、このための語句を人の手で全て考えると、時間や労力がかかり、作業者によって言葉の質が変わりうる.

この問題に対する一つのアプローチとして、計算機による語句推薦が挙げられる. 語句推薦のための技術としては WordNet [6] のような概念辞書や、word2vec [7] のような分散表現、係り受け解析などが挙げられ、これらを用いて比喩生成について研究がなされてきた. 佐藤らは word2vec とシソーラスを用いて比喩生成におけるコーパスによる出力の差異を分析した [8]. また、中條らは word2vec と単語の具象度の高さを考慮して抽象度の低い比喩生成を試みている [9].

しかし、未だ比喩辞の提案について、人の感情を刺激する [10] ような新奇性の高い比喩生成を目的とした研究は見当たらない. そこで、本研究では新奇的な比喩表現を生成することを目的とした生成手法を提案する.

な特徴を表す語を特徴語と呼ぶ.

認知心理学において、比喩では情緒・感覚的類似性、カテゴリー的類似性、特徴語との共起関係の強さが重要とされている [11]. これを踏まえて、有効な比喩の条件として、以下を考えた.

- {1} 被喩辞と喩辞が同じ意味で特徴語を用いている
- {2} 特徴語と喩辞が共起しやすい
- {3} 喩辞と被喩辞の性質的なカテゴリが遠い

{1} に関しては、特徴語が多義語の場合、文脈における意味を特定して喩辞を選ぶ必要がある. {2} に関しては、喩辞が特徴語に係る必要がある. {3} に関しては、「霜が雪のように白い」のように似た性質のカテゴリの単語で例えてもインパクトがないため、被喩辞と異なるカテゴリである喩辞を選ぶ必要がある.

本研究では、{1}~{3} を満たすことを目的として、図 1 に示した比喩生成手法を提案する.

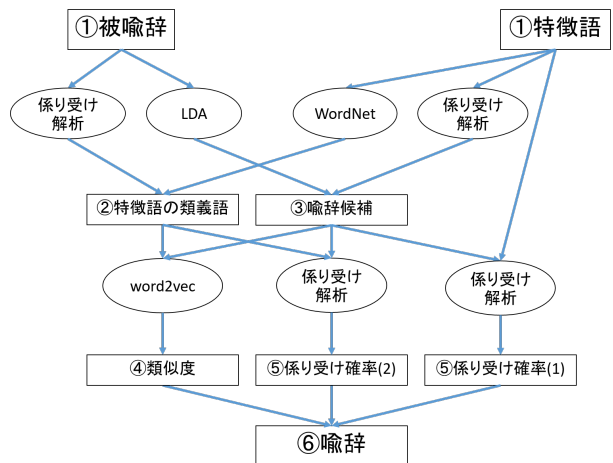


図 1: 比喩生成の流れ

## 2 新奇的な比喩の生成手法

比喩には、直喩、隠喩、換喩、提喩があるが、本研究では例えば「肌が雪のように白い」のような直喩を扱う. 直喩において「肌」のような例えられる語を被喩辞、「雪」のような例える語を喩辞、「白い」のよう

- ① 被喩辞と特徴語の入力  
 被喩辞として名詞、特徴語として形容詞または動詞を入力する.
- ② 文脈における特徴語の類義語の取得  
 前述した通り特徴語の文脈における意味を特定

するために、特徴語の辞書的な類義語の中でも用法が近いものを文脈における類義語として取得する。特徴語と WordNet において Synset を共有している語を辞書的な類義語とし、その類義語の中から被喩辞と係り受け解析で係り受けしている単語をこの文脈における類義語として取得する。

### ③ 喩辞候補の取得

新奇性のある比喩生成のためには、前述した通り喩辞と特徴語の共起関係が強く、喩辞と被喩辞がカテゴリ的に遠いことが重要であるため、喩辞は特徴語と係ること、被喩辞と異なるカテゴリであることが望ましい。そのため、以下の条件を全て満たし、特徴語と係り受け解析で係り受けする単語を喩辞候補として取得する。これより、有効な比喩条件 {2}、{3} を満たすことが出来る。

- LDA(潜在的ディリクレ配分法) で被喩辞と異なるカテゴリに属している
- WordNet に「名詞」として存在する
- 形態素解析器 MeCab で「名詞一般」に分類される

### ④ ③で取得した語と②で取得した語群の類似度の取得

先行研究 [8] [9] では被喩辞と特徴語の合成ベクトルを喩辞候補ベクトルとしているが、このように両方の単語を周辺単語に含む候補ベクトルを取ると典型的な呼応関係にある組み合わせの影響が大きくなり、既存のありきたりな比喩が生成されやすくなった。そこで本研究では、②で取得した類義語の word2vec における重心を喩辞候補ベクトルとすることで、文脈における特徴語の類義語の周辺に出現する単語を得る。そのため、文脈は考慮するが、典型的な呼応関係の影響を減らした、新奇な比喩表現の生成が出来ると考えた。それに加えて、文脈における類義語を考慮することで有効な比喩条件 {1} を満たすことが出来る。

喩辞候補の選択には喩辞候補ベクトルとのコサイン類似度を用いた。

### ⑤ 特徴語からの係り受け確率の取得

係り受け回数を係り受け確率に正規化した値を算出する。

特徴語を C、文脈における特徴語の類義語を [S] とすると、ある単語 A の係り受け確率 p は以下の式で表される。

$$p = \frac{A \text{ と } C \text{ の係り受け回数}}{C \text{ と係り受けする全単語の係り受け回数の総和}} \quad (1)$$

$$p = \frac{A \text{ と } [S] \text{ の係り受け回数の総和}}{C \text{ と係り受けする全単語と } [S] \text{ の係り受け回数の総和}} \quad (2)$$

係り受け確率 (1) は、特徴語と係り受けする全単語の係り受け回数の総和のうち、③で取得した語と特徴語の係り受け回数の占める割合を表しており、特徴語と係り受け解析で係り受けする全ての単語のうち、ある単語が特徴語と係る確率を意味している。

係り受け確率 (2) は、特徴語と係り受けする全単語の中で②で取得した文脈における類義語の係り受け回数の総和のうち、③で取得した語と②で取得した文脈における類義語の係り受け回数の占める割合を表しており、特徴語と係り受け解析で係り受けする全ての単語のうち、ある単語が②で取得した文脈における類義語と係る確率を意味している。

### ⑥ 喩辞の推薦

④で取得した類似度と⑤で取得した係り受け確率 (1) それぞれを標準化して線形和を求め、値の降順に喩辞を提示し、これを手法 (1) とする。同様に、④で取得した類似度と⑤で取得した係り受け確率 (2) それぞれを標準化して線形和を求め、値の降順に喩辞を提示し、これを手法 (2) とする。

提案手法において、word2vec としては、日本語版 Wikipedia をコーパスとした日本語 Wikipedia エンティティベクトル [12] を利用した。また、青空文庫をコーパスとして係り受け解析を行い、各単語に係る単語の品詞、回数を抽出した係り受けデータベースを構築した。そして、単語のカテゴリを判定するためのカテゴリデータベースとして日本語 Wikipedia を対象に LDA を行った。

## 3 結果と考察

特徴語として形容詞と動詞それぞれ3種類について比喩生成を試みた。手法 (1) で生成した喩辞の上位10個と手法 (2) で生成した喩辞の上位10個を出力した。

表1に喩辞の出力例を示す。

「手が氷のように冷たい」や「瞳が星のように輝く」といった典型的な比喩に加えて、「風が笑い声のように心地よい」や「人生は恋のように儂い」といった新奇だが理解できる比喩が生成された。

また、手法 (1) で生成した場合と手法 (2) で生成した場合を比較すると、手法 (1) の方が意味の通る有効な比喩が多く生成された。これは、手法 (2) において、特徴語の類義語として得られた単語とは係るが、特徴語とは係らない単語が得られたからであると考えられる。例えば、「香りが広がる」という文脈における「広がる」の類義語として「満ちる」が得られたが、そのために「心」や「力」といった「満ちる」とは係るが「広

表 1: 喩辞の出力例

特徴語の品詞	入力		出力	
	被喩辞	特徴語	喩辞(1)	喩辞(2)
形容詞	手	冷たい	水	水
			氷	氷
			雪	雪
	風	心地よい	肌	肌
			笑い声	声
				夢
	人生	儂い	夢	夢
			望み	望み
			恋	恋
少女				
動詞	瞳	輝く	星	星
			太陽	太陽
			夕陽	夕陽
			陽光	
	顔	青ざめる	死人	死人
			屍	屍
	香り	広がる	海	海
			空	空
			青空	

がる」とは係らない単語が出力された。ただし、「風が心地よい」では手法(1)では出力されなかった有効な喩辞が手法(2)で生成された。これは他の場合とは逆に、「風が心地よい」という文脈における「心地よい」の類義語として得られた「愉快」や「良い」に係りやすい単語として有効な喩辞「夢」が出力された。

## 4 結論と課題

本研究では、被喩辞と特徴語の入力に対して喩辞を生成する2つの方法を提案した。どちらの提案手法でも典型的な喩辞に加えて新奇な喩辞を生成することが出来たが、手法(1)の方が有効な語彙が得られやすかった。

今後の課題は、有効な喩辞の割合の向上である。例えば「冷たい」「白い」といった特徴を属性として普遍的に持つ喩辞を高い精度で提示することや抽象度の高い単語をいかに排除するかについて検討する必要がある。

また、どのような場合で「風が心地よい」のように手法(2)で有効な喩辞が生成されるか検討する必要がある。

## 参考文献

- [1] 柳澤秀吉, 村上存, 大富浩一, 穂坂倫佳. 感性の多様性を考慮した感性品質の定量化手法. 日本機械学会論文集 C 編, Vol. 74, No. 746, 2008.
- [2] 荒井亮人, 柳澤秀吉. 多言語の概念空間構築法の開発 (国際デザインプロセス支援への応用). 日本設計工学春季大会講演論文集, 2018.
- [3] 柳澤秀吉. 多言語意味ネットワークの開発 (曖昧な語彙の意味分析と翻訳妥当性の評価). 日本機械学会論文集, pp. 18-00129, 2018.
- [4] Yanagisawa H. Multilingual semantic networks for kansei study. *Advances in Intelligent Systems and Computing*, Vol. 739, , 2018.
- [5] 鈴木理沙, 長島愛, 互惠子, 柳澤秀吉. 感性品質の言語化のための語彙獲得手法の研究 - 化粧品の感性品質を事例として -. 日本感性工学会大会講演論文集, 2019.
- [6] Fellbaum C. Wordnet: An electronic lexical database. *The MIT Press*, 1998.
- [7] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *In Proceedings of Workshop at ICLR*, 2013.
- [8] 佐藤遼河, 杉本徹. コーパスとシソーラスを用いた喩辞生成. 人工知能学会第 32 回全国大会論文集, pp. 1-4, 2018.
- [9] 中條寛也, 松吉俊, 内海彰. 意味空間に基づく文脈情報を用いた喩辞生成. 情報処理学会研究報告, Vol. 2017-NL-231, No.14, pp. 1-10, 2017.
- [10] Daniel E Berlyne. Novelty, complexity, and hedonic value. *Perception & Psychophysics*, Vol. 8, No. 5, pp. 279-286, 1970.
- [11] 楠見孝. 喩辞理解の構造. 誠信書房, 1990.
- [12] 鈴木正敏, 松田耕史, 関根聡, 岡崎直観, 乾健太郎. Wikipedia 記事に対する拡張固有表現レベルの多重付与. 言語処理学会第 22 回年次大会 (NLP2016), 03 2016.