

印象極性による商品レビューの視覚化

漆原 駿*

町田 翔*

延澤 志保*

*東京都市大学知識工学部

*東京都市大学大学院工学研究科

1 はじめに

商品レビューとは、消費者が商品に対する感想や評価を記載した文章であり、インターネットの Web 上に存在する商品販売サイトに記載されている。商品レビューの特徴として、消費者の生の声に分かることや、商品の購入を考えているユーザに対して有用な情報を含むことが挙げられる。しかし、商品レビューの件数は膨大で、すべてを読むには時間と労力が必要になる。そこで、効率よく評価に関する情報を獲得するため、商品レビューの文章から評価の要約や抽出に関する研究が行われている。

2 評判分析のための評価表現抽出

2.1 評価表現と印象極性

商品レビューの文章といったインターネット上の膨大なテキスト情報を扱うため、テキストから評価情報を抽出、整理、提示する評判分析の研究が行われている [1]。評価表現の認識では、例えば商品によって「重い」ことが良いことと評価されるものと悪いことと評価されるものがあるなど、対象とする商品分野によって同じ評価表現の評価が異なる。そのため、評価表現の認識の研究では、対象とする商品分野を絞りこむことで精度の向上を図ることが一般的に行われる。「重い」などの評価表現が良いことか悪いことかを示す印象を印象極性と呼ぶ。例えば「重くて困った」というレビューでは「重い」の印象極性はネガティブと判断できる。

2.2 評価表現の自動抽出

小林らは意見抽出のための評価表現の収集を行っている [2]。小林らは、意見のある対象もしくは対象のある側面に対する記述者の主観的な評価を表す記述として、意見を < 対象, 属性, 評価表現 > の要素からなる組を、商品の評価情報として抽出する手法を提案している [3]。河野らは、評価表現と属性に出現する「使いやすい」、「問題ない」といった複数の形態素からなる表現に対して接続処理を行い、1つの複合語として扱う手法を提案している。河野らの手法では、例えば「使いやすい」という語句は「使う(動詞)」と「やすい(形容詞)」

の複合語「使いやすい(形容詞)」として扱う。河野らは、商品分野を特定せずに評価表現を抽出するため、評価表現の印象極性にはこだわらず、評価表現の自動認識に焦点を当てている [3]。

2.3 印象極性の判定

稲益は文脈一貫性や否定表現を利用し、化粧品のレビューを対象とした印象極性の判定を行っている [4]。文脈一貫性とは、ある対象に関して評価を述べる内容を記述する際には、好評もしくは不評の意見を列挙することが多く、好不評の極性を反転させる際には「ですが」、「しかし」といった逆説表現で明示することが多いという文脈の特性のことである。稲益は「良い」など印象極性が明確な評価表現に着目し、こういった評価表現との構文関係に基づいて他の評価表現の印象極性を推定することで、対象商品を絞らず文脈で評判分析を行う手法を提案した [4]。ただ、稲益の研究は化粧品を対象としており、対象分野を特定しない場合についての考察は行っていない。

3 商品レビューの印象極性推定手法

3.1 提案手法の概要

本研究では、商品の評価を一目で判断できるようにするため、商品レビューの評価情報に印象極性を示す色を付け表示する手法を提案する。本手法では、商品レビューに含まれている評価情報を抽出し、これに対して文脈を用いて印象極性を求める。対象分野に関わらず適切な印象極性を推定することが本研究の主眼である。

本研究のフローチャートを図 1 に表す。本研究の入力は、ある 1つの対象に対する複数の商品レビュー文章とする。出力は、商品レビュー文章の評価情報に印象極性を示す色を付けハイライト表示したものとする。印象極

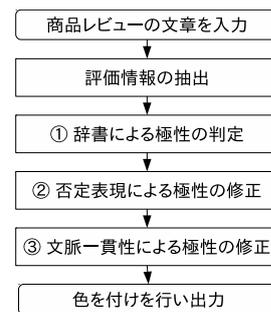
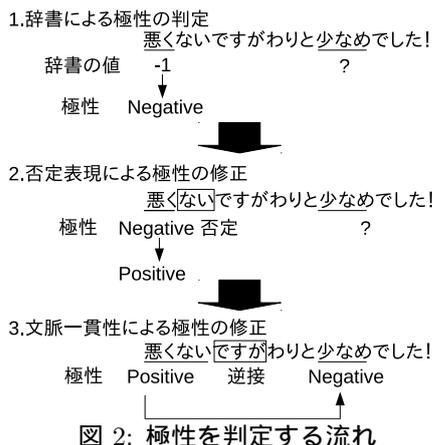


図 1: 本手法の流れ

性の推定は、3段階で行う。処理の具体例を図 2 に示す。



3.2 辞書による印象極性の判定

第1フェーズでは、高村らの単語感情極性対応表 [5] を利用して、文中の各評価表現について基本とする印象極性を得る。この辞書はさまざまな語に対して印象極性を -1 (ネガティブ) から 1 (ポジティブ) までの実数値で示したものである。本研究では $-1 \leq x < 0$ をネガティブ、 $0 < x \leq 1$ をポジティブとし、印象極性値が 0 の場合と辞書に記載がない場合とを印象極性不明とする。

図2の第1フェーズでは、与文中で認識された「悪く」と「少なめ」の2種類の評価表現に対して、単語感情極性対応表 [5] の記載から「悪く」はネガティブ、「少なめ」は印象極性不明としている。単語感情極性対応表の辞書引きには、与文を形態素解析した結果から得た各語の原形を利用する。

3.3 否定表現による印象極性の修正

第2フェーズでは、文中の否定表現を検出し、必要に応じて印象極性の修正を行う。図2では「悪く」の直後に「ない」があるため、この文では「悪く」は否定されているものと推定できる。そのため、この2語を「悪くない」とひとまとめにするとともに、「悪く」に与えられた印象極性を反転させ、この複合語の印象極性をポジティブに修正する。

3.3.1 文脈を考慮した極性の判定

第3フェーズでは文脈一貫性を用いて極性の修正を行う。これにより、辞書に載っていない単語や誤った判定をしている箇所の修正を行う。図2には「が」という逆接表現が含まれている。図2ではこの逆接表現の直前にポジティブな評価表現があるため、文脈一貫性により、この逆接表現の直後の評価表現はネガティブの可能性が高いと推定できる。ここでは逆接表現の直後の評価表現「少なめ」が印象極性不明となっていたため、これをネガティブに修正する。

同様に、順接表現の場合には、その前後の印象極性は一致すると仮定して修正を行う。

4 実験結果と考察

本手法を用いて、1つの商品に対する701個のレビュー中、逆接表現を含む91個のレビューに対して印象極性の推定実験を行った。人手で作成した正解データと比較した結果を図3に示す。

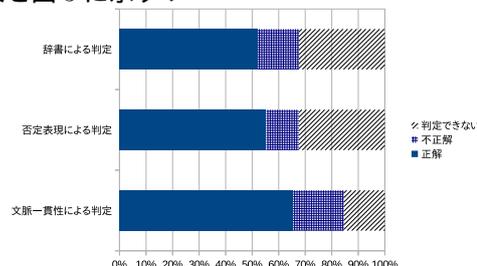


図 3: 実験結果

図3から辞書による判定から、さらに文脈を考慮して判定を行うことで正解と判定できないものの確率が向上していることがわかる。特に文脈を考慮した判定によって、辞書による判定ができなかったものが多く判定できるようになっている。

しかし、不正解が増えたのは、逆接表現や文脈を考慮して判定する際に辞書による判定で誤っている15.5%によって不正解が増えてしまったと考えられる。例を挙げると、「便利」がネガティブと判定されている。また、判定できないものが残っているというのは、1つのレビュー内に辞書による判定で評価表現の極性を1つも判定できていない場合があるため逆接表現、文脈を考慮して判定を行わず、判定できないものがあると考えられる。

5 まとめ

本研究では、文脈を考慮し評価表現の持つ印象極性を求めることで、より精度の高い推定を行うことができた。

参考文献

- 乾 孝司, 奥村 学, “テキストを対象とした評価情報の分析に関する研究動向”, 自然言語処理 Vol.13, No.3, pp.201-241, 2006.
- 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.
- 河野 一志, 町田 翔, 村松 拓実, 小林 美穂, 延澤 志保, “不特定分野の商品レビューを対象とした評価情報の自動認識”, 情報処理学会 第79回全国大会, Vol.2, pp.527-528, 2017.
- 稲益 ほなみ, “評価視点に着目したクチコミサイトの評価の視覚化”, 平成24年度 東京都市大学卒業論文, 2013.
- 高村 大哉, 乾 孝司, 奥村 学, “スピンモデルによる単語の感情極性抽出”, 情報処理学会論文誌, Vol.47, No.2, pp.627-637, 2006.