

類似度と連想度を用いた隠喩表現判定に関する研究

関口 育栄† 中島 陽子‡ 本間 宏利‡ 梶井 文人*

† 釧路高専 電子情報システム専攻

‡ 釧路高専 創造工学科

* 北見工業大学

honma@kushiro-ct.ac.jp

1 はじめに

比喩とは、ある概念を他の概念によって説明または強調する修辭的技法の一つであり、心理学や言語学など様々な分野で研究対象として取り上げられている。自然言語処理の分野でも、機械翻訳における比喩表現の誤訳の回避などの目的で比喩の抽出・解析の研究が行われている [1]。

しかしながら、比喩に関する研究はその多くは直喩を対象にしたものであり、隠喩の抽出の研究は多いとはいえない。直喩に見られる「のような」「みたいに」などの比喩指標が存在しないことが隠喩の検出を困難にさせている。

隠喩抽出は検索エンジン等で使用されている誤字修正機能等への応用が期待される。本研究では、小説からの隠喩表現の抽出を目的とし、隠喩表現の判定手法を提案する。

2 隠喩の判定

本研究で扱う隠喩表現は「彼は鬼だ」のような「AはB」という形をとる文に限定し、AとBが隠喩関係にあるかを判定する。また、AとBはどちらとも名詞であるとする。比喩表現において、Aは被喩辞、Bを喩辞と呼ぶ。我々は、隠喩における喩辞と被喩辞には大きく分けて、以下の二つの異なる関係性が存在すると思う。

強調型隠喩 喩辞の特徴によって、被喩辞の特徴を強調している形式。

連想型隠喩 被喩辞と喩辞が共通するイメージを持っている形式。

強調型の例として、「彼は鬼だ」という隠喩表現があげられる。「彼」という人物を、「鬼」という言葉の

“怖い”、“恐ろしい”といった特徴を割り付け、強調させている。

連想型の例として、「時は金なり」という隠喩表現があげられる。被喩辞である“時”と、喩辞である“金”のどちらもが、“貴重である”、“大切なもの”というイメージを共通して持っている。

これらの関係性を踏まえ、隠喩関係の判定尺度には**類似度と連想度**を用いる。被喩辞の特徴を強調するためには、喩辞と被喩辞が全く関係性が無い単語同士では成立しないので、シソーラスを用いて意味的な距離を数値化する類似度を尺度として使用する。また、共通したイメージを持つことを数値化する尺度として連想度を使用する。

2.1 類似度

類似度はシソーラスにおける2単語間の距離を0から1の数値で表す尺度である。シソーラスとは単語や概念を木構造の階層に体系的に分類した辞書のことである [2]。シソーラスの一部を図1に示す。本研究では日本語のシソーラスである『日本語語彙体系』を使用している。単語がシソーラスの複数項目に分類されている場合、類似度も複数算出される。その場合は最も高い数値を採用する。

単語組 A, B の根からの深さをそれぞれ d_A, d_B 、二つの単語で共通している親からの深さを d_c とすると、類似度 sim は式 (1) で求めることができる。

$$sim(A, B) = \frac{d_c \times 2}{d_A + d_B} \quad (1)$$

強調型隠喩の類似度を算出するとは値は小さくなり、隠喩ではない非隠喩の類似度は大きくなると考えられる。

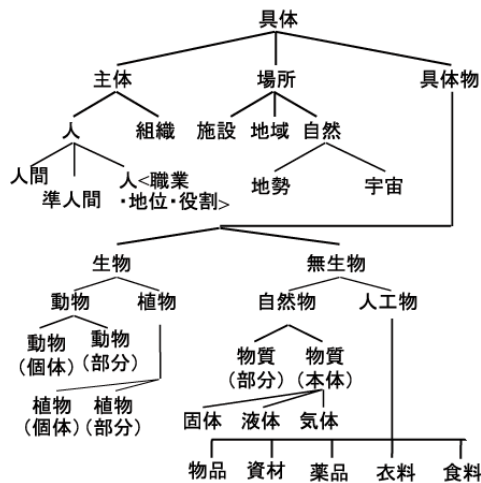


図 1: シソーラス (一部)

2.2 連想度

ある言葉から連想される言葉のことを**連想語**という。連想度は、二つの単語のそれぞれの連想語のうち、共通する連想語がどれだけあるかを表す数値である。連想語は kizasi ラボの連想語抽出ツール『keygram』を使用して抽出する [3]。連想語の例を表 1 に示す。

表 1: 連想語の例

単語	犬	猫
連想語	犬種	猫カフェ
	看板犬	黒猫
	犬小屋	家猫
	警察犬	猫缶
	飼い主	猫草
	チワワ	飼い主
	ペット	首輪
	保健所	ペット
	ドッグラン	保護
	首輪	白猫

単語 A と単語 B の連想語に全く共通しているものが無い場合の連想度を 0 とし、『共通する連想語の数』 $\times 0.5$ を連想度に追加する。また、単語 A が単語 B の連想語に含まれている、もしくは、単語 B が単語 A に含まれているときは、連想度に 1 を追加する。

表 1 の場合、共通する連想語が「飼い主」「ペット」「首輪」と三つ存在するので、連想度は $3 \times 0.5 = 1.5$ となる。

連想語は大量にとってしまうと本来の単語とは関係性が無いものの連想語としてしまうので、使用する連想語は単語一つにつき上位 15 個ずつ、単語組で計 30

個使用する。連想度は値が大きければ隠喩である可能性が高いと考えられる。

3 判定実験

3.1 類似度による強調型隠喩の判定手法

類似度によって強調型隠喩を判定できるかを確認するため、実験を行った。人手により作成した強調型隠喩に分類される単語組 100 組、連想型隠喩 100 組、非隠喩 100 組の、計 300 組の単語組を用意した。この 300 組の単語組の類似度を算出し、0.6 以下の単語組を強調型隠喩と判定する。

3.2 強調型隠喩の実験結果

類似度により強調型の隠喩判定の結果を表 2 にまとめた。強調型隠喩の判定結果は 100 組中 98 組と非常に高い精度で判定できたが、連想型隠喩も 88 組と、多く強調型隠喩と判定されてしまっている。また、非隠喩の単語組も 25 %ほど隠喩として判定されてしまっている。強調型隠喩の適合率と再現率はそれぞれ 0.46 と 0.98 となり、F 値は 0.63 となった。

表 2: 類似度による判定結果 [個]

	強調型隠喩	連想型隠喩	非隠喩
隠喩判定	98/100	88/100	25/100

3.3 連想度による連想型隠喩の判定手法

強調型隠喩と同様に、連想度による連想型隠喩判定が効果的であるかどうか実験を行った。強調型隠喩の実験に使用したものの同じ、強調型隠喩 100 組、連想型 100 組、非隠喩 100 組の 300 組の単語組を使用する。単語組の連想度を算出し、連想度が 0 より大きければ連想型隠喩と判定する。

3.4 連想型隠喩の実験結果

表 3 は評価結果をまとめたものである。三種類の単語組データで、連想型隠喩の隠喩判定個数が一番多く、連想型隠喩の判定に適性はあることがわかった。しかし、連想型隠喩 100 組の内、隠喩と判定されたものは 17 組と、隠喩表現の判定に使用するには精度が非常に低いことも判明した。評価としては適合率は 0.53、再現率は 0.17、そして F 値は 0.25 である。

表 3: 連想度による判定結果 [個]

	強調型隠喩	連想型隠喩	非隠喩
隠喩判定	6/100	17/100	9/100

3.5 隠喩表現判定実験

最後に、類似度と連想度を組み合わせた実験を行った。今までの実験で使用した 300 組の単語組データを使い、類似度による隠喩判定と連想度による隠喩判定を行う。どちらか片方でも隠喩と判定されれば、その単語組は隠喩であると判定する。

3.6 隠喩表現判定の実験結果

表 4 は類似度と連想度を組み合わせた場合の判定結果である。強調型と連想型合わせて 200 組の隠喩単語組の内、正しく隠喩と判定されたのは 189 組だった。適合率は 0.86, 再現率は 0.94, F 値は 0.90 と高い精度を示した。

表 4: 類似度と連想度による判定結果 [個]

	強調型隠喩	連想型隠喩	非隠喩
隠喩判定	99/100	90/100	30/100

これにより、類似度と連想度を組み合わせることで、隠喩判定の精度が向上することがわかった。

4 おわりに

本研究では「A は B」という文に対して、隠喩であるか非隠喩であるかの判定を行うため、隠喩表現の隠喩と被喩辞の関係性を考察し、強調型隠喩と連想型隠喩に分類した。そして、強調型隠喩の判定に類似度を、連想型隠喩の判定に連想度を、隠喩判定の尺度として設定した。

実験の結果、類似度、もしくは連想度単体での隠喩判定は高い精度ではなかったが、二つの尺度を組み合わせることで隠喩判定の精度向上に見込めることがわかった。

今後は計算式の見直しや連想度の値を考慮するなどして、連想度による隠喩判定の精度向上を行う。また、より多くの単語組データを使用した場合、精度がどのように変化するか実験していく。

参考文献

- [1] 榎井文人, 福本淳一, 椎野努, 河合敦夫. 確率的尺度を用いた比喻性検出手法. 自然言語処理, 2002.
- [2] 奥村学, 自然言語処理の基礎, コロナ社
- [3] kizasi.jp ラボ 言葉を見つける keygram, "http://kizasi.jp/lab/keygram/keygram.py/"