

概念ベースの二次属性を用いた直喩解析法

A method to analyze the simple expression using secondary attribute of Concept Base

丸山 礼文[†] 芋野 美紗子[†] 土屋 誠司[‡] 渡部 広一[‡]

Raimon Maruyama Misako Imono Seiji Tsuchiya Hirokazu Watabe

1. はじめに

近年、人のパートナーとなるコンピュータの活躍が期待されている。人は日常的に会話を用いてコミュニケーションを行っている。人とコンピュータが円滑なコミュニケーションを行うためにはコンピュータが人の会話表現を解釈することが必要である。

会話表現の一つに比喩があり、比喩とはある物事を別の物事を用いて表現する手法である。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。比喩の一種として直喩があり、これは単語と単語が「みたいな」や「のような」などの指標で接続される比喩である。本稿では、比喩の中でも直喩を対象とした意味解析を行う。

人は会話中に直喩が用いられた時、喩詞の特徴を連想することができ、喩詞の特徴を直接的に意味する別の語に置き換えることで意味解析をする。このような語の連想機能をコンピュータで実現する為に、語概念をその意味特徴を表す語(属性)の集合で定義している概念ベースが構築されている。この概念ベースを用いて、コンピュータの直喩の意味解析の実現が期待できると考えられる。

直喩表現の解析を目的とした研究として「連想メカニズムを用いた直喩の意味解析法」^[1]がある。この研究では喩詞の特徴を表す語を抽出し、直喩の意味解析を行っている。この研究の問題点としては、語の連想機能が弱い為、解析できない直喩が存在する点である。そこで、本稿では語の連想の幅を広げる為に概念ベースの二次属性を用いて、直喩の意味解析を行うことを目的とする。

2. 関連技術

2.1 概念ベース

概念ベース^[2]はある語を概念と定義し、概念の特徴を表す語である属性と、属性の重要さを表す重みの対で構成されている。ある概念 A は m 個の属性 A_i と重み $w_i (>0)$ の対によって式(1)で表現される。

$$\text{概念}A = \{(A_1, w_1), (A_2, w_2), \dots, (A_m, w_m)\} \quad (1)$$

ここで属性 A_i を概念 A の一次属性と呼ぶ。これら一次属性は、概念で構成されている。つまり概念 A の持つ属性 A_i を概念とみなし、更に属性を導くことができる。概念 A_i の持つ属性を元の概念 A の二次属性と呼ぶ。このように概念ベースは任意の次元までの属性連鎖集合により定義されている。

2.2 関連度計算方式

関連度計算方式^[3]とは2つの概念の関連の強さを定量的に表す手法である。関連度の値は0.0~1.0の実数値をとり、1.0に近づくほど関連が強い。

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

[‡]同志社大学理工学部
Faculty of Science and Engineering, Doshisha University

2.3 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、用言とその用言に関す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用言の数は約5万語である。この格フレームを用いることで、用言からその用言に結びつく名詞、格を取得できる。また、名詞からその名詞に結びつく用言、格を取得することも可能である。

2.4 感覚・知覚判断システム

感覚・知覚判断システム^[5]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と形容動詞で、「辛い」や「痛い」などである。知覚語とはある語に対する一般的な印象を表す形容詞と形容動詞で、「可愛い」や「汚い」などが存在する。

3. 直喩解析法

図1に直喩解析法の全体の流れを示す。本稿で提案する直喩解析法では、まず直喩を用いた句に対して形態素解析を行い、喩詞、指標、被喩詞に分割する。なお、本稿で扱う指標は使用頻度の高い「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。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(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。取得した候補語に対して最も喩詞の特徴を表し被喩詞を修飾する語を決めるために点数付けを行う。最後に喩詞を最も高い点数の候補語に置き換えて、意味解析の結果を出力する。

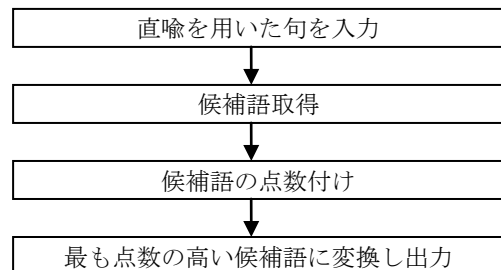


図1 直喩解析法の全体の流れ

3.1 候補語取得

まず1つ目の候補語取得方法は概念ベースの属性を用いる方法である。喩詞の属性は喩詞の特徴を表す語として考えられるため、概念ベースの属性を用いる。先行研究では喩詞の一次属性から候補語取得を行っていたが、本稿では喩詞の二次属性からも候補語取得を行う。二次属性は喩詞の一次属性の属性であるため、喩詞の特徴を

表す語が喩詞の二次属性に存在する可能性がある。よって、取得する候補語の拡張が期待できる。取得した属性のうち、喩詞の特徴を表し、被喩詞を修飾する語を候補語とする。被喩詞が体言の場合は修飾格となり得る形容詞と形容動詞を、用言の場合はそれに加えて連用修飾格となり得る副詞を候補語とする。

2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。取得した感覚語、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。

この2つの手法により、例えば「林檎のような頬」という句では、「林檎」の属性のうち形容詞、形容動詞である「甘い、丸い、赤い」と、感覚・知覚判断システムから取得した「赤い、甘い」が候補語として取得される。

3.2 候補語の点数付け

一次属性取得で得た候補語については、被喩詞と候補語の関連度を算出し、その関連度に喩詞に対する属性の重みを掛けたものをその候補語の点数とする。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との関連が高いので出力する語として適切である可能性が高いためである。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点数が大きくなると考えられる。

二次属性で得た候補語については、まず被喩詞と候補語の関連度を算出し、その関連度に一次属性の重みと二次属性の重みを掛ける。二次属性の重みは一次属性に対する重みなので、喩詞との関連性を考慮するために掛けている。図2に「大根のような足」と入力した場合の二次属性で得た候補語「太い」に対しての点数付けを行った結果を示す。

直喩：大根のような足

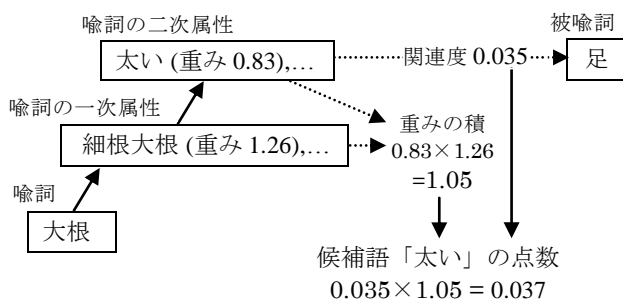


図2 二次属性で得た候補語に対しての点数付け

図2では、二次属性の候補語「太い」と被喩詞「足」との関連度 0.035 を算出する。そして、その関連度に「太い」の重み 0.83 と「細根大根」の重み 1.26 の積 1.05 を掛ける。結果、0.037 が「太い」の点数となる。

感覚・知覚判断システムから得られた候補語については被喩詞と候補語の関連度に定数（補正係数）をかけたものを点数とする。感覚・知覚判断システムの出力語は人が自然に感じる感覚や印象を表す語であるため、概念ベースから得られた候補語に比べて、喩詞の特徴をより表現していると考えられる。そのため補正係数をかけることで点数を大きくする。なお、補正係数は実験的に求めた定数 6.0 とした。

以上の方法で、取得した候補語群に対して、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る場合、その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。よって、その候補語の点数に補正係数を掛けて点数を大きくする。次に格フレームを用いて、候補語がガ格で接続する場合は点数に補正係数を掛ける。例えば「林檎のような頬」は「頬が赤い」という「被喩詞+ガ+形容詞」の形に変形することができ、意味を解釈することができる。直喩はこのような「被喩詞+ガ+形容詞」に変形することができる表現が多く見られるため、ガ格で接続する場合は点数を大きくする。

4. 評価と考察

直喩を含む 107 句用いて、本手法の評価を行った。評価は、1つの入力句に対して、出力が正しい場合は 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「×」とする。先行研究^[1]の評価結果と本稿における提案手法の評価結果を表 1 に、出力結果の例を表 2 に示す。

表1 評価結果

	○	△	×
先行研究	54.2%	10.3%	35.5%
提案手法	60.7%	9.4%	29.9%

表2 出力結果の例

入力	出力	結果
林檎のような頬	赤い頬	○
大根のような足	太い足	○
宝石のような時計	美しい時計	△
大仏みたいな顔	美しい顔	×

評価の結果、先行研究と比べて精度向上することができた。「大根のような足」は先行研究では評価×であったが提案手法では評価○になった。これは、提案手法によって二次属性から「大根」から候補語「太い」という正解となる候補語を取得することができたからである。

5. まとめ

本稿では概念ベースの二次属性を用いて、単語の持つ特徴を考慮した直喩の解析を行う手法を提案した。概念ベースの二次属性を用いることで、候補語取得の拡張を行うことができたことが精度向上に繋がった。

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] 鞠山大樹, 芋野美紗子, 土屋誠司, 渡部広一, “連想メカニズムを用いた直喩の意味解析法”, 情報科学技術フォーラム FIT2012, pp.195-196, 2012.
- [2] 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.
- [3] 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.
- [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.
- [5] 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.