

オノマトペを利用した商品の使用感の自動抽出

新里 圭司^{1,a)} 益子 宗^{1,b)} 関根 聡^{2,c)}

受付日 2014年7月10日, 採録日 2015年1月7日

概要: 本稿では商品の使用感を記述した文を商品レビューから抽出する手法について述べる。オンラインショッピングサイトでは、実店舗で買い物するときのように商品に触れたり、試したりしてから購入することができない。そのため、ユーザがいただく商品のイメージと実際に届く商品の間に、質感や食感などの使用感に関して不一致が生じることがあり、顧客満足度低下の原因となっている。購入前のユーザに対して、商品の使用感に関する情報を提供することはオンラインショッピングサイトの普及のために重要である。提案手法は、「オノマトペを含む文に出現しやすい表現は商品の使用感を記述する際に用いられやすい」という仮説に従い、単語とオノマトペのレビュー文中での共起の強さを計算し、得られた語の共起の強さをを用いてレビュー中の文が使用感を記述しているかどうか判定する。実験の結果、 F_1 値で 65.9 ポイントの精度で使用感を記述した文を抽出できることが分かった。

キーワード: オノマトペ, 商品の使用感, 商品レビューの解析

Automatic Extraction of Product Impression through Onomatopoeias

KEIJI SHINZATO^{1,a)} SOH MASUKO^{1,b)} SATOSHI SEKINE^{2,c)}

Received: July 10, 2014, Accepted: January 7, 2015

Abstract: This paper describes an automatic methodology for extracting sentences that contain product impressions (the description about “how a purchased product was after obtaining and using it”) from review data in an e-commerce site. E-commerce users cannot grasp such information before purchasing the product. This can be regarded as one of the shortcomings of e-commerce. It is important to convey this information to the users in order to prevent them from having a bad shopping experience. First, we investigate product review sentences that contain onomatopoeias, and reveal that these sentences tend to contain product impressions. Through this finding we assume that words frequently co-occurring with onomatopoeias are likely to be used for describing product impressions. According to this assumption, the proposed method calculates scores for given sentences using co-occurrence strength between words and onomatopoeias, and extracts the sentences that exceed a threshold value. The co-occurrence strength for each word is calculated from sentences in product reviews beforehand. The experimental results show that the performance of our method achieves an average F_1 score of 65.9 points and that the method outperforms its alternatives.

Keywords: onomatopoeia, product impression, product reviews analysis

1. はじめに

時間や場所を気にすることなく買い物可能なオンライン

ショッピングサイトは重要なライフラインになりつつある。しかし、実店舗で買い物するときのように、商品に触れたり、試したりしてから購入することができないという問題がある。そのため、ユーザがいただく商品のイメージと実際に届く商品の間に、質感や食感などに関して不一致が生じることがあり、顧客満足度低下の温床となっている。また、総務省が公開している情報通信白書^{*1}において、イ

¹ 楽天技術研究所
Rakuten Institute of Technology, Shinagawa, Tokyo 140-0002, Japan

² ニューヨーク大学
New York University, New York, NY 10003 USA

a) keiji.shinzato@mail.rakuten.com

b) so.masuko@mail.rakuten.com

c) sekine@cs.nyu.edu

*1 <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h23/html/nc213320.html>

インターネットで物品を購入しない主要な理由の1つに「実際に商品を見て買うことができない」点があげられており*2, ユーザの獲得という面においても問題となっている。以上より、「商品を実際に手にとってみて, または使ってみて*3どうだったか」という情報へのアクセスを容易にすることは, オンラインショッピングを今後さらに普及させるうえで重要である。この「商品を実際に手にとってみて, または使ってみてどうだったか」に関する記述を本研究では**商品の使用感**と呼ぶ。たとえば, 以下の文はシャンプーの使用感について述べていると考えられる。

- 前より洗い上がりのサッパリ感が増しました。
- 髪の毛にボリュームが出てきました!

このほかにも, 香り, 効果, 効能, 質感, 味わい, 着心地などの記述は商品の使用感を述べていると考えられる。

多くのショッピングサイトでは商品のレビューを公開している。しかしながら, レビューには使用感以外にも購入に至った経緯や, 利用目的, 店舗や商品に対する要望, クレーム, 購入者自身の情報など様々な記述が含まれるため, 使用感に関する情報を得るためにはユーザは1文1文読まなければならない。商品の使用感について記述した文をレビューから抽出し, 適切な形でユーザに提示することができれば, ユーザの負担を軽減させるだけでなく, 今までオンラインショッピングサイトの利用を躊躇していた新規ユーザの獲得にもつながる。

本稿では, オノマトペを間接的に利用することで, 商品の使用感について記述した文をレビューから抽出する手法について述べる。3章で述べるように, オノマトペを含むレビュー中の文を調査した結果, 商品の使用感が記述されやすいことが分かった。そこで, これを発展させ, オノマトペを含む文に出現しやすい表現は**商品の使用感を記述する際に用いられやすい**という仮説を設け, 商品レビューに含まれる単語とオノマトペの文中での共起の強さを計算する。そして, 各単語に対して計算された共起の強さをを用いてレビュー内の各文が使用感を記述したものであるかどうかを判定する。

本稿の構成は以下のとおりである。まず2章で本研究で用いるレビューデータについて説明し, 3章でオノマトペと商品の使用感の関係について調査した結果を報告する。続いて4章で提案手法について述べ, 5章で評価実験について報告し, 我々の仮説の有効性を実験的に示す。最後に6章で関連研究について述べる。

2. 使用するレビューデータ

本研究では, 2012年1月から2013年9月末までに楽天

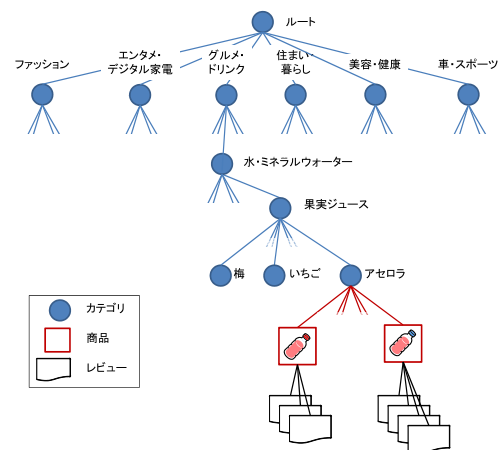


図1 楽天市場のカテゴリ, 商品, レビューの関係

Fig. 1 Overview of categories, products, and review data in Rakuten Ichiba.

が運営する「みんなのレビュー」*4へ投稿された約4,000万件のレビューデータを使用する。みんなのレビューは, 楽天市場で販売されている商品について書かれたレビューを公開しているサイトである。本稿ではこのレビューデータ全体を**レビュープール**と呼ぶ。

図1に楽天市場で販売されている商品と商品カテゴリ, レビューの関係を示す。ここでは例として「アセロラ味の果実ジュース」が登録されている様子を示した。レビューデータは楽天市場で販売されている1つの商品に紐付けられており, 商品は楽天によって設計されたカテゴリ体系*5の末端ノードに紐付けられている。つまり, カテゴリ体系の中間ノードに対して商品が紐付けられることはない。末端ノード数は44,244である。また, 商品からルートカテゴリまでのパスの長さはカテゴリによって異なっている。カテゴリ体系のルートは, サービス上では見えないようになっているため, その直下にある6つのカテゴリ(ファッション, エンタメ・デジタル家電, グルメ・ドリンク, 住まい・暮らし, 美容・健康, 車・スポーツ)が事実上の最上位である。本研究ではこの6つのカテゴリを**最上位カテゴリ**と呼ぶ。カテゴリとカテゴリ, カテゴリと商品, 商品とレビューの関係を利用することで, 「果実ジュースによせられたレビュー群」のように, カテゴリ単位でレビューをまとめることができる。

3. オノマトペを含む文と使用感の関係

商品レビュー内のオノマトペを含む文には, どの程度使用感が記述されやすいのか調査した。調査対象とするカテゴリの選定は, 最上位カテゴリそれぞれにおいて, 投稿されたレビュー中の文数に基づいて行った。カテゴリごとにレビュー文を無作為に抽出し, 2名の被験者により使用感を記述した文であるかどうかの判定を行った。以下, 調査

*2 購入しない理由の2番目にあげられている。
*3 商品が食品であれば「食べてみて」, 飲料であれば「飲んでみて」が「使ってみて」に該当する。

*4 <http://review.rakuten.co.jp/>
*5 <http://event.rakuten.co.jp/genre/>

表 1 調査対象カテゴリのレビュー数、レビュー中の文数、括弧内はオノマトベを含む文の数とその割合

Table 1 # of sentences and # of review articles in the investigated categories. # of sentences including any onomatopoeias and its ratio are shown in brackets.

カテゴリ	レビュー数	文数
レディースパンツ	320,751	1,118,288 (146,721 / 13.12%)
スマートフォン本体・アクセサリ	421,415	1,107,786 (95,639 / 8.63%)
米	170,583	456,309 (22,694 / 4.97%)
スーツケース	163,560	597,828 (42,823 / 7.16%)
シャンプー	204,114	599,590 (82,352 / 13.73%)
長袖ラッシュガード	57,013	186,449 (29,486 / 15.82%)
合計	1,337,436	4,066,250 (419,715 / 10.32%)

の詳細について述べる。

3.1 調査対象カテゴリの選択

まず、楽天市場のカテゴリ体系において、ブランド名、サイズ、味などカテゴリに対する属性値と見なせるノードを手で削除し、そのノードに紐付いている商品を親ノードに付け替えた。たとえば図 1 のアセロラ、いちご、梅カテゴリはすべて「味」と考えられるので、この操作により、これらに登録されている商品は親カテゴリである「果実ジュース」に紐付けられる。削除した結果、末端ノード数は 8,213 となった。

続いて、最上位カテゴリそれぞれについて、レビュー文数の多い上位 10 カテゴリをいったん選び出し、その中で最もオノマトベを含む文数が多いものを調査対象カテゴリとして選び出した。これは、レビュー数が多く、オノマトベを含む文数もある程度多いカテゴリを、最上位カテゴリそれぞれについて選び出すためである。レビューの文分割処理は「。」「!」「?」「♪」「…」 「(笑)」「(怒)」「(泣)」を手がかりに行った。ただし、これらの記号が鉤括弧 (「」『』) 内に出現した際は分割しない。オノマトベとしては「擬音語・擬態語の読本」[1] から、文字列長が 3 以上の見出し語、および各見出し語をカタカナに変換した 1,694 表現を利用した。これは、(1) 文字列長が 3 未満の見出し語 (たとえば「かん」など) はオノマトベではない表現にもマッチしやすい、(2) 文献 [1] にはひらがな表記しか記載されていない、(3) オノマトベはひらがな・カタカナの表記が異なっても、その意味は多くの場合変わらないと考えたためである*6。

最終的に以下の 6 つのカテゴリが選び出された (括弧内は最上位カテゴリ名である)。

- レディースパンツ (ファッション)
- スマートフォン本体・アクセサリ (エンタメ・デジタ

*6 実際はカタカナに変換することでオノマトベでない語にマッチしてしまう事例があった。たとえば笑いを浮かべる様を表す「にっ」とはカタカナに変換することで「ニット (knit)」にマッチしてしまっていた。

表 2 商品レビューにおける商品の使用感を記述した文の現れやすさ
Table 2 Likelihood of occurring sentences that describe product impressions in product reviews.

カテゴリ	使用感を記述した文の数	
	オノマトベを含む文に限定	任意の文
レディースパンツ	70	59
スマートフォン本体・アクセサリ	78	31
米	60	33
スーツケース	64	32
シャンプー	82	45
長袖ラッシュガード	79	42
合計	433 (72.17%)	242 (40.33%)

ル家電)

- 米 (グルメ・ドリンク)
- スーツケース (住まい・暮らし)
- シャンプー (美容・健康)
- 長袖ラッシュガード (車・スポーツ)

各カテゴリのレビュー数、レビュー文数、オノマトベを含む文数およびその割合を表 1 に示す。カテゴリによりオノマトベを含む文数の割合に差があることが分かる。

3.2 商品レビューにおける使用感を記述した文の割合

調査対象カテゴリの商品に対して書かれたレビューを収集し、対象カテゴリごとにオノマトベを含む文を無作為に 100 文、オノマトベを含んでいるかどうか関係なく無作為に 100 文、計 1,200 文選び出し、使用感が記述された文であるかどうか判定した*7。抽出された文の判定は、2名の被験者によって行った。被験者には、「商品を実際に手にとってみて、または使ってみてどうだったか」という記述が含まれている場合を正解とするよう指示した。ただし、例外として次の 4 点を設けた。

- 「おまけ」の使用感是不正解とする。
- 「味がさっぱりしていて良かったと喜んでいました」のようにレビューの著者と経験した人物が別の場合でも正解とする。
- 「前回のスーパーセールで買ってみて、味が良かったので今回も購入してみました。」のように、同じ商品であれば過去の購入に基づく記述でも正解とする。
- 何が評価の対象となっているか分からない場合は不正解とする。

被験者間の κ 統計量 [2] は 0.813 であり、これは very good agreement とされる値である。両被験者がともに正解と判定した文のみを正解、つまり使用感が記述された文とした。

調査結果を表 2 に示す。表中の「オノマトベを含む文

*7 このとき、文の誤分割の影響を減らすために、文字数が 10 文字以上かつ 100 文字以下の文に限定した。

きちんと、きつと、ぼつちり、ガッカリ、がっかり、ぱつちり、キチンと、ドキドキ

図 2 不正解の文により多く含まれていたオノマトペ

Fig. 2 Onomatopoeias which tend to be contained in wrong sentences.

に限定」は、オノマトペを含む文を無作為に選び出した場合、「任意の文」はオノマトペを含んでいるかどうかに関係なく無作為に文を選び出した場合の結果を表す。カテゴリによって差があるものの、表よりオノマトペを含む文の平均 72.17%が商品の使用感を記述しているものであることが分かる。一方で、オノマトペの出現を前提としない場合は 40.33%であり、2つの数字を比べるとオノマトペを含む文には使用感が記述されやすいことが分かる。

図 2 に正解よりも不正解の文に多く含まれていたオノマトペの一覧を示す。不正解数が最も多かったオノマトペは「きちんと」であった。この理由は、「きちんと」は以下の文のように店舗の対応や、商品の発送に関する情報を記述する際に用いられることが多いためである。

- 個人的な質問にもきちんと答えて頂き感謝しております。
- きちんと商品が届きました。

本節の調査で得られた商品レビュー中でオノマトペを含む文の特徴を整理すると以下ようになる。

特徴 1：オノマトペを含む文では商品の使用感が記述されやすい。

特徴 2：商品レビューにおいて、オノマトペを含む文は出現しにくい (10%程度)。

以上よりオノマトペを手がかりにすることで精度良く使用感を記述した文を収集することができるが、その数は多くないことが分かる。

4. 提案手法

提案手法の概要を図 3 に示す。本手法では 3.2 節で得られた知見 (特徴 1) を発展させ、オノマトペを含む文に出現しやすい表現は商品の使用感を記述する際に用いられやすいという仮説を設ける。そしてこの仮説に基づき、任意のオノマトペと単語のレビュー文中での共起の強さを計算し、得られた共起の強さを使得レビューから使用感について書かれた文を発見する。提案手法は以下の 3つのステップからなる。

Step1：レビューの文分割とオノマトペに基づく文の分類

Step2：オノマトペと語の共起の強さの計算

Step3：文のスコアリング

以下、各ステップについて述べる。

4.1 レビューの文分割とオノマトペに基づく文の分類

本ステップでは、入力として与えられたカテゴリに紐付

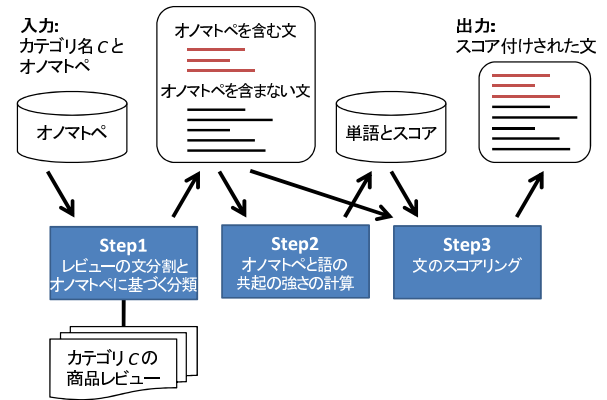


図 3 提案手法の概要

Fig. 3 Overview of the proposed method.

く商品のレビューを文単位に分割し、オノマトペを含む文と含まない文に分類する。商品レビューを文に分割する処理は 3.1 節と同じ方法で行う。

続く文の分類についても、基本的には 3.1 節同様、文献 [1] の見出し語のうち文字列長が 3 以上のもの、およびそれらのカタカナ表記を利用する。しかし、図 2 にあげたオノマトペは使用感を記述していない文に出現しやすいため除く。カテゴリごとに異なるオノマトペの集合を準備することで、提案手法の性能の向上が考えられるが、今回は全カテゴリにおいてこのオノマトペ集合を用いる。

4.2 オノマトペと語の共起の強さの計算

続いて、先述の仮説に基づき、レビューに含まれる各単語とオノマトペの共起の強さを求める。本研究では、名詞、動詞、形容詞^{*8}のうち、レビュープール内の 100 件以上のレビューに含まれている語を対象とした。形態素解析器には MeCab^{*9}、形態素解析用辞書としては NAIST Japanese Dictionary version 0.6.3b^{*10}を用いた。未知語については、多くの場合名詞と考えられるので、名詞同様 100 件以上のレビューに含まれている語を対象に加えた^{*11}。各単語とオノマトペの共起の強さには、相互情報量をもとにした以下の値を用いる。

$$\text{Score}(w) = \log(1 + \text{cooc}_{w,O}) \times \log \frac{p(w,O)}{p(w) \cdot p(O)}$$

ここで $\text{cooc}_{w,O}$ は語 w と任意のオノマトペをともに含む文の数^{*12}、 $p(w,O)$ は任意のオノマトペと語 w をともに含む

^{*8} ただし、NAIST Japanese Dictionary version 0.6.3b において、品詞の細分類が非自立、もしくは接尾の語は除く。これらに加え、名詞については細分類が数、代名詞、副詞可能となっている語も除く。

^{*9} <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

^{*10} <http://sourceforge.jp/projects/naist-jdic/>

^{*11} ただし、ひらがな、カタカナ、漢字、アルファベットを 1 文字も含まないものは利用しない。

^{*12} 「びつたり」のように形態素解析の結果によっては語 w がオノマトペになることがある。共起しているオノマトペと語 w が同じ場合であっても「共起」として扱った。

表 3 語とそのスコアの例. 括弧内の数字は獲得された単語の異なり数

Table 3 Examples of words and their scores. # of words for each category is shown in brackets.

レディースパンツ (9,966)		スマートフォン本体・ アクセサリ (10,259)		米 (4,846)		スーツケース (7,262)		シャンプー (7,707)		長袖ラッシュガード (4,433)	
スコア	単語	スコア	単語	スコア	単語	スコア	単語	スコア	単語	スコア	単語
17.476	びったり	19.435	びったり	15.057	びったり	17.963	作り	17.130	サラ	12.915	びったり
15.706	ブカブカ	17.714	キラ	14.286	ギリギリ	17.189	ギリギリ	14.119	ギシギシ	11.161	生地
14.720	ぶかぶか	16.501	めちゃくちゃ	14.263	めちゃくちゃ	16.852	パンパン	13.448	ゴワゴワ	10.397	ブカブカ
14.030	ピタピタ	16.273	はまる	13.673	モリモリ	15.947	びったり	13.252	感	9.810	ピタピタ
:	:	:	:	:	:	:	:	:	:	:	:
4.499	厚手	4.486	遊び	4.503	蓋	4.498	遊ぶ	4.498	つける	4.491	イヤ
4.466	固め	4.480	密着	4.473	好み	4.424	詰めこむ	4.497	洗い上り	4.484	感じ
4.453	めだつ	4.470	ピカ	4.467	炊きたて	4.390	度	4.497	ワサ	4.400	かぶれる
4.422	素肌	4.463	ガツツキ	4.465	忘れる	4.368	キャスター	4.497	ささる	4.385	ストン
:	:	:	:	:	:	:	:	:	:	:	:
0.001	着用	0.004	マイメモ	0.003	すすめる	-0.002	探せる	0.003	ハマる	0.004	もらえる
0.000	品物	0.001	謎	-0.001	国内産	-0.002	モスグリーン	-0.001	体験	0.004	お知らせ
-0.001	気持	-0.002	ヒビ	-0.005	華	-0.003	足りる	-0.002	DS	-0.005	深い
-0.002	頻繁	-0.003	できる	-0.005	昼食	-0.004	含む	-0.002	申す	-0.009	少し
:	:	:	:	:	:	:	:	:	:	:	:
-6.390	リピート	-6.641	対応	-5.504	玄関	-4.582	パープル	-8.563	安い	-5.441	楽しみ
-6.489	リピ	-6.644	注文	-5.662	無料	-4.878	男女	-8.812	無料	-5.683	沖縄
-6.665	対応	-6.858	楽しみ	-5.878	送料	-4.979	レッド	-8.894	対応	-6.040	対応
-6.698	発送	-7.499	購入	-6.718	助かる	-6.545	購入	-9.232	送料	-7.440	行く

文が出現する確率, $p(w)$ は語 w を含む文が出現する確率, $p(O)$ は任意のオノマトベを含む文が出現する確率である. スコアを計算する際, 個々のオノマトベを区別していないことに注意されたい. また相互情報量単体では出現頻度の低い単語に対して大きなスコアが与えられる傾向があったため, オノマトベとの共起頻度が高い単語ほどスコアが大きくなるように $\log(1 + coc_{w,O})$ を乗じて補正した. 加えてゼロ頻度問題に対応するため, $p(w, O)$ を算出する際の共起頻度に対して 1 を加算した.

表 3 に対象カテゴリそれぞれに対して得られた単語とそのスコアの例を示す. 表より「びったり」「ギリギリ」「ブカブカ」「ゴワゴワ」などのオノマトベに対して高いスコアが与えられていることが分かる. これらの語はオノマトベであるが, その品詞が名詞, または未知語であるためオノマトベとの共起の強さが計算されている. スコアが中程度の部分を見ると「厚手」「密着」「炊きたて」「詰めこむ」「洗い上り」「かぶれる」などの単語が並んでおり, オノマトベとの共起を見ることで使用感を記述する際に用いられるであろう単語をうまく獲得できていることが分かる. その一方で「発送」「購入」「リピート」「無料」「対応」「助かる」など, 使用感以外の情報を記述する際に用いられるであろう単語には低いスコアが与えられており, 本研究で用いた仮説が有効にはたらいっていることが分かる.

4.3 文のスコアリング

最後に前節で計算した単語のスコアを用いて文 s が商品の使用感について記述しているかどうかを表すスコア $S(s)$ を計算する. ここでは文を形態素解析し, 文に含まれる各単語のスコアの総和を文 s のスコアとして用いる.

$$S(s) = \sum_{w \in s} \text{Score}(w)$$

スコアが同点の場合は, 同点の文をその文長 (文字数) に従って降順に順位付けし, 文長も同じ場合はランダムとした. 提案手法により抽出された文の応用を考えた場合, 多くの単語を使って使用感を記述した文の方がユーザに対してより具体的な商品のイメージを与えやすいとの考えから, 単語数 $|s|$ による $S(s)$ の正規化は行わない.

5. 評価実験

本章では提案手法の評価について述べる. 5.1 節では評価データ, 5.2 節では比較実験に用いるベースライン手法について述べる. 5.3 節では, 提案手法単体の性能評価, ベースライン手法との比較実験に加え, 提案手法の汎用性を確認するために行ったオノマトベを含む文の割合と精度の関係の調査についても述べる. 5.4 節では文のスコアが 0 より大きいものを出力としたときの false-positive/negative の事例について考察する.

5.1 評価データ

評価データには 3.2 節の調査で用いた、「オノマトペを含んでいるかどうかに関係なく無作為に選び出した文」を用いた。このデータはカテゴリごとに 100 文、計 600 文からなる。2 名の被験者により使用感の記述を含むかどうかの判定がなされているが、56 文について判定の不一致が見られた。そこで、この不一致を被験者同士の話し合いにより解消した後、評価実験に用いた。話し合いの結果、新たに 8 件の正解が加わり、使用感の記述を含む文の数は 250 件になった。

5.2 ベースライン手法

以下の 3 つの手法をベースラインとし、提案手法との比較を行った。

手法 1. この手法ではカテゴリと単語の関連度の強さに基づいて文をスコアリングする。たとえばお米カテゴリに対する「炊きたて」やシャンプーカテゴリに対する「洗いり」などの単語は使用感を記述した文に現れやすいと考えられると同時に、各カテゴリと関連の強い単語と考えられる。その一方で使用感を記述した文に現れにくいであろう「購入」や「発送」などの単語は様々なカテゴリに出現すると考えられ、特定のカテゴリと関連が強くなりにくいと考えられる。以上より、カテゴリと単語の関連の強さをを用いることで使用感を記述した文が抽出できるのではないかと考え、カテゴリと単語の関連度を基にしたスコアリング手法をベースライン 1 とした。

カテゴリと単語の関連度は相互情報量をもとに求めた。

$$\text{Score}_{\text{base1}}(w) = \log(1 + \text{cooc}_{w,c}) \times \log \frac{p(w,c)}{p(w) \cdot p(c)}$$

ここで $\text{cooc}_{w,c}$ は、カテゴリ c に登録された商品のレビューのうち語 w を含むものの数、 $p(w,c)$ はカテゴリ c に登録された商品のレビューに語 w が出現する確率、 $p(w)$ は語 w を含むレビューが出現する確率、 $p(c)$ はカテゴリ c に登録された商品のレビューが出現する確率である。これらの値はレビュープール全体から計算した。また、提案手法同様、ゼロ頻度問題回避のため、 $p(w,c)$ を計算する際に用いる共起頻度には 1 を加えた。文のスコアには、提案手法と同じく文中に含まれる各単語のスコアの総和を用いる。

手法 2. この手法ではオノマトペを 1 つ以上含む文を使用感を記述した文として抽出する。提案手法同様、オノマトペ辞書には 4.1 節で用いたものを用いる。

手法 3. 評価表現を含む文は使用感について述べていると考えられるため、評価表現が 1 つ以上含まれる文を出力する手法を準備した。評価表現としては、小林

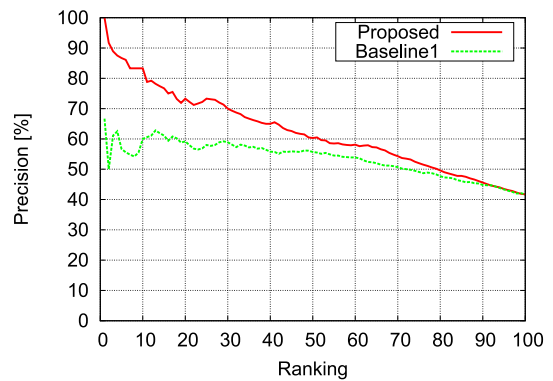


図 4 評価結果 (各カテゴリの平均)

Fig. 4 Evaluation results (average).

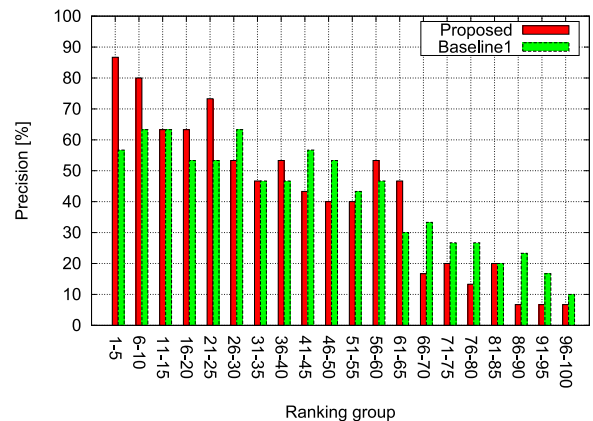


図 5 順位グループごとの精度 (各カテゴリの平均)

Fig. 5 Precision of each ranking group (average).

ら [3] の辞書および東山ら [4] の辞書^{*13}に収録されていた 13,560 語を用いた。

5.3 実験結果

5.3.1 提案手法の評価およびベースライン手法 1 との比較

スコアに従って文を降順にソートしたときの順位を横軸、当該順位までの文を出力としたときの精度を縦軸としたときの結果を図 4 および図 6 に示す。また、スコアに従って文を降順にソートした後、5 位区切りで順位グループを作成し、各グループの精度を示したものを図 5 および図 7 に示す。図 4 および図 5 は各カテゴリの評価結果を平均化したもの、図 6 および図 7 はカテゴリごとの個別の評価結果である。また図 6 では、提案手法、ベースライン手法 1 のどちらの場合にもスコアが 0 の部分に「▼」を印した。

図 4 から提案手法の精度を示したグラフの線が右下がりになっていること、図 5 からは順位が下がるにつれて各順位グループの精度が低くなっていることが分かる。図 6 においても、グラフの線の滑らかさがカテゴリごとに異なるものの図 4 と同様の傾向が見られ、また図 7 において

*13 ともに <http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary> から取得可能。東山らの辞書は p と n のみを用いた。

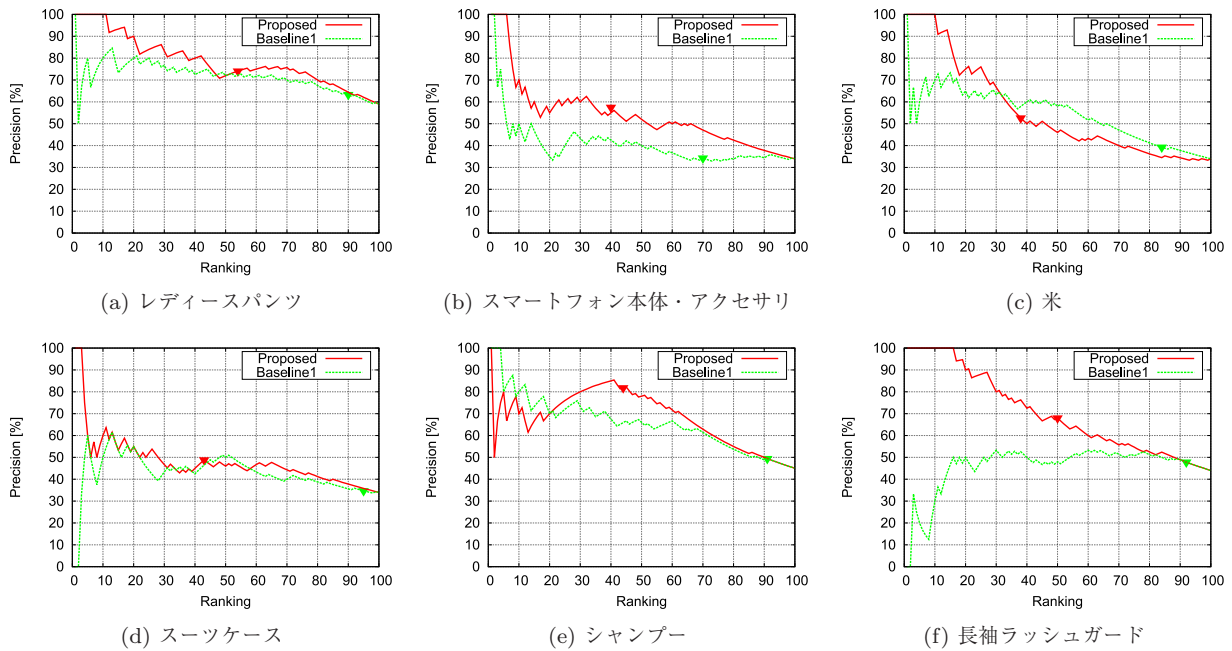


図 6 カテゴリごとの評価結果

Fig. 6 Evaluation results (each category).

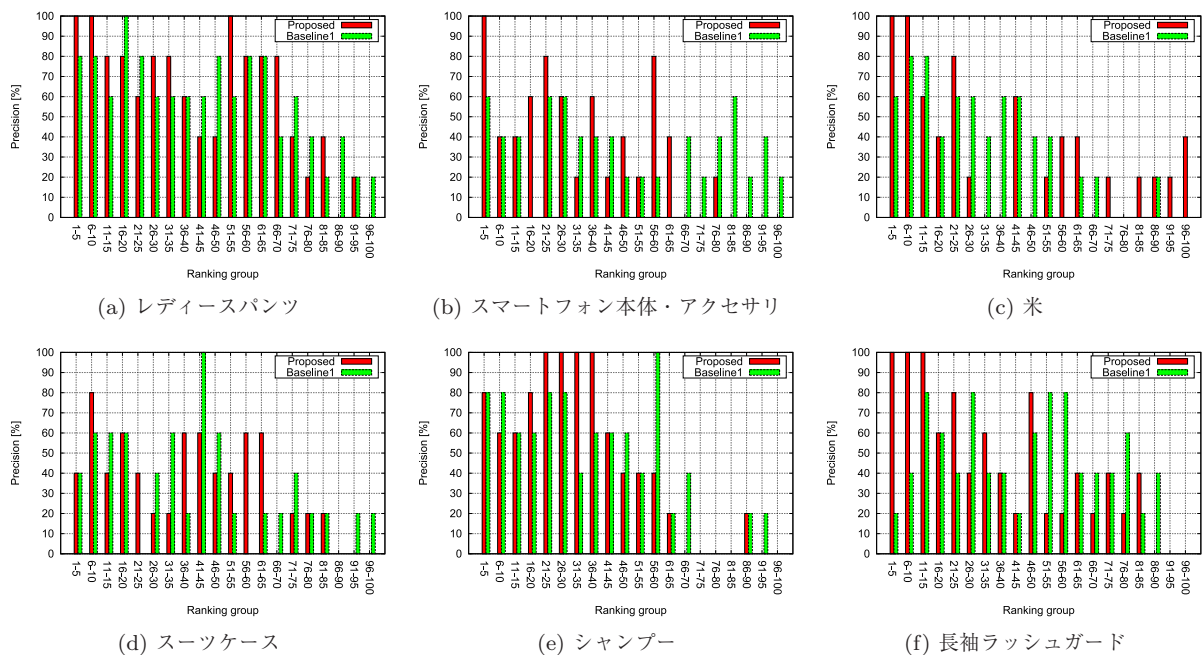


図 7 順位グループごとの精度 (カテゴリ別)

Fig. 7 Precision of each ranking group (each category).

も、スーツケース以外のカテゴリについては、順位の若いグループの方が精度が高い傾向にあることが分かる。以上より、本稿で提案したオノマトベと単語の共起の強さを利用した文のスコアリング手法が使用感を記述した文の抽出に効果があるといえよう。さらにこの結果から、本手法で用いた仮説「オノマトベを含む文に出現しやすい表現は商品の使用感を記述する際に用いられやすい」の有効性が、間接的にはあるが実験的に示されたと考えられる。図 4, 5, 6, 7 には比較のために、ベースライン手法 1 の性能も

示している。各図より提案手法の方が使用感を記述した文の抽出に適していることが分かる。

図 4, 図 5 より、入力として与えた文全体の 10%にあたるスコア上位 10 文を出力した場合、その精度は平均で 80%以上と非常に高いことが分かる。実際の応用を考えた場合、使用感が記述された数文だけを利用できればよい場合も少なくない。たとえばスマートフォンユーザに対して使用感を記述した文を提示することを考えたとき、表示領域が限定されているためスコアの高い文を商品ごとに数文

表 4 文とそのスコアの例 (シャンプー)

Table 4 Example of sentences with scores (shampoo).

判定	スコア	文
o	38.41	泡立ちはとても良く、洗い上がりは髪はきしまず、リンスいらす。
x	33.93	1 回目の洗いはオイリーでしっかり油分を取って、2 回目の洗いはドライで髪にやさしく洗っています。
o	33.44	前より洗い上がりのサッパリ感が増しているように思います。
o	32.41	仕上がりもよく髪がサラサラになりました。
o	24.46	ノンシリコンはあまり泡立たなかつたり、洗髪後髪がキシキシしたりする印象が強いのですが、このシャンプーは泡立ちもよく、キシキシしません。
:	:	:
x	-23.88	お店の方の対応は迅速でとても良いです♪
x	-24.01	こちらですと、大変お安く買えるのでリピです。
x	-32.38	今回、安い価格にポイントを使用して購入できたので、とても得した気分です。
x	-40.14	送料無料なので、これと一緒に買って他の商品も無料になるのは助かります♪
x	-49.94	某解析サイトを拝見し、コスパも良いし絶賛だったのでお試し購入☆楽天ポイントを使ったら更にお安くなりました

表 5 ベースライン手法 2, 3 との比較結果

Table 5 Comparison results with the baseline methods.

カテゴリ	ベースライン手法 2			ベースライン手法 3			提案手法		
	P[%]	R[%]	F ₁	P[%]	R[%]	F ₁	P[%]	R[%]	F ₁
レディースパンツ	100.0	22.0	36.1	73.8	76.3	75.0	74.1	67.8	70.8
スマートフォン本体・アクセサリ	60.0	17.6	27.3	46.3	91.2	61.4	56.1	67.6	61.3
米	100.0	5.9	11.1	46.0	85.3	59.8	52.6	58.8	55.6
スーツケース	40.0	5.9	10.3	39.4	76.5	52.0	48.8	61.8	54.5
シャンプー	72.7	17.8	28.6	55.0	73.3	62.9	81.8	80.0	80.9
長袖ラッシュガード	100.0	31.8	48.3	50.0	79.5	61.4	68.0	77.3	72.3
マクロ平均	78.8	16.8	27.0	51.8	80.4	62.1	63.6	68.9	65.9

出力できればよい。このような応用においては、提案手法の性能は十分実用的であると考えられる。

最後にシャンプーカテゴリの文に対して計算されたスコアを表 4 に示す。表より使用感について記述された文に対しては高いスコアが、一方で使用感の記述を含まない文に対しては低いスコアが与えられていることが分かる。

5.3.2 ベースライン手法 2, 3 との比較

次にベースライン手法 2, 3 と提案手法の比較実験を行った。その結果を表 5 に示す。提案手法の性能はスコアが 0 より大きい文のみを出力としたときのものである。表よりベースライン手法 2 は精度が高く、再現率が極端に低いことが分かる。これは 3.2 節で述べた商品レビュー内のオノマトベを含む文の特徴と照らし合わせて考えると妥当な結果であろう。提案手法はオノマトベと単語の共起の強さに基づいて文の抽出を行うことで、ベースライン手法 2 に比べ精度は劣るものの高い再現率を達成しており、結果的に F₁ 値を大きく改善している。また提案手法とベースライン手法 3 を比べると、精度および F₁ 値の平均値について提案手法の方が高いことが分かる。その一方で、シャンプー以外のカテゴリではベースライン手法 3 の方が再現率は高いことが分かる。

評価データに含まれる 250 件の使用感を述べた文のう

表 6 提案手法およびベースライン手法 2, 3 で獲得できた/できなかった文の数

Table 6 # of sentences extracted and not extracted by the proposed method and the baseline method 2 and 3.

		ベースライン手法 2		ベースライン手法 3	
		獲得	→ 獲得	獲得	→ 獲得
提案手法	獲得	41	133	144	30
	→ 獲得	4	72	55	21

ち、ベースライン手法 2 (オノマトベの有無に基づく方法) で獲得できた文は 45 件、ベースライン手法 3 (評価表現の有無に基づく方法) で獲得できた文は 199 件、提案手法で獲得できた文は 174 件であった。ベースライン手法で獲得でき提案手法で獲得できなかった文、およびベースライン手法で獲得できなかったが提案手法で獲得できた文の数を表 6 に示す。表よりベースライン手法 2 で獲得できる使用感を述べた文の 9 割以上は提案手法で獲得できること、およびベースライン手法 2 では獲得できない使用感を記述した文を数多く獲得できていることが分かる。またベースライン手法 3 の方が提案手法よりも再現率が高いため、ベースライン手法 3 で獲得できて提案手法で獲得できない文数は、提案手法で獲得できてベースライン手法 3 で獲得できない文数よりも多いことが分かる。

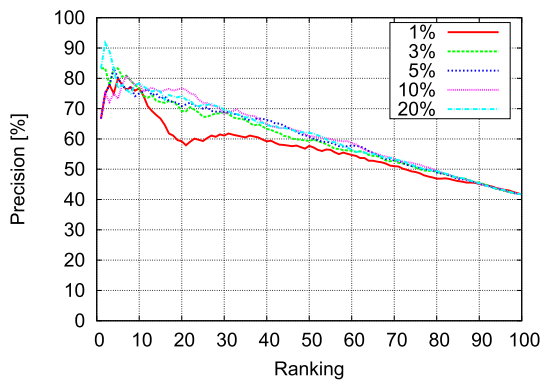


図 8 オノマトペを含む文の割合と精度の関係 (各カテゴリの平均, $N=10,000$)

Fig. 8 % of sentences including onomatopoeias vs. precision (average, $N=10,000$).

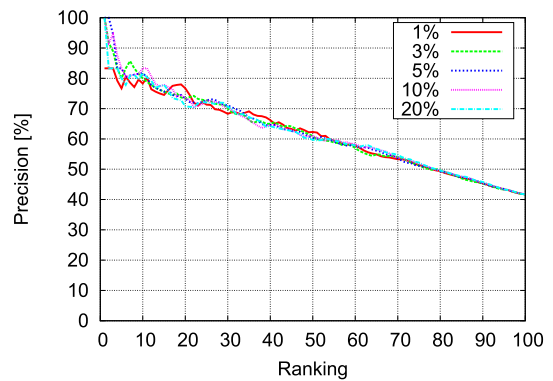


図 9 オノマトペを含む文の割合と精度の関係 (各カテゴリの平均, $N=100,000$)

Fig. 9 % of sentences including onomatopoeias vs. precision (average, $N=100,000$).

提案手法で獲得できてベースライン手法 2, 3 両方で獲得できなかった文は全部で 27 文あった。その一例を以下に示す。

- 粒 は割と大きめで、緑のお米も少ない印象です。[米]
- すごく伸び がいいのではきやすいです！ [レディースパンツ]
- きしみ もそんなに気になりません。[シャンプー]
- 軽い、移動 しやすい。[スーツケース]
- 何か リボン の取り付けかたごつ。[スマートフォン]
- 生地 はよいと思います。[ラッシュガード]

下線は文中で最もスコアの高かった単語を表す。これを見ると、お米カテゴリに対する「粒」や、シャンプーカテゴリの「きしみ」、スーツケースカテゴリの「移動」など評価表現とは見せない語に対して計算されたスコアが使用感を記述した文の抽出に有効であることが分かる。このように提案手法はオノマトペや評価表現以外の単語に対しても適切なスコアが計算できるため、これらを含まない文であっても使用感を記述した文として抽出することができる。

5.3.3 提案手法の汎用性の調査

ここまでの実験は、楽天市場の最上位カテゴリからレビュー文数が多く、オノマトペを含む文数もある程度多いカテゴリを対象に行っているため、提案手法に有利な設定となっている可能性があり、実験に用いたカテゴリ以外のデータに対して手法がどれだけ有効なのかは不明である。しかしながら、一般にオンラインショッピングサイトのカテゴリ数は楽天に限らず数が多いため、いくつかのカテゴリについて実験すれば汎用性が示せるのか、という点は自明でない。そこで、実験に用いたカテゴリについて辞書構築を行う際、オノマトペを含む文と含まない文の割合を調整することで擬似的にオノマトペを含む文が少ない、または多い環境を作り提案手法の精度を調査した。具体的には、まず、オノマトペを含む文の割合が 1%, 3%, 5%, 10%, 20% となるように各カテゴリから無作為に N 文抽出し、こ

表 7 false-positive/false-negative の数

Table 7 # of false-positive/false-negative errors.

カテゴリ	false-positive	false-negative
レディースパンツ	14	19
スマートフォン本体・アクセサリ	18	11
米	18	14
スーツケース	22	13
シャンプー	8	9
長袖ラッシュガード	16	10
合計	96	76

の文を使ってオノマトペと単語の共起の強さを計算した。そして、計算された共起の強さをういたときに使用感を記述した文の抽出精度がどう変化するか調べた。実験では N として 10,000 および 100,000 を用いた。

実験結果を図 8, 図 9 に示す。実験結果より、 $N=10,000$ かつ、オノマトペを含む文の割合が 1% のときのみ精度が悪くなっていることが分かる。レビュー文数が 1 万以下かつ、オノマトペを含む文の割合が 3% 未満のカテゴリ数を調べたところ、8,213 件中 468 件 (5.7%) であった。これは、使用感を記述した文の抽出が、今回評価実験に用いたカテゴリと同程度の精度で他の多くのカテゴリに対しても行えることを示唆していると考えられる。

5.4 考察

本節では、文のスコアが 0 より大きいものを出力したときの false-positive/negative の事例について考察する。各カテゴリにおける false-positive/negative の数を表 7 に、エラーのタイプごとの事例数を表 8 に示す。

5.4.1 false-positive の分析

購入目的・背景、価格や店舗の対応に対する評価・感想、身長や体型といったレビュー筆者自身の情報を述べる中で、オノマトペと共起の強い語が用いられ、結果として文のスコアが高くなっている事例が最も多かった。以下に例を示す (下線部は文中でオノマトペとの共起が最も強い語

表 8 False-positive/false-negative のエラータイプおよび事例数
Table 8 Error types of false-positive/false-negative and # of their instances.

false-positive		false-negative	
タイプ	事例数	タイプ	事例数
購入背景・評価・感想	40	買い物	28
文脈少	21	色・デザイン	25
推測・願望	10	購入目的・背景	8
心配・不安	5	その他	15
その他	20		

を表す)。

- お米を洗う際には気を付けようと思います。[米]
- 下半身太めでLサイズの10分丈を購入しました。[レディースパンツ]
- くせ毛でカラーリングで傷んだ髪なので3年ほど使い続けています。[シャンプー]
- 商品自体には問題ありませんでしたが、丁寧な梱包も商売人として当たり前のことだと思います。[スマートフォン]

このような事例を除くためには、「使用感の記述」以外にもレビュー文のタイプを推定するモデルを構築し、各モデルの出力を総合的に用いて文のタイプを判定するなどの処理が必要だと考えられる。

次に多かった事例は、文単体では文脈が少ないために、「使用感」を記述したものであるかどうか判断できないものであった。以下に例を示す。

- その汚れがなかなか取れませんでした [スーツケース]
- その点ではとてもいい商品だと思います！ [ラッシュガード]
- ですが、物が悪いって事ではなく、詰めが甘いって事です。 [スマートフォン]

これらを改善するためにはレビューにおける談話の構造を解析し、その結果に基づいて抽出すべきかどうか判断する必要があるだろう。また21件中10件は、以下の例のように文分割処理が失敗したために文脈が少なくなっていた。

- と疑問符ばかりが頭に浮かびました。
- 170 cm の身長に約6？

文分割処理の精度を改善することでこれらの事例は減らせると考えられる。

以下の文のように、推測・願望などを述べている文に対しても高いスコアが与えられていた。

- ストラップをつければ、夏のレジャーにも活躍しそうです♪ [スマートフォン]
- 早くこれを着て泳ぎたいです。 [ラッシュガード]

これらの文を出力から除くためにはモダリティ解析が必要であり、その導入は今後の課題である。

最後に、単語「心配」「不安」を含む文に対しても高いス

コアが付与されていた。これは「心配」「不安」が以下の例文のようにオノマトペ「ちょっと」とともに用いられやすく、両単語に対して高いスコアが計算されていたためであった。

- ヒップと太ももがかなり立派な体型なので、ちょっと心配でしたが、Lサイズでぴったり。
- 食べ物の通販はちょっと不安でしたが、全く問題なくとても美味しく頂いています。

このような事例を除くためには、「心配」「不安」をストップワードとして登録する、もしくはカテゴリごとに利用するオノマトペを選別する必要があると考えられる。

5.4.2 false-negative の分析

オノマトペと共起しにくい語を含んでいるために、文のスコアが低くなっているケースがほとんどであった。オノマトペと共起しにくい語はいくつかのパターンに分類することができた。最も多いパターンは“買い物”特有の語であり、「価格」、「安い」、「購入」、「届く」、「注文」、「発送」、「リピート」などの語が該当する。複文にこれらの語が含まれている事例について、そのスコアを下げる原因となっていた。以下に例を示す（下線は文中で最もスコアが低かった単語）。

- 近所のドラッグストアで買った 安い お米と似たような味でした [米]
- 冷めてもおいしいし、家まで届くし、3度目の リピート です。 [米]
- 白を 購入 しましたが、スケ感もなく、日焼けも完全に防げ、本当に良い買い物だったと思います。 [ラッシュガード]
- 届いて、あまりの軽さにびっくりです。 [スーツケース]
- メタリックグリーンを 注文 しましたが、届いて思ったよりステキな色合いでしたので大満足です。 [スーツケース]

次に多かったパターンは、単語「色」や「デザイン」、「ブラック」などの色名を表す語であった。これらの語もオノマトペと共起しにくく、低いスコアが割り当てられていた。結果、以下に示す例のように、単語「色」「デザイン」や色名を含む文のスコアは低くなっていた。

- 色 もきれいで発色もよく、サイズもぴったりでした。 [レディースパンツ]
- デザイン はもちろんアルミ製バンパーなのに脱着が工具無しで出来るのでとても楽で満足しています。 [スマートフォン]
- 黒 系のスーツと合わせても違和感は無いです。 [スーツケース]

3つ目のパターンは「購入目的・背景」を述べる際に使われやすい語を含むものであった。たとえば、スーツケースに対する「出張」、ラッシュガードに対する「海」「日焼け」などが該当する。以下の文は、これらを含んでいるた

めに文としてのスコアが低くなっていた。

- 今までも同じ 72L でしたが、一回り小さく、重量も非常に軽くて海外出張先であちらこちら移動が多いので使い勝手よさそうです。[スーツケース]
- 乾きも早いし日焼けも防げて、大変便利でした！[ラッシュガード]

ここであげたような事例を獲得するためには、単純に単語のスコアを考慮するだけでなく、同一文中の単語の組合せ、複文かどうかの情報などを考慮する必要がある。

6. 関連研究

従来よりレビューを対象とした評判分析の研究は数多く行われている [5]。以前はレビュー内の意見を好評・不評に分類するタスクが主なものであったが、最近はこの軸に加え、より実用面を考慮した要望やクレームといった軸で分類する試みも多い [6], [7]。従来とは異なる分類軸をレビューに対して提起したものとしては、安藤ら [8] の研究がある。安藤らは「商品」「売り手」「買い手」の視点から、商品レビューに記述されている情報を 23 種類のカテゴリに人手で分類し、商品カテゴリごとにどのような種類の情報が出現しやすいか調査した。本研究で注目している「使用感」は彼女らが定義した「判断」と近く、本研究はその自動抽出手法と見なすことができる。

ブログなどの Consumer Generated Media から人々の経験に関する情報の抽出を試みる「経験マイニング」というタスクがある。Inui ら [9], Abe ら [12] は、ブログ記事からトピック、経験主、事態タイプ、事実性情報といった情報を抽出し、記事を構造化することで、任意のトピックに対するトラブル・要望の検索や、「商品 A に興味はあるが未購入のブロガー」といった複雑な検索を実現している。倉島ら [10] は Inui ら, Abe らとは異なり、経験を < 時間, 空間, 動作, 対象, 感情 > の 5 つ組で定義し、各スロットの値をブログ記事から抽出し、記事の構造化を行っている。また、ブログ記事中から経験が記述された文を抽出することに主眼をおいた研究もある [11]。経験マイニングでは、テキストに書かれているすべての経験の抽出を試みるため、商品レビューからは購入背景や、店舗の対応、使用するまでの準備、使用方法といった使用感以外の情報も抽出対象となるのに対し、提案手法は使用感に関する情報だけを対象とする。1 章で述べたように、使用感に関する情報へのアクセスを容易にすることはオンラインショッピングをより普及させるうえで重要であり、使用感に関する情報はレビュー中に記述された購入者の経験の中でも特に重要なものである。従来の経験マイニング手法で使用感だけを抽出するためには、抽出された経験の分類が必要であり、提案手法で構築した辞書はそのための重要な資源になると考えられる。

オノマトベに注目した研究もさかに行われており、オ

ノマトベの自動獲得 [13] や用例の獲得 [14], オノマトベの類型化 [15], [17], 音象徴を利用した評判分析 [19] など多岐にわたる。その中であって、商品レビューに出現するオノマトベとその商品が登録されているカテゴリの関係を調査した研究はあるが [20], 3.2 節で示した「オノマトベと商品の使用感に関する記述の関係」について調査した研究はなく、この点で本研究は新しいといえる。また、料理レシピやレストランに対して書かれたレビューからオノマトベを抽出し、レシピ、レストランの検索を支援する研究もある [16], [18]。これらの研究はテキスト中のオノマトベを直接用いているが、表 1 に示したようにオノマトベを含む文数は少ないためカバレッジに限界がある。本研究ではこの点を補うため、オノマトベを直接用いるのではなく、オノマトベが出現する文脈を利用しており、従来研究とはオノマトベの利用方法という点でも異なる。

7. おわりに

本稿では商品の使用感を記述した文をレビューから抽出する手法について述べた。オンラインショッピングサイトにおいて、使用感に関する記述へのアクセスが容易になることで次の効果が期待でき、その抽出は重要である。

顧客満足度の向上: 商品に対してユーザがいただくイメージをより正確なものにすることができるため。

新規ユーザの獲得: 総務省の情報通信白書によれば「実際に商品を見て買うことができない」点が、オンラインショッピングを利用しない主な理由の 1 つであるため。

提案手法は、オノマトベを含む文に出現しやすい表現は商品の使用感を記述する際に用いられやすいという仮説を用いており、この仮説は我々が行ったオノマトベと使用感を含まれる文の関係の調査結果に基づいている。この仮説に従い、提案手法ではオノマトベと単語の文中での共起の強さを商品レビューから算出し、得られた共起の強さをを用いてレビュー中の文が使用感を記述しているかどうかを判定する。実験の結果、 F_1 値で 65.9 ポイントの精度で使用感を記述した文を抽出できること、および上記の仮説が使用感を記述した文の抽出に有効であることが分かった。

エラー分析の結果、談話構造やモダリティを解析することで性能が改善されるという見通しを得た。そのため、より深い言語解析結果を取り込めるよう手法を拡張することは今後の課題と考えられる。また、日本語は他の言語に比べてオノマトベが豊富にあるため、そのオノマトベを利用した本手法も日本語に特化したものと思われるかもしれない。しかしながら、日本語において求めた単語とオノマトベの共起の強さは、単語を翻訳することで他言語にも利用可能だと考えられる。他言語のレビューに対して本手法を適用し、その性能を調査することも今後の課題である。

参考文献

[1] 尚学図書・言語研究所(編): 擬音語・擬態語の読本, 小学館(1991).

[2] Landis, R. and Koch, G.: The Measurement of Observer Agreement for Categorical Data, *Biometrics*, Vol.33, No.1, pp.159-174 (1977).

[3] 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).

[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回年次大会発表論文集*, pp.584-587 (2008).

[5] Pang, B. and Lee, L.: *Opinion mining and sentiment analysis*, Foundations and Trends in Information Retrieval (2008).

[6] Kanayama, H. and Nasukawa, T.: Textual Demand Analysis: Detection of Users' Wants and Needs from Opinions, *Proc. 22nd International Conference on Computational Linguistics - Volume 1*, pp.409-416 (2008).

[7] 乾 孝司, 梅澤佑介, 山本幹雄: 評価表現と文脈一貫性を利用した教師データ自動生成によるクレーム検出, *自然言語処理*, Vol.20, No.5, pp.683-706 (2013).

[8] 安藤まや, 関根 聡: レビューには何が書かれているのか?, *ALAGIN & NLP 若手の会合同シンポジウム(2013)*.

[9] Inui, K., Abe, S., Hara, K., Morita, H., Sao, C., Eguchi, M., Sumida, A., Murakami, K. and Matsuyoshi, S.: Experience Mining: Building a Large-Scale Database of Personal Experiences and Opinions from Web Documents, *Proc. 2008 IEEE/WIC/ACM International Conference on Web Intelligence*, pp.314-321 (2008).

[10] 倉島 健, 藤村 考, 奥田英範: 大規模テキストからの経験マイニング, *電子情報通信学会論文誌 D, 情報・システム*, Vol.92, No.3, pp.301-310 (2009).

[11] Park, K.-C., Jeong, Y. and Myaeng, S.-H.: Detecting Experiences from Weblogs, *Proc. 48th Annual Meeting of the Association for Computational Linguistics*, pp.1464-1472 (2010).

[12] Abe, S., Inui, K., Hara, K., Morita, H., Sao, C., Eguchi, M., Sumida, A., Murakami, K. and Matsuyoshi, S.: Mining personal experiences and opinions from Web documents, *Web Intelligence and Agent Systems*, Vol.9, No.2, pp.109-121 (2011).

[13] Okumura, M., Okumura, A. and Saito, S.: Automatic Construction of a Japanese Onomatopoeic Dictionary Using Text Data on the WWW, *Proc. 11th International Conference on Applications of Natural Language to Information Systems*, pp.209-215 (2006).

[14] Asaga, C., Mukarramah, Y. and Watanabe, C.: ONOMATOPEEDIA: Onomatopoeia Online Example Dictionary System Extracted from Data on the Web, *APWeb, Lecture Notes in Computer Science*, Vol.4976, pp.601-612, Springer (2008).

[15] 市岡健一, 福本文代: Web上から取得した共起頻度と音象徴によるオノマトペの自動分類, *電子情報通信学会論文誌 D, 情報・システム*, Vol.92, No.3, pp.428-438 (2009).

[16] ラートサムルアイパンカンウィパー, 渡辺知恵美, 中村聡史: オノマトペロリ: オノマトペを利用した料理推薦システム, *情報処理学会研究報告 DD, [デジタル・ドキュメント]*, pp.1-7 (2009).

[17] 鍛冶伸裕, 宇野良子, 喜連川優: ウェブテキストにもとづくオノマトペのカテゴリ化とその工学的支援, *じんもんこん 2010 論文集*, Vol.2010, No.15, pp.11-18 (2010).

[18] Kato, A., Fukazawa, Y., Sato, T. and Mori, T.: Extraction of Onomatopoeia Used for Foods from Food Reviews

and Its Application to Restaurant Search, *Proc. 21st International Conference Companion on World Wide Web*, pp.719-728 (2012).

[19] Igarashi, T., Sasano, R., Takamura, H. and Okumura, M.: Use of Sound Symbolism in Sentiment Classification, *Journal of Natural Language Processing*, Vol.20, No.2, pp.183-200 (2013).

[20] 内田ゆず, 長谷川大, 荒木健治, 米山 淳: 商品レビュー文におけるオノマトペと商品カテゴリの関係についての予備調査, *言語処理学会第19回年次大会発表論文集*, pp.810-813 (2013).



新里 圭司 (正会員)

2006年北陸先端科学技術大学院大学情報科学研究科博士後期課程修了。博士(情報科学)。京都大学大学院情報科学研究科特任助教, 特定研究員を経て, 2011年から楽天技術研究所。自然言語処理, 特に, 知識獲得, 情報抽出, テキストマイニングの研究に従事。



益子 宗 (正会員)

2002年IPA未踏ソフトウェア創造事業開発代表者。2006~2008年日本学術振興会特別研究員。2008年筑波大学大学院システム情報工学研究科博士課程修了。博士(工学)。同年楽天技術研究所入所。2011年より筑波大学大学院非常勤講師を兼任。エンタテインメントコンピューティング, HCI, CGアニメーション等の研究に従事。



関根 聡 (正会員)

1987年東京工業大学応用物理学科卒業。1987年松下電器東京研究所入社。1992年に英国マンチェスター工科大学計算言語学科で理学修士(MSc)を取得。1994年ニューヨーク大学研究助手。1998年同大学でPh.D.を所得, 研究助教授就任。2007年同大学准教授。2010年より2014年楽天技術研究所ニューヨーク所長を兼任。自然言語処理, 特に, 情報抽出, 知識獲得, 言語解析の研究に従事。現在, 言語処理学会理事, AFNLP役員等を務める。