

単語分散表現と喩辞の意味カテゴリを用いた喩の作成支援

佐藤 遼河[†] 浅見 一樹[‡] 杉本 徹[†]

芝浦工業大学工学部[†] 芝浦工業大学大学院理工学研究科[‡]

1. 研究背景と目的

近年、小説投稿サイトの利用やセルフパブリッシングにより個人での執筆活動を公開する機会が増えている。それに伴い、読者にイメージが伝わりやすい表現を用いることへの需要が高まっている。イメージが伝わりやすい表現方法の一つとして喩が挙げられる。しかし喩える語（喩辞）と喩えられる語（被喩辞）の関連性を考慮した上でことば選びをする必要があるため、喩を作成することは難しい。

本研究は文章中のユーザが指定した語に対応する喩を生成し、ユーザに提案するシステムを構築することが目的である。システムが喩の作成を支援することで、ユーザは容易にイメージが伝わりやすい表現を用いることができるようになる。喩は直喩、隠喩、換喩、提喩に分類できるが本研究では直喩のみを扱う。

喩の自動生成を目指した研究としては[1][2]などがあるが、本研究では喩辞を求めるときに特徴語と喩辞のカテゴリを考慮するという点に特徴がある。

2. 提案手法

提案手法の処理の流れを図1で示す。

はじめに、ユーザは被喩辞と喩として表現したい特徴（特徴語）を入力する。システムは特徴語・カテゴリ行列を参照して、特徴語と共起しやすい単語のカテゴリを得る。ここから被喩辞のカテゴリを除いたものが喩辞のカテゴリとなる。また、システムは被喩辞の単語ベクトル V 被喩辞と特徴語の単語ベクトル V 特徴語を取得し、それらを加算してベクトル V 喩辞候補を得る。このベクトルとコサイン類似度が高い単語のうち、喩辞のカテゴリに属す単語を喩辞として出力する。

2.1 特徴語・カテゴリ行列

特徴語を行、分類語彙表[3]に基づく名詞のカテゴリを列とする共起行列である。

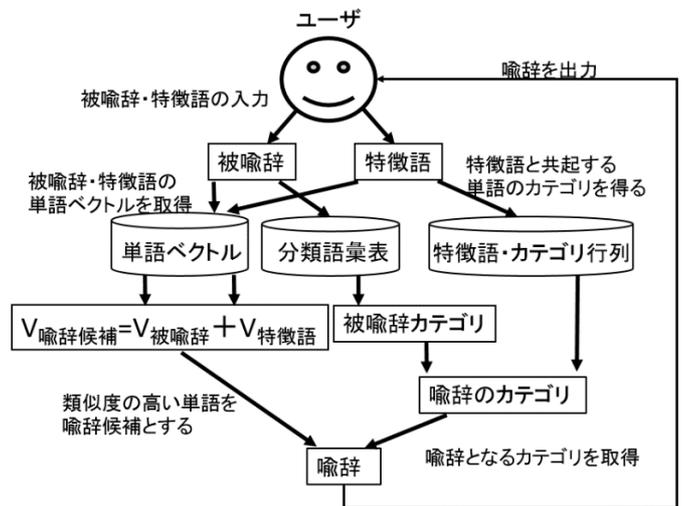


図1 提案手法の概要

行列を作成するために、まずコーパス中の文章を係り受け解析し、特徴語と共起する名詞を含むカテゴリを抽出する。特徴語が形容詞と形容動詞の場合は係り先、動詞の場合は係り元と係り先の文節中の名詞を共起する名詞とする。本研究ではコーパスとして青空文庫の文章を約313MB用いた。また係り受け解析にはCaboCha[4]を用いた。

行列の要素 a_{ij} は以下の式で求める。

$$a_{ij} = \frac{\text{特徴語とカテゴリ}j\text{の単語が共起する回数}}{\text{カテゴリ}j\text{の単語の総出現回数}}$$

特徴語を表す行において要素の値が大きいものから順に喩辞のカテゴリの候補とする。

2.2 単語ベクトル

単語の分散表現ベクトルはWord2Vec[5]を用いて200次元のベクトルを生成した。Word2Vecはコーパスからニューラルネットワークの学習を行い、単語の意味を表す固定長ベクトルを獲得する手法である。本研究ではコーパスは前節で述べた青空文庫の文章において単語の活用形を基本形に直したものをを用いた。

Support for metaphor creation using distributed representations of words and semantic category of vehicles

[†]Ryoga Sato, [‡]Kazuki Asami, [†]Toru Sugimoto

[†]College of Engineering, Shibaura Institute of Technology

[‡]Graduate School of Engineering and Science, Shibaura Institute of Technology

3. 評価実験

3.1 生成された喩辞に対する評価

3.1.1 評価方法

生成された喩辞に対してアンケートによる定量的な評価を行った。10種類の被喩辞と特徴語のペアをシステムに入力し、出力された喩辞を5人の被験者にすばらしい(4)、適切である(3)、判断が難しい(2)、不適切(1)の4段階で評価をしてもらった。

本実験ではシステムが生成する喩辞をコサイン類似度が0.5以上かつ各カテゴリの上位5位以内の単語とし、最大で50語を出力した。

3.1.2 結果

表1に喩辞の出力結果の例を示す。

表1 喩辞の出力例

被喩辞：ひとみ 特徴語：光る		
カテゴリ	生成された喩辞	
天地	二日月	夜空
物質	黒曜石	夕焼け
資材	たいまつ	イルミネーション

次に喩辞の評価結果を表2に示す。

表2 喩辞に対する評価の割合

喩辞数	4	3	2	1
477個	4.2%	13.1%	12.3%	70.4%

生成された喩辞の17.3%が適切であるという結果が得られた。また入力した被喩辞と特徴語のペアのうち、コサイン類似度が0.5を上回る喩辞の候補が少なかったため、50個の喩辞が得られないものが3組あった。

3.2 比喩の有用性の評価

3.2.1 評価方法

本実験では前節の実験で評価が高かった喩辞を被喩辞と特徴語のペアごとに上位10個ずつ用いて有用性の評価を行う。比較対象として文芸サークルに所属する大学生が作成した喩辞を被喩辞と特徴語のペアごとに10個ずつ用いる。6人の被験者に被喩辞と特徴語を提示し、システムが生成した10個の喩辞と人が作成した10個、合計20個の喩辞をランダムに提示する。被験者は比喩として用いたいと思う表現を提示された喩辞の中から5個選択し回答する。

本実験で用いた喩辞の例を表3に示す。

表3 評価に用いた喩辞の例

被喩辞：少女 特徴語：美しい	
システム	人
令嬢	サクラ
姫	女神
聖女	満月
騎士	初雪
女王	オーロラ
ヒロイン	星座
女優	花火
舞姫	宝石
花	夜空
人形	ダイヤモンドダスト

3.2.2 結果

被験者が選んだ喩辞のうちシステムが生成した喩辞は35.3%、人が作成した喩辞は64.7%であった。このことからシステムが生成した喩辞に一定の有用性があることが分かった。

システムが生成した喩辞と人が作成した喩辞のうち、共通して得られたのは「被喩辞：ラジオ、特徴語：鳴る」に対する喩辞「雷」のみだった。この点から本システムは人が比較的思いつきにくい喩辞を提案できることが分かった。

4. 結論

本研究では被喩辞と特徴語の入力に対して意味カテゴリを考慮して喩辞を生成する方法を提案した。また、今回行った実験により提案手法に一定の有用性があることが分かった。提案手法を用いることでユーザは容易に比喩を作成できるようになると期待できる。

今後は出力にあたって喩辞のカテゴリと喩辞ベクトルのコサイン類似度の2種類のパラメータを設定できるため、より良い結果が得られるパラメータの値を調査し、比喩の適切さを向上したい。

参考文献

- [1]北田純弥,萩原将文:電子辞書を用いた比喩による文章作成支援システム,情報処理学会論文誌,Vol.42,No.5,p.1232-1241(2001)
- [2]中條寛也,松吉俊,内海彰:意味空間に基づく文脈情報をを用いた比喩生成,情報処理学会研究報告,Vol.2017-NL-231,No.14,pp.1-10(2017)
- [3]国立国語研究所:分類語彙表 増補改訂版,大日本図書(2004)
- [4]CaboCha <https://taku910.github.io/cabocha/>
- [5] T. Mikolov, K. Chen, G. Corrado, and J. Dean: Efficient Estimation of Word Representations in Vector Space. In ICLR Workshop(2013)