

評判分析における品詞情報と意味類型情報の有効性比較

岡田 正平 山本 和英

長岡技術科学大学 電気系

{okada, yamamoto}@jnlp.org

1 序論

日本語における用言は動詞と形容詞に分類される¹。その分類基準はその終止形が母音ウを持つ拍で終わる(動詞)、拍「い」で終わる(形容詞)という形態的なものである。また、一般的に動詞は客観的に観測可能な人や物の動きや変化を、形容詞は物事の特徴や性質を表すとされている。しかし形態的な基準に基づく品詞分類は、その品詞に属する語の意味を完全に分離しているわけではない。例えば動詞の中には「優れる」のような物事の特徴を表す語や「むかつく」のような人の感情を表す語が存在している。したがって、テキストから動作のみを抽出する目的ですべての動詞を抽出すると動作とは無関係な上記のような動詞が含まれてしまう、などの問題が生じることになる。

このような問題に対し、本研究室の先行研究において中山ら [1] は形態上の特徴に関係なく、その意味によって用言を「作用性用言」と「形状性用言」に分類することを提案した。そのうえで、分類の前段階として動詞、形容詞と作用性用言、形状性用言の中間概念と言える4種類の意味類型(動作, 変化, 感覚・感情, 形容)を定義した。また中山らは、用言のうち形容詞は常に意味類型の「形容」のみを取り、動詞はその語義や用法により同じ動詞でも異なる意味類型を取り得るとし、IPA 品詞体系日本語辞書⁽¹⁾中の全ての動詞に対して、取り得る意味類型を人手により付与したSNOW D1: 動詞意味類型辞書を作成し、公開した²。

また我々 [2] は、文中の動詞の意味類型はその文脈から1つに決定できると仮定し、形態素解析結果と既存の言語資源から得られる情報を素性とした教師あり機械学習を用いて文中の動詞の意味類型を1つに決定する意味類型曖昧性解消器を作成した。

これらの先行研究により形態素解析済みの日本語文中のすべての用言に意味類型を付与することができるようになった。用言に付与された意味類型情報は動詞・

形容詞という表層的な情報を用いた自然言語処理を行うタスクにおいて、品詞情報の代わりに利用することができる。応用例としては評判分析や感情解析、テキスト中からの動作や変化の抽出などが挙げられる。本研究では評判分析の例をとり意味類型の有用性を示す。具体的には、品詞の代わりに意味類型を用いることでブログ記事中から評判表現を含む文の抽出の精度を4.7ポイント、F値を27.2ポイント向上させた。

2 関連研究

品詞情報の代わりに意味類型情報を利用することは、語の意味をより正確に扱うことにつながり、性能向上に寄与できると考える。本研究では意味類型を評判分析に応用することでその有用性を示す。

近年、個人がWeb上に意見を発信することが可能になったことに伴い、テキストを対象とした評判分析・感情解析への関心が高まっている。品詞情報を用いている研究としては以下のような例が挙げられる。Scholzら [3] は語ごとに計算されるエントロピーに基づく指標を品詞毎に平均し素性とする手法で、新聞記事中の文に対する肯定/否定/中性(客観)の極性推定で最高65%の精度を達成している。また、小林ら [4] は意見を(対象, 属性, 評価値)という3つ組で定義、共起パターンに基づく属性・評価値表現の半自動抽出方法を提案しており、抽出時のノイズ軽減を目的としたフィルタリングの1つに品詞情報を用いている。感情解析においては、Nakagawaら [5] は依存構造木のノードの極性を表す隠れ変数と、事前極性や極性反転の有無、品詞等の組み合わせを素性として用いた条件付き確率場を用いて文の感情極性を2値分類する手法を提案し、日本語を対象に最高84.7%の精度を達成している。

品詞情報を用いない手法の例としては、Socherら [6] は再帰的ニューラルネットワークをテンソルの導入によって改良したRecursive Neural Tensor Networkによって、文レベルでの感情極性2値分類で85.4%の精度を達成している。

¹本稿では特に言及しない限り形容動詞を形容詞に含めて考える。

²<http://www.jnlp.org/SNOW/D1>

3 意味類型の評判分析への応用

中山らは日本語の用言を意味的に分類することを目的として、動作、変化、感覚・感情、形容という4種類の意味類型を定義した。文中の用言はその用言が持つ意味によって1つの意味類型を持つと考える。また、同じ語であっても文脈によって持つ意味類型が異なる場合がある。例えば「コップに水を満たす」の「満たす」は意味類型の動作を、「条件を満たす」の「満たす」は意味類型の形容を持つ。現在までの我々の先行研究により、形態素解析結果から文中の用言に対してその用言が持つ1つの意味類型を付与することが可能になっている。そこで本研究では意味類型を実際の自然言語処理タスクに応用し、その有効性を示す。

応用例として本研究では評判文抽出を扱う。評判表現の多くは「良い」「～が多い」などの形容表現が用いられる。意味類型の情報をを用いることで「～に優れる」といった動詞の形容表現を扱うことができる、また「～を感じる」「～だと思う」のような意味類型の感覚・感情に対応する表現も評判になりやすい。このようなことから、意味類型情報を利用することでより網羅性高く評判表現を扱うことができ、性能向上に貢献できるという仮説を立てた。

対象とするコーパスには249記事、4186文からなる人手による解析済みブログコーパスであるKNBコーパス⁽²⁾を用いた。本研究における評判表現の定義はKNBコーパスに従い「ある対象に対して述べられた肯定的、もしくは否定的判断や態度、叙述」とする。KNBコーパスは評判表現についてアノテーションされており、「感情」「批評」など7種の評判タイプや評価の極性、評価保持者、評価対象といった詳細な情報も付与されている。しかし、本研究の目的は純粋な品詞情報と意味類型情報の比較であるため、評判表現の詳細な情報についての予測・同定は行わない。本研究ではKNBコーパス中で評判表現を含む文を評判文、含まない文を非評判文と定義し、各文を評判文/非評判文に分類する2値分類タスクを考える。

品詞情報のみを用いた場合と意味類型情報を用いた場合とを比較するため以下に示す手法を用いて実験を行う。

3.1 評判文抽出手法

まず、入力文をMeCab⁽³⁾を用いて形態素解析したあと以下に示す形態素解析の後処理を行う³。

³KNBコーパスは形態素アノテーション済みであるが、意味類型曖昧性解消器の入力がIPA品詞体系日本語辞書の品詞体系に基

- 直後に原形が「する」の動詞を伴うサ変接続の名詞は動詞として扱う
- 直後に助詞「を」と原形が「する」の動詞を続けて伴うサ変接続の名詞は動詞として扱う
 - － 上記2つに当てはまった場合、動詞「する」は用言として扱わない
- 直後に原形が「だ」「です」の助動詞を伴う形容動詞語幹の名詞は形容詞として扱う
- IPA品詞体系日本語辞書上の品詞細分類が自立でない動詞(「～している」の「いる」や「～し得る」の「得る」など)は用言として扱わない

この後処理はそれぞれ次に示す目的に基づく⁴。

- 意味類型曖昧性解消を「する」ではなく、語義を担うサ変名詞に対して行う
- 形容動詞(形容動詞語幹(名詞)+「だ/です」(助動詞))を形容詞として扱う
- 事実上機能語として振る舞う動詞を用言から除外する

その後、用言として扱われる動詞に対しては意味類型曖昧性解消器を用いて意味類型を付与し、形容詞に対しては意味類型の「形容」を常に付与する。

次に品詞情報、または意味類型情報を用いてコーパス中の全文から評判文の抽出を行いそれぞれの結果を比較する。先述のとおり評判表現の多くは形容表現が用いられていること、また、動作を表す表現は結果として客観的な事象を表していると考えられることから、以下に示す規則を用いて評判文の抽出を行う。

品詞情報のみを用いた規則 文中の最後に出現する用言の品詞が動詞であれば非評判文であると、形容詞であれば評判文とする。

意味類型情報を用いた規則 文中の最後に出現する用言の意味類型が動作であれば非評判文であると、変化、感覚・感情、形容のいずれかであれば評判文とする⁵。

づいていることを前提としているため再度形態素解析を行っている。

⁴なお、この後処理を行わずに品詞情報のみを用いて評判文抽出を行う場合、後述する結果と比較して悪い結果を示した。具体的には分類精度が0.574、F値が0.310となった。

⁵意味類型が変化、感覚・感情の場合に評判文とするのは予備実験の結果に基づく。

表 1: 品詞情報のみを用いた評判文抽出

正解\予測	評判文	非評判文	再現率 [%]
評判文	533	1512	26.1
非評判文	177	1964	91.7
適合率 [%]	75.1	56.5	59.7

評判文抽出の観点からの F 値=38.7%

表 2: 意味類型情報を用いた評判文抽出

正解\予測	評判文	非評判文	再現率 [%]
評判文	1442	603	70.5
非評判文	889	1252	58.5
適合率 [%]	61.9	67.5	64.4

評判文抽出の観点からの F 値=65.9%

文中の1つの用言だけを見るという単純な規則を用いるのは、結果に寄与する要因を減らし品詞情報と比較した意味類型情報の特徴を明らかにする意図に基づいている。また、最後の用言に注目しているのは、日本語文の主要部は後ろに出現しやすいという仮定に基づいている。

なお {品詞が動詞の用言} ∩ {意味類型が動作の用言} であるので、意味類型を用いた場合は必ず評判文に対する再現率が上昇することに留意されたい。

3.2 結果・考察

表 1, 2 に品詞情報のみを用いた場合と意味類型情報を用いた場合の評判文抽出結果をそれぞれ示す。意味類型情報を用いた場合、品詞情報のみを用いた場合と比べて評判文/非評判文の分類精度において 4.7 ポイント、評判文抽出の観点から見た F 値において 27.2 ポイント上昇している。以下、各種誤りについて解析した結果について述べる。

品詞情報のみでは誤りで、意味類型情報を扱うことで正しく評判文であると判断された例は 909 文存在した。誤りには「～がわからなかったです。」「～と違います」「～と感じました」など、意味類型が感覚・感情の動詞で終わる文を多く含んでおり、動詞をその意味によって分類するという意味類型導入の意図が反映されており、有用性が確認できた。

品詞情報、意味類型情報どちらを用いた場合にも共通して誤りだった例は 780 文存在した。うち評判文を非評判文としてしまった例は 603 文、非評判文を評判文と判定してしまった例は 177 文だった。さらに前者について誤りの原因の解析を行った。全 603 文から 50 文を無作為に抽出した解析を行った結果の内訳は以下のとおりである。

意味類型が動作でも評判表現になる例 (16 件) 以下に示す表現においては下線部の動詞⁶の意味類型によらず評判表現になり得る。このような誤りはパターンマッチで個別に扱うことで解決できる。

⁶下線部が 1 動詞を表す。「～する」表記だがサ変動詞に限らない。

「～してしまう」「～れた」「～してくれる」「～しか～しない」「～してみたい」「～しやがって」「～しないでくれ」「～しない」「～すべきではないだろうか」「～して行きたい」

また、「普段から使っている」「観光客が押し寄せる」「騒ぎ散らす」のように特定の用法で評判表現になる例も存在した。

用言以外が評判表現 (のトリガ) になる例 (12 件) 用言以外の語にも評判表現になる、あるいはその語を伴うことではじめて評判表現になる、という語が存在する。具体的には「それ以上のもの」「宝もの」「人だかり」「バカ」「抜群」といった名詞、「ぜひ」「あっさり」「日常的に」といった副詞、「ありがとう」「ジーザス」といった感動詞、「超」といった接頭辞が確認された。「抜群」や「バカ」といった名詞に関しては形容動詞語幹として辞書に登録された方が良いと思われる。

評判表現が文の途中に出現する例 (11 件) 1 文に複数の節が含まれているなどの理由により、文の途中の用言が評判表現になっている場合がある。このような事例は係り受け解析や述語項構造解析を行い、その情報を用いることで扱うことができる。

形態素解析誤りに起因する誤り例 (7 件) うち 4 件はブログ記事の文体に起因するものであった (話し言葉, 方言, 小文字化, カタカナ化)。

形容動詞語幹の扱いに起因する誤り例 (3 件) 直後に「だ」や「です」を伴わない場合でも、形容表現として形容動詞語幹が用いられている場合がある。「～がお洒落。」のように文末に出現する場合が該当する。

意味類型付与誤りに起因する例 (1 件) 意味類型を付与する際に誤った意味類型が付与されてしまい、結果として誤った例が 1 件存在した。

表 3: KNB コーパス中の評判表現に含まれていた動詞「ある」の項

意味\項	名詞単体	名詞句
(属性を)有する	楽しみ, 問題, 雰囲気, 制限, 弱点, 開き, 魅力, おそれ, 価値, 歴史, 見所, ご利益, 種類, 威厳, 義務, 由来, メリット, 距離, 人気, チャンス, オチ, 使い道, 違和感, 文化, 割引, 安心感	例) 見直す必要, 得るもの, 参考になる点, QOL を高める要素, 興味深いもの, 引き込まれるもの, 歴史を感じさせるもの, 色々な機能
存在する	自然, 名所, 見所	例) いいところ, 有用なもの, 素敵なカフェ, お勧めスポット
(事象)が起こる	火事	例) 迷うこと, 無料になる場合, 不安になるとき
感情	興味, 文句, 思い入れ, 不満, 意見	例) 納得した覚え, そんなに使わなかった記憶, 面白いことになりそうという意識, 多いという印象

品詞情報のみでは正しかったが意味類型情報を扱うことによって本来は非評判文である文を評判文と判断し誤りとなった例は 712 文存在した。その中で最も大きい割合 (2 割程度) を占めていたのが, 文中の最後の用言が動詞「ある」の場合である。動詞「ある」は存在を表すので意味類型は常に形容であるが, 実際に評判表現になるか否かは項の名詞 (句) に依存する。例えば「レトロで綺麗なお庭があります。」は評判文であるが「~, その先に町家風の店内がある」は非評判文であった。表 3 に KNB コーパス中の評判表現に含まれていた動詞「ある」の項を示す。表中で「例」と書かれていない項目は, コーパス中で見つかった全ての項を挙げている。項が名詞単体の場合, 評判表現になるか否かは名詞自身の持つ極性によって決まると考える。「~がある」を伴った際の名詞の評価極性情報は日本語評価極性辞書 (名詞編)⁽⁴⁾ に含まれており, これを利用することで評判表現か否かを判別することが可能になると考えられる。実際に表 3 中の名詞単体のうち「由来」「使い道」「オチ」「割引」「意見」を除いた全ての名詞が日本語評価極性辞書 (名詞編) に含まれていた。項が名詞句の場合は主辞である名詞に依存する場合と, 修飾部に依存する場合に分けられる。項の主辞が「機能」「印象」などの特定の名詞の場合は修飾部に依らず評判表現になりやすいが「もの」や「ところ」あるいは「カフェ」のような単体では評判に結びつきにくい名詞が主辞の場合は評判表現か否かが修飾部に依存すると考えられる。

4 まとめ

本研究では, 先行研究において日本語の用言を意味に基づいて分類することを目的として提案された意味類型の有用性を評判文抽出タスクで確認した。その結果, 品詞情報のみを用いた場合と比較して評判文/非評判文の分類精度で 4.7 ポイント, F 値で 27.2 ポイントの上昇を確認した。また誤り解析を通じて意味類型

情報だけでは扱うことのできない評判表現は, パターンマッチや項の名詞の属性情報を利用することで扱えるようになることを確認した。

意味類型情報は, これまで「動詞」「形容詞」という表層的な情報を用いてきた手法において品詞情報の代わりに用いることにより, 用言の意味をより正確に反映させることが可能になる。

使用した言語資源およびツール

- (1) IPA 品詞体系日本語辞書。Ver. 2.7.0。http://mecab.sourceforge.net/src/
- (2) 橋本力, 黒橋禎夫, 河原大輔, 新里圭司, 永田昌明。構文・照応・評判情報つきブログコーパスの構築。自然言語処理 Volume 18, Number 2, pp. 175-201。2011。
- (3) 工藤拓。形態素解析器 MeCab。Ver. 0.996。http://mecab.sourceforge.net/
- (4) 東山昌彦, 乾健太郎, 松本裕治。述語の選択選好性に着目した名詞評価極性の獲得。言語処理学会第 14 回年次大会論文集, pp. 584-587。2008。

参考文献

- [1] 中山匠, 山本和英。用言の新しい意味類型 - 作用性用言と形状性用言 -。言語処理学会 第 17 回年次大会 発表論文集, pp. 560-563, 2011。
- [2] Shohei Okada and Kazuhide Yamamoto. Semantic Type Disambiguation for Japanese Verbs. In *Proceedings of the IALP 2014*, pp. 6-9, 2014.
- [3] Thomas Scholz and Stefan Conrad. Opinion Mining in Newspaper Articles by Entropy-Based Word Connections. In *Proceedings of the EMNLP 2013*, pp. 1828-1839, 2013.
- [4] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一。意見抽出のための評価表現の収集。自然言語処理, Vol. 12, No. 3, pp. 203-222, 2005.
- [5] Tetsuji Nakagawa, Kentaro Inui, and Sadao Kurohashi. Dependency Tree-based Sentiment Classification using CRFs with Hidden Variables. In *Human Language Technologies: NAACL 2010*, pp. 786-794, 2010.
- [6] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *Proceedings of the EMNLP 2013*, pp. 1631-1642, 2013.