

意見極性の系列に基づくレビュー記事の評価値推定手法の提案

堀内 雅人[†] 佐藤 哲司^{††}[†] 筑波大学情報学群知識情報・図書館学類 〒 305-8550 茨城県つくば市春日 1-2^{††} 筑波大学大学院図書館情報メディア研究科 〒 305-8550 茨城県つくば市春日 1-2E-mail: [†]s0911649@u.tsukuba.ac.jp, ^{††}satoh@ce.slis.tsukuba.ac.jp

あらまし レビューからの情報獲得において、レビューのすべてを閲覧することは時間的負担のかかる作業である。このため、レビューを閲覧するかどうかを、評価値を見て決定する場合が少なくない。しかし、この評価値をレビューが客観的に付与しているとは限らない。本研究では、この問題の解決を目的に、レビュー記事の内容からレビューが取るであろう評価値を推定する手法を提案する。具体的には、評価表現辞書の拡張、拡張後の辞書による文単位での極性付与、極性の系列による評価値算出を行った。提案法により、評価表現辞書に含まれる表現数を15倍～30倍に拡張し、文への極性付与においては、拡張前の辞書と比較して付与数で約3倍、付与精度で10.3ポイント向上がみられた。評価値推定においては、文単位の極性の系列を考慮することで、系列を考慮しない場合に比べて最大で26ポイント推定精度が向上した。

キーワード レビュー, 評価表現, 文書分類

Masahito HORIUCHI[†] and Tetsuji SATOH^{††}[†] College of Knowledge and Library Sciences, School of Informatics, University of Tsukuba

1-2 Kasuga, Tsukuba, Ibaraki, 305-8550 Japan

^{††} Graduate School of Library, Information and Media Studies, University of Tsukuba

1-2 Kasuga, Tsukuba, Ibaraki, 305-8550 Japan

E-mail: [†]s0911649@u.tsukuba.ac.jp, ^{††}satoh@ce.slis.tsukuba.ac.jp

Abstract In getting information from reviews, browsing all of the reviews is a time consuming task. That is why, it is not uncommon to see the evaluation values to decide whether view reviews or not. However, the evaluation values reviewers give is not always objectively. In this study, we propose the method to estimate the evaluation value of the review, expand the number of evaluative expressions, give a polarity, for example, positive or negative or neutral to sentence, and calculate the evaluation values of the reviews by using the sequence of the polarity. By the proposed method, we increase the number of evaluative expressions by 15~30, increase the number of the giving polarity to sentence 3 times, improve the precision of giving polarity to sentence 10.3 points. By considering the sequence of the polarity, we improve the precision of estimating evaluation values of the reviews up to 26 points in comparison with not considering it.

Key words Reputation, Evaluative expression, Document classification

1. はじめに

近年、レビューサイトやSNSなどで個人が発信する、特定の商品やサービスに関する評判情報が増加し続けている。人々が商品の購入を検討する際には、これらの他者が発信する評判情報を参考にすることが一般的になってきている。例えば、

「Amazon.co.jp」^(注1)、「価格.com」^(注2)、「楽天トラベル」^(注3)などの、オンライン商品購入サイトでは、商品ごとに多くのレビューが集まっており、各レビューが参考になったかどうかといった投票も盛んに行われている。これらのサイトにおいて、レビューは当該商品そのものの情報を共有するだけでなく、

(注1) : Amazon.co.jp, <http://amazon.co.jp/>

(注2) : 価格.com, <http://kakaku.com/>

(注3) : 楽天トラベル, <http://travel.rakuten.co.jp/>

商品の新しい使い方やメンテナンス方法、他の商品との組み合わせ方など、未知な情報や価値観に接触する機会を提供する有用な情報になり得る。しかし、レビューが増加するに伴い、有用な情報に到達するまでに閲覧するユーザの時間的負担が増加する。

商品のレビューには文章（以下、レビュー記事）の他に、レビュアーによる評価値が付与されていることがある。評価値は、ユーザが商品への評価を概観したり、閲覧するレビューを選択する際に役立つ。しかし、レビュアーによる評価値の付与基準は曖昧であり、同じような内容のレビュー記事に異なる評価値が付与されるなど、レビュー記事と評価値が不一致であることも多い。レビュアーが付与した評価値が曖昧であることで、ユーザが混乱することも少なくない。レビューを選択することが困難な場合、ユーザは既存のランキングで上位に現れているレビューを集中して閲覧すると考えられる。そのため、ランキング上位のレビューにさらに「参考になった」などの投票がなされる可能性が高く、ユーザが獲得する情報は偏ってしまう。様々な評価値のレビューを閲覧することで、獲得する情報が偏ることを避ける要求がある。しかし、大量のレビューが評価値ごとに集まっても、評価値の付与基準が曖昧であり、閲覧しているレビューが、レビュー集合中でどこに位置しているのが把握しづらくなっている。統一的な評価値で位置づけることで、レビューの肯定、否定の程度による閲覧を可能にし、ユーザの負担を軽減することが期待できる。

本研究では、商品のレビューを評価値によってレビュー集合の中で位置づけ、効率よく閲覧するための手法を提案する。ユーザがレビューを閲覧する際に、そのレビューが肯定的なのか否定的なのかは重要である。レビューの評価値を統一的に自動付与することで、レビューの概観把握を容易とし、情報をより有効に活用することができる。例えば、評価値が高いレビューがほとんどである商品は、良い商品であると推測できるが、数が少なくとも評価値の低いレビューを閲覧することで、良いとされている商品の評価の低い部分の情報を獲得することが期待できる。また、実際のレビュー記事では、同じ肯定的な文でも、レビュー記事中の肯定的な文の数や直前の文が肯定的、否定的どちらの極性であるかによって肯定の強さに差があると考えられ、極性の系列を考慮した評価値推定を行うことでより正確な推定が期待できる。

本研究におけるレビュー記事に対する評価値の推定手法の手順は、まず、各レビュー記事を文単位で分割する。次に、文単位で肯定または否定の極性を付与することで、レビュー記事を文の極性が系列を成している極性の集合として捉える。そして、意見極性の系列、特に連続や反転に着目して対象物への評価値を推定する。本手法により、レビュー内容とレビュアーが付与した評価値が一致しないレビューや、評価値のついていない評判情報に対して、レビュアーが付与した評価値とは独立かつ客観的に評価値を付与することで、レビューや評判情報の当該商品のレビュー集合中における位置づけを得ることができる。

ユーザは自らの考えに即した情報は受け容れやすく、そうでない情報は受け容れづらいことが知られている。そのため、閲

覧中の情報がレビュー集合全体の中でどこに位置しているかを確認することで、ユーザが目を向けていなかった獲得する機会が生まれる可能性がある。

本研究の構成は以下の通りである。第2章で関連研究について述べ、本研究の位置づけを示す。第3章で本研究での提案手法について述べ、第4章で評価実験の概要、結果および考察を述べる。第5章で本論文を総括する。

2. 関連研究

評判情報を扱う研究は、テキスト評価分析の要素技術に関する研究と、テキスト評価分析の応用研究に大別される[1]。前者は、評価表現辞書の構築、評判情報を観点とした文書分類、評判情報を含む文の抽出、評判情報の要素組の抽出に関する研究に分類される。後者は、分類、抽出した評判情報を要約、または可視化して提示するための手法に関する研究である。以下では、本研究に特に関連のある、評価表現辞書の構築に関する研究と、評判情報を観点とした文書分類に関する研究について述べる。

評価表現辞書の構築、特に評判情報を語句単位で肯定/否定に自動分類する研究について、Kampsら[2]やHuら[3]は、WordNetの類義関係や反義関係などの情報を利用して形容詞の評価極性を判定する手法を提案している。Turney[4]は、コーパスから得られる共起情報から語句の評価極性値を判定している。Huら[3]、藤村ら[5]、立石ら[6]は「広い」、「大きい」、「良い」、「満足した」などのように、商品の特徴や評価の視点、個人の主観的な感情を表す表現を評価表現とし、評価表現に着目する手法を提案している。小林ら[7]は用言の評価表現の獲得手法を、東山ら[8]は名詞の評価極性の獲得手法を提案しており、これらの成果に基づいて、乾らは日本語評価極性辞書（用言編、名詞編）^(注4)を作成している。また、情報通信研究機構(NICT)では、評価表現に対する評価極性の判定を組み込んだ、意見（評価表現）抽出ツール^(注5)を公開している。

評判情報を観点とした文書分類とは、評判情報を文書単位で肯定極性、否定極性のいずれの極性を持つかを判定する課題である。ここでは広義に、評判情報の文書、評価点や参考になったかなどの情報により、文書を何らかの指標で分類する研究について述べる。Turney[4]は、文書に含まれる評価表現を評価値極性値付きで抽出し、極性値の平均値を求め、平均極性値の符号によって文書全体の評価極性を決定する手法を提案している。Taboadaら[9]は、評価表現が出現する位置に応じて、人手で定めた重みを評価極性値に乘じることにより評価極性値を修正する手法を提案し、文書の後半2/3の位置に現れる評価表現への重みを最も高くした場合にもっとも良い分類精度が得られたと報告している。江崎ら[10]は、日本語のWeblog記事を対象にして、形容詞の出現位置を考慮した同様の手法を提案している。ここでの結果によると、Weblog記事の前方側に現れ

(注4)：日本語評価極性辞書、<http://cl.naist.jp/inui/research/EM/sentiment-lexicon.html>

(注5)：意見（評価表現）抽出ツール、<http://alaginrc.nict.go.jp/opinion/>

る評価表現への重みを最も高くした場合に最も良い分類精度が得られたと報告している。Pangら[11]は、教師あり機械学習に基づく分類手法が、評価文書分類にも有効であるかどうかを実験的に検証し、形容詞以外の単語も、評価文書の評価極性を判定する際によい指標となっていることを示している。Pangら[12]は、評価文書間の肯定極性類似度を提案し、人手で付与した3値と4値の評価点と高い相関を持つとしている。清水ら[13]は、レビューの中には記事内容がほとんど同じであるのに評価点が異なっているものが存在すると仮定し、ゲームソフトのレビューの評判を主成分分析とクラスター分析を用いた手法で把握しようとしている。小倉ら[14]は、レビューがユーザから得た評価を元に支持率と参考度を算出し、レビューの品質を考慮したスコアリング、およびランキングの手法を提案している。

本研究では、文ごとに付与した極性の系列を利用して、文書の評価値を推定する手法を提案する。評判情報を扱う研究としては、評価表現辞書を利用した、評判情報を観点とした文書分類に関する研究である。本研究の前提として、評判情報の文単位での系列情報を用いるため、乾らの作成した日本語評価極性辞書、およびNICTが提供している日本語WordNet^(注6)を用いて、レビュー記事の各文の評価極性を判定した。

Turney[4]の手法に端を発する評価表現の比率を用いて評価値を推定する手法が知られている。本研究は、評価値を推定する手法として、文書中の極性を持つ形態素の出現位置に重みを持たせる点でTaboadaら、江崎らの手法と共通している。しかし、複数の品詞を用いて極性を付与する点と、2値から4値までに限定されずに評価値を付与する点で異なる。レビュー記事の内容と評価点のずれ等の理由で、レビューの概観把握が困難であることに対して、清水ら、小倉らと同様に、現状と異なる評価軸を与えて評判情報を把握しようとする手法である。評価値付与において、評価点と肯定/否定の系列に着目している点があり、ブログなどの商品紹介や評価点のついていないレビューに対しても容易に評価値を付与できる特徴を持つ。

3. 提案法

本章では、評価値推定を行うための手法について述べる。3.1節では提案手法の概要、および全体の処理を概観する。3.2節では文に対して評価極性を付与する手法について、3.3節では付与した評価極性を用いて文書への評価値を推定する手法について述べる。

3.1 提案手法の概要

評価値推定の手法を提案するにあたり、最初に用語を定義する。

評判情報 レビューやブログ、SNS等で書かれた、商品やサービスの情報

評価対象 評判情報の対象となっている特定の商品やサービス

評価極性 評判情報中に出現する単語の

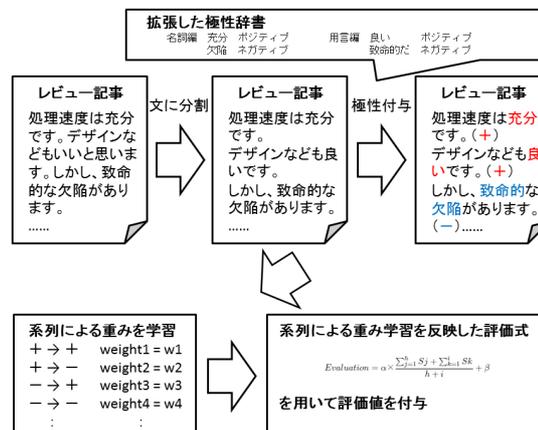


図1 提案手法の概要図

ポジティブ/ネガティブ/ニュートラル
極性系列 評価極性の並び

評価値 レビューに対して付与された点数

本研究では、レビュー記事の文ごとに付与した極性の系列を利用して、レビュー記事の評価値を推定する。ここでの評価値の推定は、文書分類において、評価値による多値分類に該当する。単純にレビュー記事の分類を行う場合、レビュー記事に含まれる単語に極性および極性の強さを表すスコアを付与し、レビュー記事全体を1文として扱い、評価値を付与する手法が考えられる。また、文ごとに極性やスコアを付与する場合でも、単純には、文中にポジティブとネガティブ、いずれの極性を持つ単語が多いかによって付与することができる。これらの手法で極性やスコアを付与する際には、文のスコアは文中の極性を持つ単語に付与されたスコア、および極性を持つ単語の数に依存する。具体的には、単語に付与されたスコアや単語の数による評価値付与は、文ごとの意見の強さを単語のみで判断している。人が書いたレビュー記事は、文ごとに意見の強さが異なる可能性があり、単語の極性のみでは、レビュー記事に評価値が上手く付与できていない可能性がある。

本研究では評判情報を扱う粒度を文とし、文ごとに極性を付与することにより、レビュー記事を極性の系列と捉える。そして、極性が連続もしくは反転した際に、連続も反転もしていない場合と比べ、文の重みが増えていると仮定する。

レビュー記事を文単位に分割するため、次の手順で文を分割した。

1. 文中にて、「。」「.」「!」「!」「?」「?」「♪」が1つ以上、または改行が2つ以上連続した場合、そこで分割する
2. 分割した単位の中で、文末を意味しない「,」「,」「,」に接続して現れる改行を除去する
3. 分割した単位内に、改行が3つ以上ある場合は、全ての改行位置で分割する

この手順で分割された単位を文として扱う。手順3を行った理由は、評判情報は統一の表記法で表記されていないため、

(注6) : 日本語 WordNet, <http://cl.naist.jp/inui/research/EM/sentiment-lexicon.html>

文末に「。」「.」などの記号を用いずに、改行を用いることにより文の区切りを表しているレビューが多く見られたためである。

提案手法の概要を図1に示す。入力は、評価対象のレビュー記事である。与えられたレビュー記事に評価表現辞書により極性を付与する。付与された極性を用いて、評価値を算出する。正解データとして用意した評価値と算出した評価値を比較することで、系列による重みを学習し、評価式に反映する。最後に、極性系列の重みを反映した評価式を用いて評価値を付与する。提案手法は、評価表現辞書による評価極性の付与、およびレビュー記事の評価値推定に大別される。以下では、この順で提案手法の詳細について述べる。

3.2 評価表現辞書による評価極性の付与

3.1節で分割した文に対して、乾らの日本語評価極性辞書を拡張し、拡張語の辞書を用いて評価極性を付与する。以下に文に評価極性を付与する手順を示す。

1. 文に対して形態素解析を行い、形態素に分割する
2. 分割した形態素と、評価表現辞書の表現を正規表現でマッチングさせる
3. マッチした場合、辞書の表現が持つ極性を分割した形態素に付与する
4. 文中の複数の形態素に極性が付与された場合、以下の式で文の極性を判定する

$$Polarity_{sentence} = N_{positive} - N_{negative} \quad (1)$$

$Polarity_{sentence}$ は文の極性、 $N_{positive}$ はポジティブな極性が付与された形態素の数、 $N_{negative}$ はネガティブな極性が付与された形態素の数である。 $Polarity_{sentence} > 0$ ならば文の極性はポジティブ (p)、 $Polarity_{sentence} < 0$ ならば文の極性はネガティブ (n)、 $Polarity_{sentence} = 0$ ならば文の極性はニュートラル (e) と判定する。

本研究では、評価極性の付与にあたって、乾らの日本語評価極性辞書を拡張して使用している。本節では、評価極性の付与に用いる評価表現辞書と、評価表現辞書の拡張手法について述べる。

3.2.1 評価表現辞書

評価表現辞書とは、評価を表すために用いられる可能性のある表現を収集した辞書を指す。評価表現辞書の構築に関しては、評価表現辞書に収集した表現に、様々な手法で評価極性を付与したものが存在する。近年では、語句単位で肯定/否定の2値に分類して収集する研究が盛んに行われている。特に、Turneyらに代表される、大量コーパスから語句の評価極性を判定する手法など、機械的に極性を付与した評価表現辞書の自動作成に関する研究は多い。

3.2.2 日本語評価極性辞書

日本語の評価表現辞書の作成において、小林らが用言の評価表現の獲得手法を、東山らは名詞の評価極性の獲得手法を提案しており、これらの成果に基づき、乾らが日本語評価極性辞書

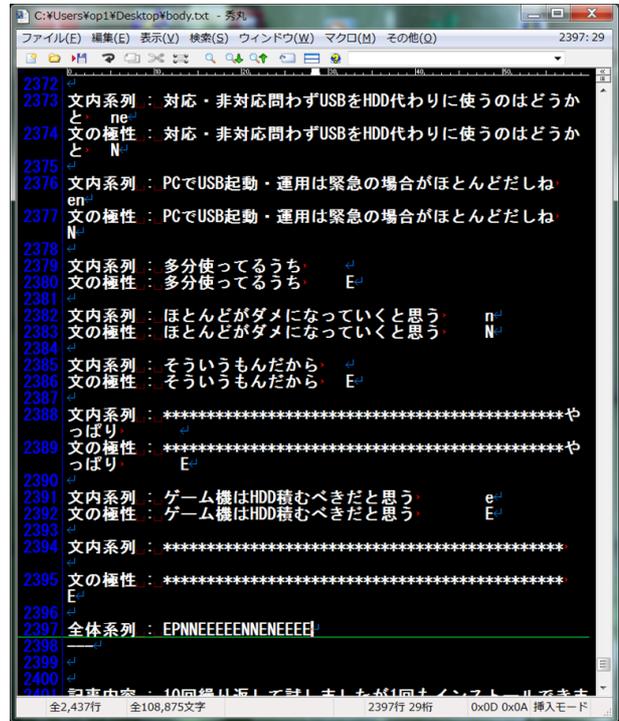


図2 評価極性の付与例

(用言編, 名詞編)を公開している。乾らの日本語評価極性辞書は、自動収集した評価表現に対して、ルールを設けて人手で評価極性を付与している。人手で評価極性を付与することは、人間の感性に沿って評価極性を精度良く付与できる利点があるが、付与に大きなコストがかかる。そのため、評価表現辞書としては小規模なものになり、評価表現の網羅性が低い。本研究では、日本語 WordNet における類義語、対義語を用いて、乾らの日本語評価極性辞書を拡張している。そして、拡張前後の辞書を用いてレビュー記事の文単位での極性付与を行い、拡張前後で文単位での極性付与の精度が上昇することを確認した。具体的には、以下の手順で辞書の拡張を行った。

1. 日本語評価極性辞書の各表現を日本語 WordNet で検索し、類義語を辞書に追加する
2. 類義語を追加した辞書から、以下の語を除外する
 - 重複する語
 - 拡張前の辞書にある品詞以外の語
 - ポジティブとネガティブ両方に出現する語
3. 追加した類義語に、元の表現の極性を付与する

辞書の拡張により、評価極性の付与コストを大きく上げることなく、評価表現の網羅性を向上することができる。

表 1 日本語評価極性辞書（用言編）の内容一例

極性（経験 or 評価）	極性語
ネガ（経験）	壊れる
ネガ（経験）	不満がある
ネガ（経験）	もたつく
ネガ（評価）	危ない
ネガ（評価）	今ひとつ
ネガ（評価）	使いにくい
ポジ（経験）	気持ちいい
ポジ（経験）	楽しい
ポジ（経験）	捗る
ポジ（評価）	お買い得だ
ポジ（評価）	お気に入りです
ポジ（評価）	ほほえましい



図 3 日本語 WordNet, Web 上での検索例

表 2 拡張後の日本語評価極性辞書（用言編）の内容一例

極性	極性語	備考
ネガ	壊れる	
ネガ	いかれる	拡張語
ネガ	打ちこわれる	拡張語
ネガ	潰れる	拡張語
ネガ	割れる	拡張語
ネガ	崩れる	拡張語
ネガ	……	……
ネガ	不満がある	名詞辞書の「不満」が拡張される
ネガ	もたつく	拡張語なし
ネガ	危ない	
ネガ	危い	拡張語
ネガ	危うい	拡張語
ネガ	険しい	拡張語
ネガ	峻しい	拡張語
ネガ	おぼつかかなげ	拡張語
ネガ	疑わしい	拡張語
ネガ	今ひとつ	拡張語なし
ネガ	使いにくい	拡張語なし
ポジ	気持ちいい	拡張語なし
ポジ	楽しい	
……	……	……
……	……	……

3.3 レビュー記事の評価値推定

3.2 節で付与した文の極性を用いて、レビュー記事に評価値を付与する。レビュー記事において、文の極性が連続、あるいは反転した際は、文の意見としての極性が弱まっている、もしくは強まっている傾向があると考えられる。そこで、文の極性

が連続あるいは反転した際の重みを考慮した評価値の推定式を以下のように定義する。

$$Evaluation = \alpha \times \frac{\sum_{j=1}^h S_j + \sum_{k=1}^i S_k}{h + i} + \beta \quad (2)$$

$Evaluation$ は評価値、 S_j はポジティブな極性が付与された文のスコア、 S_k はネガティブな極性が付与された文のスコア、 h はポジティブな極性が付与された文の数、 i はネガティブな極性が付与された文の数である。ただし、

$$S_j = \begin{cases} Wnp & (\text{直前の文の極性がネガティブ}) \\ Wpp & (\text{直前の文の極性がポジティブ}) \\ Wep & (\text{直前の文の極性がニュートラル}) \end{cases}$$

$$S_k = \begin{cases} Wnn & (\text{直前の文の極性がネガティブ}) \\ Wpn & (\text{直前の文の極性がポジティブ}) \\ Wen & (\text{直前の文の極性がニュートラル}) \end{cases}$$

であり、 Wnp , Wpp , Wep , Wnn , Wpn , Wen はそれぞれの場合の重みを考慮した文のスコアである。本研究では $\alpha = 2$, $\beta = 3$ とし、文のスコアは 0.1~2.0 の間で変動するものとする。 $Evaluation$ の計算結果が 1 未満となった時は 1, 5 より大きくなった時は 5 として扱う。また、レビュー記事の先頭の極性も系列と捉えるために、極性系列の先頭にニュートラルの極性を挿入している。

4. 評価実験

4.1 実験の概要

3 章で提案した、拡張した評価表現辞書による評価極性付与と、レビュー記事の評価値推定を実装し、評価実験を行うことで提案手法の有効性を確認する。以下、4.2 節では、拡張前後の評価表現辞書による評価極性付与の精度、極性付与数の比較について述べる。実験用データについて説明した後、評価極性付与結果の比較を行い、考察を述べる。4.3 節では、付与した評価極性の系列を用いて、レビュー記事に評価値を付与する。実験方法について説明した後、評価値推定の評価、考察を述べる。

本実験では、Amazon.co.jp の商品ページのレビューを利用する。Amazon.co.jp のレビューを利用する理由は、評価項目などに分かれていない状態の、平文に近いレビュー記事が取得できるためである。Amazon.co.jp のレビューを図 4 に示す。図のように、参考になったかどうかの投票数、評価値、レビュータイトル、日付、レビュアー名、レビュー記事などが記入されている。

4.2 評価表現辞書の拡張および文への評価極性付与の評価

Amazon.co.jp の様々なカテゴリ、商品から、各々 100 件のレビューを取得し、拡張前後の辞書を用いて、1 文中において抽出されるポジティブな評価表現 1 つにつき +1 を与え、ネガティブな評価表現 1 つにつき -1 を与え、入力文中の全ての評価表現の値の和が正、負、ゼロのいずれになるかで文単位の極性を判断した。ただし、この場合、1 つ 1 つの評価表現への極性付与が必ずしも正しく行われていなくても、文全体の極性の判定は正しく行うことができている場合がある。

10人中、10人の方が、「このレビューが参考になった」と投票しています。

★★★★ 美味しく炊けます。2011/4/10

By Ipse Dixit - レビューをすべて見る

レビュー対象商品: ZOJIRUSHI マイコン炊飯ジャー 飯の炊き5.5合 NS-WB10-CA ベージュ (ホーム&キッチン)

購入する前、いくつかのレビューを読んで、多くのネガティブなコメントがあり、買おうか買うまいか悩みました。あまり高いものを買うつもりもなかったため、他にあまり選択肢もなく、ZOJIRUSHIというブランドもあり、買ってみました。

実際の炊きあがり具合はともよく、美味しく炊きあがります。また、白米、七分づき米、玄米の3種類しか炊いていませんが、どれもとても美味しくできました。

購入して1週間なので、そこまでの情報ですが、6,000円弱でとても良い買い物をさせていただきました。

本当におすすです。

図4 Amazon.co.jpのレビュー

表3 拡張前後の評価表現辞書の表現数

辞書	拡張前	拡張後
名詞編	12,214語	182,445語
用言編	5,280語	158,142語

表4 拡張前後の評価表現辞書による文への極性付与数、極性付与精度

極性付与数	拡張前	拡張後
	44,025	132,608
極性付与精度	拡張前	拡張後
	63.4%	73.7%

本実験に際して、家電、炊飯器、冷蔵庫、PSP（ゲームカテゴリ内でPSPで検索）、お菓子の各カテゴリ、各商品からレビューを各100件、計500件取得し、3名の実験協力者が文ごとにポジティブ、ネガティブ、ニュートラルの極性を付与した。本実験では、実験協力者が3名ともポジティブ、またはネガティブを付与した文について、拡張語の評価表現辞書により付与した文の極性と、実験協力者が付与した文の極性を比較し、極性を正しく付与できたかを評価する。

4.2.1 評価極性付与結果の比較

まず、評価表現辞書を3.2節で述べた手法で1回拡張した前後の辞書の表現数を示す。表3に示す通り、名詞編では約15倍、用言編では約30倍の表現に極性を与えることができた。

次に、拡張前後の評価表現辞書を用いて、文ごとに評価極性を付与した際の、極性付与数について述べる。Amazon.co.jpから取得した500件のレビューを、3.1節の手法で文に分割した結果の281,249文に、名詞編、用言編の両方の拡張後辞書を用いて、文ごとに3.2節の手法で極性を付与した。この時の辞書拡張前後での極性付与数を比較する。表4に示す通り、拡張前と比較して、拡張後には約3倍の文に極性を付与することができた。続いて、文ごとに手動で付与した極性と、辞書を用いて付与した極性が一致しているかを辞書の拡張前後で比較する。この際、極性を持つ表現が文中に出てこない場合は、その文を母数から除いている。表4に示す通り、拡張前と比べて、拡張後の方が文ごとの極性付与精度は10.3ポイント向上した。

4.2.2 評価極性付与についての考察

極性辞書を拡張することで、文単位での評価極性付与数、評価極性付与精度がともに上昇することを示した。文単位での評価極性付与精度が上昇した理由として、辞書拡張前には、少な

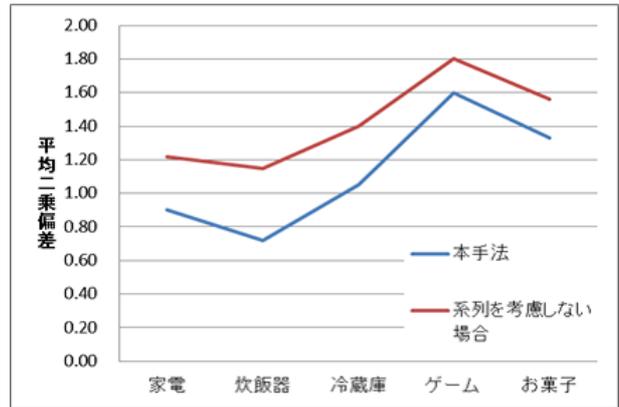


図5 レビュー評価値との平均二乗偏差

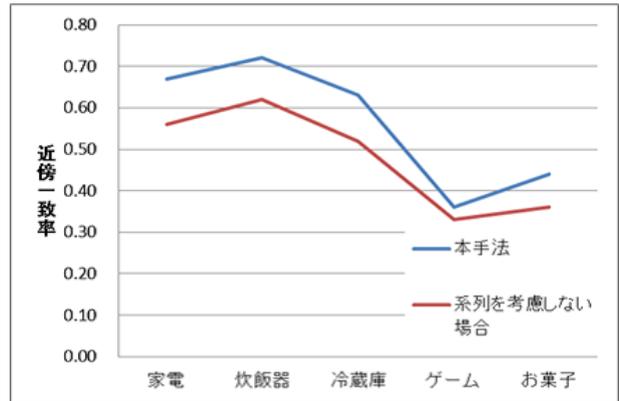


図6 レビュー評価値との近傍一致率

い表現に付いた極性で1文の極性を判定するため、極性が付く表現が1文に1つしかないときに、文の極性を誤って付与することが多いことが挙げられる。辞書拡張後には、複数の極性で1文の極性を決定することが増え、正しい評価極性を付与できるようになったと考えられる。

4.3 評価値推定の評価

4.3.1 レビュー評価値との比較

評価極性を用いて文書への評価値を推定する手法の実験方法について説明する。本手法の評価として、まず本手法で推定した評価値とレビュー評価値の比較を行った。レビュー評価値とは、レビューが付与している評価値である。評価値の推定精度と、推定精度が最も良くなったときの極性の連続、反転に与えられたパラメータを算出した。すなわち、レビューが付与した評価値を推定できたかを評価した。

また、カテゴリや商品ごとに、推定式の重みを0.1~2.0の間で変動させ、最適な重みがカテゴリや商品ごとに異なるかについて検証する。評価するレビューは、家電、炊飯器、冷蔵庫、PSP（ゲームカテゴリ内でPSPで検索）、お菓子の各カテゴリから、売れている順にカテゴリ、商品のレビューを上位各200件、計1000件取得し、カテゴリごとに上位100件と下位100件ごとに分けて、2つのグループを作成したものである。

評価の指標には、平均二乗偏差と近傍一致率を用いた。

表 5 カテゴリ、商品ごとの平均二乗偏差が最小となったパラメータ

カテゴリ, 商品	Wnp	Wpp	Wep	Wnn	Wpn	Wen
ランキング上位 100 件						
家電	0.6	0.5	0.6	0.4	0.6	0.6
炊飯器	0.6	0.5	0.6	0.4	0.5	0.6
冷蔵庫	0.5	0.6	0.6	0.6	0.7	0.6
PSP	0.9	1.1	0.8	1.1	0.7	0.8
お菓子	1.0	0.7	0.6	0.6	0.8	0.6
ランキング下位 100 件						
家電	0.7	0.5	0.5	0.4	0.5	0.6
炊飯器	0.5	0.5	0.6	0.4	0.6	0.6
冷蔵庫	0.7	0.5	0.7	0.6	0.6	0.6
PSP	0.8	0.9	0.9	0.9	0.7	0.7
お菓子	0.6	0.6	0.5	0.5	0.7	0.6

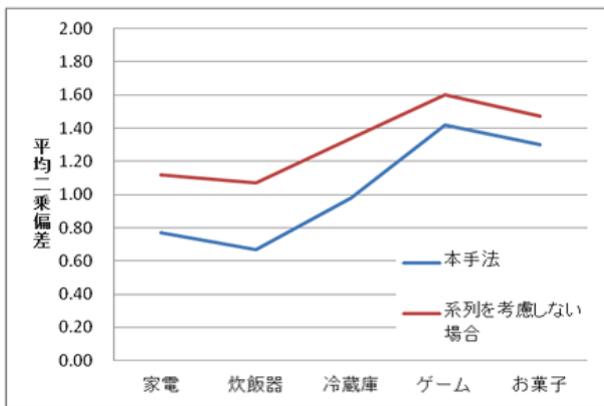


図 7 実験協力者評価値との平均二乗偏差

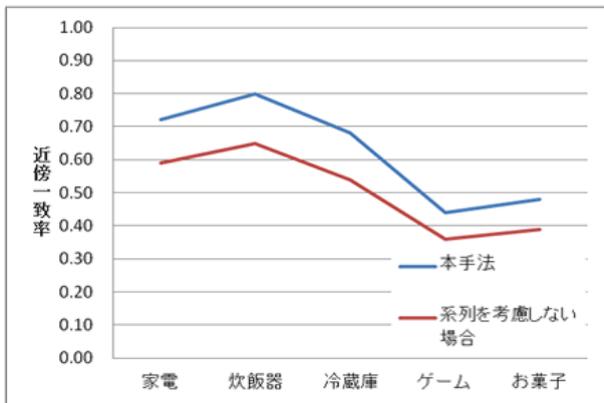


図 8 実験協力者評価値との近傍一致率

平均二乗偏差は、真の値からの誤差を表すもので、 N 個の数 $\{x_1, x_2, \dots, x_N\}$ について、

$$x_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (3)$$

と定義される。ここで、 x_i は 1 レビューごとの、本手法で推定した評価値とレビューア評価値との差の絶対値である。

近傍一致率は、推定評価値と正解となるレビューア評価値の差が 1.0 未満となる評価値を推定できた割合を示す。

4.3.2 実験協力者評価値との比較

次に、本手法で推定したパラメータを用いた評価値と、実験協力者がレビューアの文を読み付与した評価値（実験協力者評価値）との比較を示す。評価の指標はレビューア評価値との比較の際と同じである。

実験協力者評価値は、3 名の実験協力者がカテゴリ、商品の 300 件のレビュー記事を読み、レビューに付与されているであろう評価値を 1~5 の 5 値で付与したものである。実験協力者評価値においては、3 名の付与した評価値が一致したものを正解とした。

実験協力者が付与した評価値 300 件のうち、3 名の評価値が一致したものは 227 件だった。この 227 件のレビューの実験協力者評価値を正解データとして、本手法の系列の連続、反転の重み（最もレビューア評価値の推定精度が良かったパラメータ）を学習した評価式で精度を算出した。3 名の実験協力者が付与した評価値が一致しなかった 73 件のうち、各々が付与した評価値の差が 1 であったものが 66 件あった。残りの 7 件は差が 2 以上で、少数ではあるものの、評価者により判断が大きく異なるレビューが存在した。本手法により推定した評価値を実験協力者評価値と比較した場合、全てのカテゴリ、商品でレビューア評価値と比較した場合よりも差異が小さい。これは、レビューアが付ける評価値と実験協力者が付けた評価値の間で明確に差があることを示している。実験協力者評価値は 300 件中 227 件が一致、66 件が評価値の差が 1 であり、実験協力者によって評価値の付与基準が大きく異なるとは考えにくい。レビューアが付与する評価値と、レビューの閲覧者との間で評価値に対する認識が異なっていることが、本実験の結果から読み取れる。実際に学習に用いたレビューア評価値と比較した場合よりも実験協力者評価値と比較した場合の差異が小さいことから、本手法により閲覧者の受け取る評価値に近い評価値を推定できているといえる。

4.3.3 評価値推定についての考察

本手法による評価値推定では、系列による重みを考慮しない場合と比較して、平均二乗偏差で 0.2~0.43、近傍一致率で最大 26 ポイント評価値推定の精度が向上することを示した。

本手法を用いて、閲覧者の受け取る評価値に近い評価値を推定できることを示した。その際、類似したカテゴリ、商品のランキング上位、下位のグループ間では、類似していないカテゴリ、商品間と比べて、推定式のパラメータの差が小さいことがわかった。このことから、カテゴリや商品によって、文の書かれ方の傾向が異なると予想される。また、家電や炊飯器、冷蔵庫では、系列が反転した際の重みが、PSP やお菓子では、系列が連続した際の重みが比較的高くなった。PSP やお菓子は、本手法とレビューア、実験協力者とを比較した場合でも評価値推定の精度が他に比べて低く、レビューに人の主観が入りやすいカテゴリ、商品の場合は、レビュー記事の書かれ方そのものが他のカテゴリ、商品と異なっている可能性がある。

5. ま と め

本論文では、意見極性の系列に基づくレビュー記事の評価値推定について、類義語を用いて評価表現辞書を拡張、拡張した辞書を用いて文に極性を付与し、付与した系列の連続、および反転の重みを考慮した評価値推定手法を提案した。類義語を用いた評価表現辞書の拡張においては、名詞編で約 15 倍、用言編で約 30 倍の表現に極性を与えることができた。拡張後の辞書を用いた文への極性付与を行い、実際のレビュー記事に対して、拡張前の約 3 倍の文に極性を付与でき、なおかつ極性付与の精度が 10.3 ポイント向上した。極性系列の傾向による評価値推定においては、系列の傾向を考慮しない場合に比べて、平均二乗偏差で 0.2~0.43、近傍一致率で最大 26 ポイント推定精度が向上した。また、カテゴリ、商品ごとの比較により、本手法を用いて、閲覧者の受け取る評価値に近い評価値を推定できることを示した。この際、類似したカテゴリや商品では、類似していないカテゴリや商品と比べて、推定式のパラメータの差が小さかった。このことから、レビューはカテゴリや商品によって、文の書かれ方の傾向が異なるといえる。家電や炊飯器、冷蔵庫では、系列が反転した際の重みが、ゲームやお菓子では、系列が連続した際の重みが比較的高くなった。人の主観が入りにくいカテゴリと、入りやすいカテゴリで、極性の系列の傾向が異なっている可能性があり、本手法による推定精度にも差が出ている。

本手法による評価値の推定が上手くいかなかったレビュー記事、特に推定評価値とレビュー評価値の差が大きいものには、主に以下のようなものが存在した。極端に短いレビュー記事で、ごく少ない極性しか付与されず、その少ない極性の付与が誤っていたもの。レビュー記事とレビュー評価値が明らかに一致しないもの。このうち、前者は本手法を用いて、複数回辞書を拡張することや、コーパスによる機械学習などを用いて、極性の付与数を増やすことで一定の改善が見込まれる。後者は、スパムに近いレビュー記事である場合もあり、そのような場合は、レビュー記事の閲覧支援の観点から見ると、閲覧するレビューから外すことも考えられる。

辞書による評価値付与では、辞書に出てこない語に極性が付与されない。本手法では、この問題に辞書を拡張して対応したが、辞書の拡張で追加された表現が、必ずしも拡張元の表現と同じ極性を持つとは限らない。拡張後の極性付与の精度が約 70 %であることを考えると、極性の付与手法の検討は重要な課題である。同時に、高い付与精度で極性付与をした場合の評価値推定を今後評価をする必要がある。また、表現がどの程度肯定的、もしくは否定的かを考慮していない。今回は、1 文ごとに文中の表現の重みによるスコアリングをしていないが、表現のスコアリングにより、推定精度が向上することが考えられる。今回は隣り合う極性系列の連続や反転だけに注目したが、文単位で極性に重みを付与する場合には、広域的な系列も考慮することが重要である。また、本手法では評価式によるマッピングで評価値を推定している。今後の課題として、マッピング以外の手法による評価値推定、具体的には、極性系列を利用した、

機械学習によるレビューの多値分類を検討していく。

謝 辞

本研究の一部は筑波大学図書館情報メディア系プロジェクト研究による助成を受けたものである。

文 献

- [1] 乾孝司, 奥村学. テキストを対象とした評価表現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.
- [2] Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten De Rijke. Using wordnet to measure semantic orientations of adjectives. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC-2004)*, Vol. IV, pp. 1115-1118, 2004.
- [3] Minqing Hu. and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining (KDD2004)*, pp. 168-177, 2004.
- [4] Peter D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-2002)*, pp. 417-424, 2002.
- [5] 藤村滋, 豊田正史, 喜連川優. Web からの評判および評価表現抽出に関する一考察 (夏のデータベースワークショップ: Dbws2004). 電子情報通信学会技術研究報告. DE, データ工学, Vol. 104, No. 177, pp. 141-146, 2004.
- [6] 立石健二, 福島俊一, 小林のぞみ, 高橋哲朗, 藤田篤, 乾健太郎, 松本裕治. Web 文書集合からの意見情報抽出と着眼点に基づく要約生成. 情報処理学会研究報告 自然言語処理研究会報告, Vol. 2004, No. 93, pp. 1-8, 2004.
- [7] 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. 2, pp. 203-222, 2005.
- [8] 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.
- [9] Maite Taboada and Jack Grieve. Analyzing appraisal automatically. In *AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, pp. SS-04-07, 2004.
- [10] 江崎晃司, 松井藤五郎, 大和田勇人. Weblog 上の評判情報における形容詞の出現位置を考慮した賛否分類. 情報処理学会第 67 回全国大会, Vol. 2, pp. 393-394, 2005.
- [11] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2002)*, Vol. 10, pp. 79-86, 2002.
- [12] Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL-2005)*, pp. 115-124, 2005.
- [13] 清水航. レビューの語の重みを考慮したテキストマイニングによるゲームソフトの評判分析. Master's thesis, 法政大学大学院工学研究科, 2011.
- [14] 小倉達矢, 宮野晋平, 永井慎介, 山口実靖, 浅谷耕一. 良質なレビューの特性とそれを考慮した評判情報の抽出. 情報処理学会全国大会講演論文集, Vol. 70, No. 5, pp. 197-198, 2008.