

連想メカニズムを用いた直喩の意味解析法 Semantic Analysis Method for Similes Using Association Mechanism

鞠山 大樹[†] 芋野 美紗子[†] 土屋 誠司[‡] 渡部 広一[‡]
Daiki Mariyama Misako Imono Seiji Tsuchiya Hirokazu Watabe

1. はじめに

近年、人間の日常生活において、コンピュータが良きパートナーとして機能することが求められてきている。それに伴い、コンピュータの知的化が進められているが、その一方で複雑化も進んでいる。そこで、誰にでもコンピュータを容易に扱うことのできるインターフェースが必要である。そのようなインターフェースの実現には、人間同士が日常的に行っている会話を、人間とコンピュータでも同様に行えることが望ましい。そのためにはコンピュータが人間の意図を理解し、曖昧な表現にも適宜対応することが求められる。

会話には様々な表現方法があり、その一つに比喩がある。比喩とは、ある単語の持つイメージを別の単語が持つイメージに置き換えて表現する手法である。人間は会話の中に比喩が用いられていても、単語の持つイメージを常識的に連想し、意味を解釈することができる。しかし、常識的な知識を持たないコンピュータにとって、人間と同様に比喩を理解することは困難である。そこで本稿では、人間の連想を模した連想メカニズムをコンピュータに与えることで、比喩の意味解析を行う手法を提案する。これにより、人とコンピュータのコミュニケーションをより円滑に行うことが可能になると期待される。

比喩では、例えられる語を被喩詞、例える語を喩詞と呼び、比喩であることを明示する語を指標と呼ぶ。また、比喩には直喩と隠喩がある。直喩とは、単語と単語が「みたいな」や「のような」などの指標で接続される比喩であり、「地獄みたいな生活」や「雪のような肌」のように用いられる。隠喩とは、「地獄の生活」や「雪の肌」のように、比喩であることを明示しない比喩である。本稿では、入力を比喩の中でも一般的な直喩に限定する。指標は出現頻度の高い「のような、のように、みたいな、みたいに、のごとく、のごとき、じみた」の7種とする。

本稿では、連想メカニズムを用いることで喩詞のイメージを表す語を抽出し、それらの語に重み付けを行うことで直喩の意味解析を行う。これにより、喩詞と被喩詞のイメージや関係性を考慮した意味解析が可能となる。

2. 関連技術

2.1 連想メカニズム

連想メカニズムとは、ある語から別の関連する言葉を連想したり、複数の言葉の間にある関連性を判断するといった人間が自然に行う思考をコンピュータで実現するために作られたものである。概念ベース^[1]と関連度計算方式^[2]を基に連想メカニズムは構築されている。

[†]同志社大学大学院理工学研究科
Graduate School of Science and Engineering, Doshisha University
[‡]同志社大学理工学部
Faculty of Science and Engineering, Doshisha University

2.1.1 概念ベース

概念ベースとは国語辞書などから機械的に構築した単語(概念)とその意味特徴を表す単語(属性)からなる知識ベースである。概念には属性とその重要性を表す重みが付与されており、87242語の概念が収録されている。概念と属性と重みの組み合わせの例を表1に示す。

表1 概念ベースの例

概念	属性, 重み
林檎	(果樹, 0.72), (赤い, 0.19), (花, 0.06), ...
雪	(大雪, 0.73), (白い, 0.15), (舞う, 0.06), ...

2.1.2 関連度計算方式

関連度計算方式とは、概念ベースに定義されている二つの概念間の関連の強さを定量的に表現する手法である。関連度は0.0から1.0の間で値が変動し、概念間の関連が強いほど大きな値を示す。

2.2 感覚・知覚判断システム

感覚・知覚判断システム^[3]とは入力する名詞に対してその名詞から連想される感覚語や知覚語を出力するシステムである。感覚語とは人間が日常的に用いる五感に関する形容詞あるいは形容動詞で、98語が定義されている。知覚語とはある語に対する一般的な印象を表す形容詞あるいは形容動詞で、104語が定義されている。例えば、「宝石」という入力に対して感覚語として「硬い」、知覚語として「美しい」、知覚語として「大切な」が得られる。

2.3 Web から自動構築した大規模格フレーム

Web から自動構築した大規模格フレーム^[4](以下格フレーム)とは、動詞とその動詞に係る名詞を用法ごとに整理したデータベースである。この格フレームは、Web上の約5億文の日本語テキストから自動的に構築されている。また格フレームに含まれる動詞の数は約5万語である。

この格フレームを用いることで、動詞からその動詞に結びつく名詞と用言の用法である格(デ格, ガ格など)を取得できる。また、名詞からその名詞に結びつく動詞格を取得することも可能である。

3. 直喩の意味解析法

提案手法では、まず入力された句を形態素解析の結果を基に喩詞、指標、被喩詞に分割する。次に、直喩を用いた句を変換する際に喩詞および指標と置き換わる語の候補となる語(候補語)を概念ベースおよび感覚・知覚判断システムを用いて取得する。それぞれの候補語に対して重み付けを行い、最後に重みが最も高い候補語を喩詞および指標に置き換えて、意味解析の結果を出力する。

3.1 候補語取得

直喩では喩詞の持つイメージで被喩詞の持つイメージを修飾する。例えば「林檎のような頬」という直喩では

「林檎」の持つ「赤い」というイメージで「頬」の色に関するイメージを修飾し、「赤い頬」という意味を示す。よって喩詞のイメージを表す語を候補語とすることが望ましい。候補語は二つの方法で取得する。

一つ目は概念ベースの属性を用いる方法である。概念ベースを用いる理由としては、喩詞の属性は喩詞の特徴を表す単語と考えられるからである。喩詞の属性を概念ベースより取得し、候補語とする。ただし、候補語の被喩詞への係り方を考慮し、被喩詞が体言の場合は形容詞、形容動詞を、用言の場合は形容詞、形容動詞、副詞を候補語とする。

二つ目は感覚・知覚判断システムを用いる方法である。喩詞に対して感覚・知覚判断システムを用いて、感覚語および知覚語を取得する。取得された感覚語、知覚語も喩詞の特徴を表す語と考えられるので候補語とする。

二つの手法により、例えば「林檎のような頬」という句では、喩詞である「林檎」の属性のうち形容詞、形容動詞である「甘酸っぱい、甘い、丸い、赤い、古い、高い、多い」と感覚・知覚判断システムから取得した「赤い、甘い、丸い」を候補語として取得できる。

3.2 重み付け

取得した候補語から最も被喩詞に係る語としてふさわしい語を選出するため、各候補語に重み付けを行う。

まず、各候補語に以下の二つの方法で重みを付与する。

一つ目の方法は属性取得で得られた候補語について行うものである。関連度計算方式を用いて被喩詞と候補語の関連度を算出する。その関連度に喩詞から取得した属性の重みを掛けた値を重みの基準値とする。関連度計算方式を使う理由としては、被喩詞と候補語の関連度が高いほど、その候補語は被喩詞への係り方が自然で、出力する語として正当である可能性が高いと考えられるからである。さらに、属性の重みを掛けることで喩詞と関係の深い語ほど重みを大きくする。例えば、属性取得から得られた候補語「高い」は被喩詞「頬」との関連度が 0.008 で、概念「林檎」の属性としての重みは 0.017 であるため、それらを掛け合わせた 0.00013 が本手法における重みとなる。

二つ目の方法は感覚・知覚判断システムを用いて得られた候補語について行うものである。被喩詞と候補語の関連度の算出を行った上で、その値に定数(補正係数)を掛けることで重みの補正を行う。なお、補正係数は実験的に求めた定数 6.0 とする。例えば、感覚・知覚判断システムから得られた候補語「赤い」は被喩詞「頬」との関連度が 0.031 であるため、補正係数を掛けた 0.186 が本手法における重みとなる。

次に二つの手段で得られた候補語の重みに対して以下の補正を掛ける。まず、被喩詞の属性を取得し、候補語の中に被喩詞の属性と一致するものがあれば、その候補語の重みに補正係数を掛ける。この場合、候補語は喩詞と被喩詞が共通して持つ特徴と考えられるため、重みを大きくする。次に格フレームを用いて、候補語が被喩詞にガ格で接続するかを調べ、ガ格で接続する場合は重みに補正係数を掛ける。例えば「林檎のような頬」は「赤い頬」と変換されることが期待される。この「赤い頬」は「頬が赤い」という「被喩詞+ガ+形容詞」の形に変

形できる。直喩においてこのような変形が多く見られるためガ格で接続する場合は重みを大きくする。例えば候補語「赤い」は被喩詞の属性に存在し、「頬が赤い」という表現も存在するため、補正係数を 2 回掛けた 6.696 が最終的な「赤い」の重みとなる。

4. 評価

直喩の入力句を合計 107 個用意し、本システムの評価を行った。評価は、一つの入力句に対して、出力結果が正しい場合は 2 点、どちらとも言えない場合は 1 点、正しくない場合は 0 点とし、被験者 3 名により評価した。一つの入力句に対して、合計点が 5 点～6 点の場合は○、3 点～4 点の場合は△、0 点～2 点の場合×とした。評価結果を表 3 に、出力結果の例を表 4 に示す。

表 3 評価結果

○	△	×
54.2%	10.3%	35.5%

表 4 出力結果の例

入力	出力結果	評価
林檎のような頬	赤い頬	○
炎のような恋	熱い心	○
薔薇のごとく赤い	美しく赤い	△
牛みたいな人	鈍い人	×
大仏みたいな顔	美しい顔	×

結果は、○が 54.2%、△が 10.3%、×が 35.5%となった。例えば、「薔薇のごとく赤い」という入力に対して「美しく赤い」と出力されるが、「とても赤い」という出力の方が自然であると考えられたため評価は△となった。また、「大仏みたいな顔」という入力に対して「美しい顔」と出力されるが、本来は「優しい顔」と出力されることが望ましいため評価は×となった。

5. おわりに

本稿では連想メカニズムを用いることで単語が持つイメージを反映し、単語同士の関係性を考慮して直喩の意味解析を行う手法を提案した。本手法の評価実験を行った結果約 54%という精度が得られた。精度向上などの課題はまだ残されているが、人間が連想によって直喩を理解するプロセスをシステムとして構築することができた。

謝辞

本稿の一部は、科学研究費補助金(若手研究(B)24700215)の補助を受けて行った。

参考文献

- [1] 奥村紀之, 土屋誠司, 渡部広一, 河岡司, “概念間の関連度計算のための大規模概念ベースの構築”, 自然言語処理, Vol.14, No.5, pp.41-64, 2007.
- [2] 荻原寛, 渡部広一, 河岡司, “概念ベース内の共起情報に着目した概念間関連度計算方式”, 信学技報, Vol.106, No.587, pp.17-22, 2007.
- [3] 渡部広一, 堀口敦史, 河岡司, “常識的感覚判断システムにおける名詞からの感覚想起手法”, 人工知能学会誌, Vol.19, No.2, pp.73-82, 2004.
- [4] 河原大輔, 黒橋禎夫, “格フレーム辞書の漸次的自動構築”, 自然言語処理, Vol.12, No.2, pp.109-131, 2005.