

Webからのメタファ表現候補の抽出システム

Extraction System of Metaphor Expression from Web.

間中 宏修十 山田 剛一十 絹川 博之十
Hironobu Manaka Koichi Yamada Hiroshi Kinukawa

1. はじめに

実世界に存在する言葉を理解する上で、既存の知識やイメージを用いたメタファ表現を使用することが有効である。Web上には数多くのメタファ表現が存在するが、それらを自動的に抽出する手法は確立されていない。人間が利用しやすい形式でメタファ表現を提供するシステムが求められている。

本研究では、Webからメタファ表現を自動的に抽出することを目的とする。そのために、最初のステップとしてメタファ表現と近い意味で使用される直喩表現の抽出パターンを生成する。その上で、それを元にしたメタファ表現候補の抽出パターンを導き、メタファ表現候補を抽出するシステムを開発し評価する。

2. 直喩表現の利用

直喩表現は「あたかも」、「のよう」などの定型語を用いて、その言葉が比喩であることを明示している表現である。それに対してメタファは、定型語を使用せず、比喩であることが明示されていない表現であり、表現構成語の特徴からメタファを抽出することは困難である。

メタファ表現には対応する直喩表現が存在することから、直喩表現の特徴(文型や被喩辞と喩辞との関係など)を分析することにより、メタファ表現を抽出するためのパターンを導くことができると考えられる。

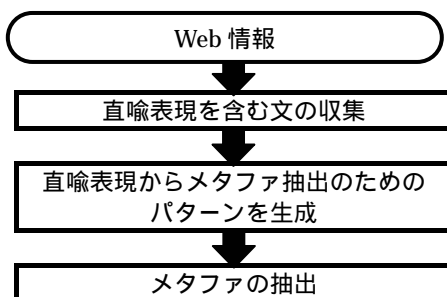


図1. メタファ抽出の処理過程

3. 直喩表現の特徴の分析

まず、Web情報から直喩表現になりえる文の収集を行い、その中から直喩表現の抽出を行う。その過程で得られるパターンを元にメタファの抽出パターンを検討する。

3.1 直喩表現を含む文の収集

Web検索エンジン Google[1]より、直喩表現の構文パターンをクエリとした検索を行い、その結果から直喩表現を含む文を収集する。典型的な直喩表現は、図2に示すような直喩表現パターン「AはBのよう」となっている。ここで、Bの概念(喩辞)がAの概念(被喩辞)を喩えている。

| | | | |
|-------|------|------|-------|
| 言葉 | は | マジック | のよう |
| [被喩辞] | [助詞] | [喩辞] | [定型語] |

図2. 典型的な直喩表現パターン

Googleの検索エンジン[1]は、品詞情報を考慮した文字列の検索をすることができない。よって品詞情報以外の構文情報から得られる特徴を元に直喩表現を収集する。AとBの部分は形態素3語以内の文字列とする。AとBの間に存在する助詞は「は、が、を、も」、定型語は「みたいに、のようだ」など多くのパターンがあるが、今回は直喩表現で最も使用頻度が高い「AはBのよう」を処理の対象とする。よって検索クエリは「"***は***のよう"」とする。

3.2 連辞的な直喩表現の抽出過程

3.1で収集した文の中から連辞的な直喩表現の抽出を行う。収集した文の中にはAとBの間に存在する助詞が「は」だけではないものが存在する。これは収集時に「***」の文字列の品詞情報を考慮していないためである。今回の調査ではAの直後の助詞が「は」以外のものは対象外の文型であるため扱わない。

3.2.1 被喩辞・喩辞候補の抽出

まず、直喩表現を含む文におけるAとBの部分の絞込みを行う。ここでは、助詞「は」の直前に存在する(複合)名詞を被喩辞候補、定型語「のよう」の直前に存在する(複合)名詞を喩辞候補とする。ただし、以下のように被喩辞・喩辞候補の対象が何を示しているか判断ができない場合は被喩辞・喩辞候補としない。

(1) 代名詞 例: あなたは水のような女性ですね。

(2) 指示語 例: 生活習慣病は次のようなものです。

代名詞は形態素解析ソフト MeCab[2]による品詞情報を用いて、指示語はあらかじめリストを用意することで判定を行う。

3.2.2 連辞的な表現の判定

AとBが連辞的な関係にある文の判定を行う。ここで

† 東京電機大学大学院, Tokyo Denki University

は、図3のようにAとBの係り先が同じである文のみを連辞的な表現とする。図4のようにAとBの係り先が異なる文は連辞的な表現ではないものとする。文節の係り受け関係は、Cabocha[3]を用いて判定を行う。

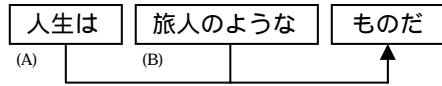


図3. 連辞的な表現の例

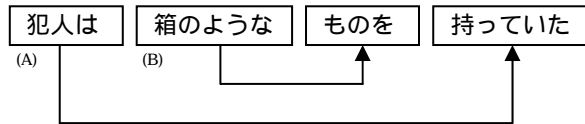


図4. 連辞的ではない表現の例

3.2.3 直喩表現の判定

「のような」の用法が直喩表現として使用されているかの判定を行う。3.2.2 までの処理において絞り込まれた文はすべて「AはBのようなC」(A, B, Cはすべて名詞)である。よってA, B, Cの関係から直喩表現とそれ以外の表現の分類を行った。

(1) 直喩表現である文

(a) AとBの共通の属性がCである表現
この場合、「AのCはBのCのようだ」と言い換えることができる。

例：マンステールのニオイは納豆のニオイのようだ。
(A) (C) (B) (C)

(b) Cが「もの」である表現
AとBが全体として類似している場合、あるいはAとBの類似点が明示されない場合に用いられる表現である。

例：戦争はオセロのようなもの。
ただし、AとBの概念が近い場合には、直喩表現ではあるがメタファ表現にはなりにくい。
例：ライオンはトラのようなもの。

(2) 直喩表現でない文

(a) 例示
Aと類似する一例がBによって示される表現である。例示は直喩表現と異なり、BとCの概念が包含関係になる特徴を持っている。

例：ソウルは東京のような大都市なので...

(b) 「AはCだ」がコピュラ文となっており、AとCの概念が包含関係にある表現

「Aは(BのようなC)だ」という構造で、「Bのような」がなくても文が成り立つ。

例：PlayStation3は夢のようなマシンだ。

(c) おおよそ等しいことを示す表現
「ほとんどB」であることを表現している文であり、Bで喩えているわけではない。

例：東京は嵐のような日でした。

4. 実験

ここでは3章で述べた手法を元の実装した直喩表現候補の抽出システムの処理結果と人手で分類した結果とを比較する実験を行った。結果を表1に示す。

表1. 「AはBのような」の文における分類結果

| 項目 | | |
|----------------|-----|----|
| 直喩表現候補(下記以外の文) | 74 | 42 |
| 被喩辞・喩辞の候補がない文 | 105 | 0 |
| 連辞的な表現ではない文 | 74 | 0 |
| 合計 | 253 | 42 |

: 実装したシステムによる処理結果

: 連辞的な直喩表現の件数

表1より、直喩表現候補を74件にまで絞り込むことができた。今後は42件の連辞的な直喩表現を抽出するためにさらなる絞り込みを行う。

5. おわりに

本論文では、直喩表現パターンの特徴を分析し、その分析結果を元に連辞的な直喩表現を自動的に抽出するためのシステムを実装した。

今後は「のように」、「みたいだ」など、他の直喩表現の定型語についての調査を行い、パターンの生成を行い、メタファ抽出プログラムを作成する。作成したプログラムについて実験評価し、改良していく予定である。

6. 謝辞

本研究で使用している、形態素解析エンジン MeCab[2]、係り受け解析ソフト Cabocha[3]を開発された方々に感謝致します。

参考文献

- [1]Google ウェブ検索, <http://www.google.com>
- [2]形態素解析エンジン MeCab, <http://mecab.sourceforge.net>
- [3]係り受け解析ソフト Cabocha, <http://chasen.org/~taku/software/cabocha/>
- [4]山梨正明, 岩田純一: 比喩と理解, 東京大学出版会
- [5]George Lakoff and Mark Johnson. : Metaphors We Live By. University of Chicago Press, 1980.
- [6]S. Krishnakumaran, X. Zhu : Hunting Elusive Metaphors Using Lexical Resources, In NAACL 2007 Work-shop on Computational Approaches to Figurative Language, pp.13-21, 2007.