

# オノマトペを利用した商品の使用感の抽出

新里 圭司 関根 聡

楽天技術研究所

{keiji.shinzato, satoshi.b.sekine}@mail.rakuten.com

## 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時のように、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問題がある。そのため、ユーザが抱く商品のイメージと実際に届く商品の間に、質感や食感等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<sup>1</sup>において、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」点が挙げられており、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上で重要である。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。例えば、以下の文はチーズケーキの使用感について述べていると見なせる。

- 口に入れたらとろっとして甘酸っぱく、私好みのお味でした。

味わいの他にも、質感、香り、効果、効能、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、ユーザに使用感を伝えるため商品のレビューを公開<sup>2</sup>している。しかし、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レビューには含まれるため、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、既存ユーザの負担を軽

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

<sup>2</sup>みんなのレビュー (<http://review.rakuten.co.jp/>) など。

表 1: オノマトペを含む文と商品の使用感の関係

カテゴリ	使用感についての		オノマトペ ではない	正解率 [%]
	記述である	記述でない		
ワンピース	43	4	3	91.4
シャンプー	43	5	2	89.6
チーズケーキ	39	10	1	79.6
合計	125	19	6	86.8

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、オノマトペが出現する文脈を学習し、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。オノマトペ<sup>3</sup>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わかった。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、使用感の記述に用いられる単語をレビューから収集する。そして、収集した単語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。

## 2 オノマトペを含む文と使用感の関係

商品レビューにおいて、オノマトペを含む文のうち、どのくらいが使用感を記述したものであるのかを調査した。楽天が運営する「みんなのレビュー」から、ワンピース、シャンプー、チーズケーキカテゴリの商品に対して書かれたレビューを収集し、そこからオノマトペを含む文を無作為に50文ずつ、計150文選出した。レビューからオノマトペを含む文を抽出する方法については後述の3.1節を参照されたい。

調査結果を表1に示す。表中の「オノマトペではない」は、下の文のように、文字列としては同じだがオノマトペではない表現にマッチした事例の数を表す。

- 探し続けていた商品 にやっと 出会えました。

<sup>3</sup>本稿では擬音語・擬態語をオノマトペと呼ぶ。

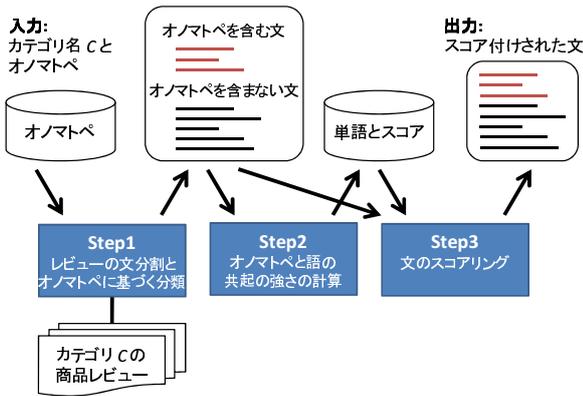


図 1: 提案手法の概要

「オノマトペではない」の数を考慮せずに正解率を計算した結果、オノマトペを含む文の 86.8%が商品の使用感を記述しているものであった。この 3 カテゴリについて、オノマトペの有無を区別することなく、レビュー内の文が使用感を記述したものであるかどうかを調べたところ、平均で 42.8%であった。この数字と比べると、オノマトペを含む文は使用感を述べやすいことがわかる。

オノマトペを含んでいるにも関わらず使用感を記述していない文の多くは「ちょっと」を含むものであり、「ちょっと高い」など価格に関する表現が多かった。

### 3 提案手法

提案手法の概要を図 1 に示す。本手法では、2 節で得られた知見を発展させ、「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、レビューから使用感について書かれた文を抽出する。具体的には、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。

#### 3.1 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の商品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、「。」「!」「?」「♪」「…」「(笑)」「(怒)」「(泣)」を手がかりに行った。

表 2: 語とそのスコアの例

チーズケーキ		シャンプルー		ワンピース	
スコア	単語	スコア	単語	スコア	単語
1.991	感	1.881	ごわごわ	1.941	ゴワゴワ
1.618	重い	1.881	かんじ	1.941	ぶかぶか
1.568	後味	1.881	めちゃくちゃ	1.941	ニット
1.446	口当たり	1.881	ギシギシ	1.941	めちゃめちゃ
:	:	:	:	:	:
0.005	最高	0.001	感想	0.001	ゆう
0.003	十分	0.001	良い	0.000	加工
-0.002	売る	-0.004	すき	-0.001	切る
-0.008	分	-0.004	優しい	-0.001	シルク
:	:	:	:	:	:
-1.681	機会	-1.880	送料	-1.696	手数料
-2.095	親戚	-2.104	無料	-1.704	問い合わせ
-2.133	中元	-2.380	安値	-1.909	キャンセル
-2.195	物産	-2.454	親切	-1.951	雑誌

続いて得られた文をオノマトペを含むものと、含まないものに分類する。入力として与えるカテゴリごとにオノマトペを準備しても構わないが、本手法では全カテゴリにおいて共通のオノマトペを用いた。オノマトペは「擬音語・擬態語の読本」[6]の見出し語になっている 849 表現、および各表現をカタカナに変換したものを利用した。これは、(1) 文献[6]にはひらがな表記しか記載されていない、(2) オノマトペはひらがな・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<sup>4</sup>。

#### 3.2 オノマトペと語の共起の強さの計算

続いて、上述の仮説に基づき、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<sup>5</sup>を対象とした。各単語とオノマトペの共起の強さには以下の式で求められる相互情報量を用いる。

$$\text{Score}(w) = \log \frac{p(w, O)}{p(w) \cdot p(O)}$$

ここで  $p(w, O)$  は任意のオノマトペを含む文に語  $w$  が出現する確率、 $p(w)$  は語  $w$  を含む文が出現する確率、 $p(O)$  は任意のオノマトペを含む文が出現する確率である。スコアを計算する際、個々のオノマトペを区別していないことに注意されたい。

表 2 にチーズケーキ、シャンプルー、ワンピースカテゴリに対する単語とそのスコアを示す。「後味」「口当たり」「めちゃくちゃ」「ごわごわ」「ぶかぶか」など使用感の記述に用いられるだろう単語には高いスコアが、反対に「中元」「送料」「無料」「キャンセル」など、

<sup>4</sup>実際はカタカナに変換することでオノマトペでない語にマッチしてしまう事例があった。例えば笑いを浮かべる様を表す「にっこ」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。

<sup>5</sup>ただし、細分類が非自立、もしくは接尾の語は除く。さらに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。

表 3: 評価に用いたレビューデータ. 括弧内の数字はオノマトペを含む文数を表す.

カテゴリ	レビュー数	文数
ワンピース	468,059	1,786,651 (234,217)
シャンプー	249,095	788,619 (111,417)
チーズケーキ	44,906	141,131 (13,413)

表 4: 比較結果

カテゴリ	ベースライン			提案手法		
	P	R	F <sub>1</sub>	P	R	F <sub>1</sub>
ワンピース	0.567	0.723	0.636	0.678	0.851	0.755
シャンプー	0.470	0.816	0.596	0.667	0.816	0.734
チーズケーキ	0.513	0.908	0.656	0.667	0.805	0.729

使用感以外の情報を記述する際に用いられるだろう単語には低いスコアが与えられていることがわかる.

### 3.3 文のスコアリング

最後に前節で計算したスコアを用いて文  $s$  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.

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

## 4 評価実験

楽天が運営する「みんなのレビュー」のデータを使って評価実験を行った. 具体的には, 2012年1月から2013年10月までにワンピース, シャンプー, チーズケーキカテゴリの商品に寄せられたレビューを対象に行った. 各カテゴリのレビュー数, 文数, オノマトペを含む文数を表3に示す. 提案手法をカテゴリごとに適用し<sup>6</sup>, 得られた結果から無作為に200文を選び, 文が商品の使用感を記述しているかどうか人手で評価した. 評価は第一著者と, 本論文とは関係のない被験者の2人で行った. 被験者間の  $\kappa$  統計量 [2] は 0.774 であり, これは good agreement とされる値である. 両被験者が共に正解と判定したものをのみを正解とした.

評価実験の結果を図2に, チーズケーキカテゴリのレビュー文に対して計算されたスコアを表5に示す. 図の横軸はスコアに従って文を降順にソートした時の順位, 縦軸は当該順位までの文を出力とした時の正解率を表す. ワンピース, シャンプー, チーズケーキいずれのカテゴリにおいても, グラフの線が右下がりになっていることから, 本稿で提案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ことがわかる.

<sup>6</sup>形態素解析には MeCab (<http://code.google.com/p/mecab/>) を用いた.

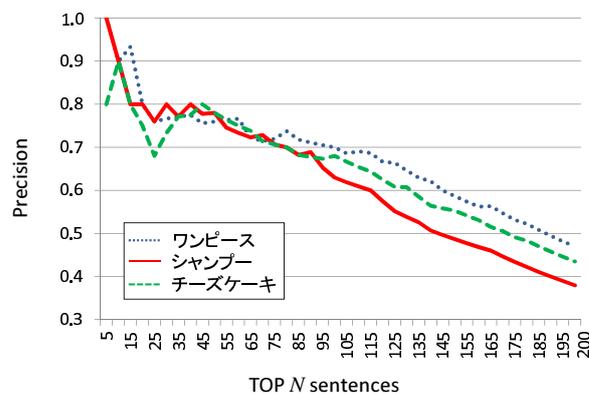


図 2: 評価結果

次に, 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ベースラインとし, 提案手法との比較を行った. 評価表現としては, 小林ら [5] の辞書および東山ら [7] の辞書<sup>7</sup>を用いた. 提案手法はスコアが0以上の文を出力とした. 比較実験の結果を表4に示す. 表より提案手法はベースラインに比べ精度が高いことがわかる. その一方で, チーズケーキカテゴリに関してはベースラインの方が再現率が高い. ベースラインで獲得できて, 提案手法でできなかった文を以下に示す (下線は評価表現).

- たまたま見つけて購入しましたがとても 美味しい!
- 相手の方には, こんなチーズケーキは食べたことないと 絶賛 でした.

どちらの文もオノマトペと共起しにくい語(「購入」「絶賛»)が含まれており, これらのスコアが文全体のスコアを下げる原因となっていた. このような文を獲得するためには, 文よりも細かい単位を解析対象とする, ベースラインで用いた評価表現辞書の情報を利用する, などが考えられる.

## 5 関連研究

安藤ら [4] 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を調査した. 本研究で注目している「使用感」は, 彼女ら定義した「判断」と近く, 本研究はその情報の自動抽出手法とみなすことができる.

レビューから商品の使用感を抽出する本研究は評判抽出の一種と考えられる. 教師データを用いない評判

<sup>7</sup>共に <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.

表 5: 文とそのスコアの例 (チーズケーキ)

スコア	文
9.27	チーズケーキのとろけるような食感と濃厚な味わい、ブラウニーのナッツの香ばしさとチョコの甘みが濃縮され、どちらもおいしかったです。
5.38	口に入れたらとろっとして甘酸っぱく、私好みのお味でした。
5.06	口の中に入れると香りだけ残してすーっと溶けていく感じ。
-4.44	北海道物産展で購入したことがあり、とっても大好きです。
-5.05	いつもクリスマスプレゼントを贈ってくれる義兄のお母さんにお中元で贈ってみました。
-5.61	お店のオンラインショッピングでも物産展でも何度も購入し、味はわかっています。

抽出の多くが評価表現辞書に基づくのに対し [3], 本手法はオノマトペの出現する文脈を学習し, その結果得られた単語を評判 (正確には商品の使用感) 抽出に利用しており, この点が従来手法と異なる。

また近年はオノマトペに注目した研究も盛んに行われている。例えば人工知能学会全国大会では, 2011 年より 3 年間にわたりオーガナイズドセッション「オノマトペの利活用」が開催された<sup>8</sup>。言語処理分野でも, オノマトペの音象徴を利用した評判分析 [1] や, 大規模コーパスを用いたオノマトペの意味的な分類 [9], オノマトペと商品カテゴリの関係の調査 [8] など多岐にわたる。しかし, オノマトペの出現する文脈を商品の使用感の抽出に利用した研究は, 筆者らの知る限りない。

最後に教師なし機械学習の観点から本手法を考えてみたい。本手法はオノマトペの出現の有無を手掛かりに商品の使用感を記述した文を抽出する。オノマトペは辞書に記載されているものを使っており, それを生テキストに適用することで擬似教師データを構築している。この点で *distant supervision* の一種と見みなすことができる。*distant supervision* の枠組みでは, シードおよび最終的に出力されるデータの型が一致している。例えば, 固有表現獲得であれば, シードも出力されるデータも固有表現であるし, 関係抽出であればシードおよびその出力はある関係 (例えば, 企業-創業者) にある表現のペアである。これに対し, 本手法はシード (オノマトペ) および最終的に出力されるデータ (使用感を記述した文) の型が異なっている。両者の型が違うにも関わらず, ある程度の精度を達成できた点は興味深い。

## 6 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。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る。

**顧客満足度の向上:** 商品に対してユーザが抱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要な原因の 1 つであるため。

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, オノマトペと文中で共起しやすい語を商品レビューから獲得し, 得られた語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,  $F_1$  値で約 74 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わかった。

今後の課題としては, 今回は限られたカテゴリのレビューを対象に実験を行ったため, より多くのカテゴリに対して手法の有効性を確認することが挙げられる。

## 参考文献

- [1] Takuma Igarashi, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura. Use of sound symbolism in sentiment classification. *Journal of Natural Language Processing*, 20(2):183–200, 2013.
- [2] Richard Landis and Gary Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.
- [3] Bo Pang and Lillian Lee. *Opinion mining and sentiment analysis*. Foundations and Trends in Information Retrieval, 2008.
- [4] 安藤まや, 関根聡. レビューには何が書かれているのか? In *ALAGIN & NLP 若手の会合同シンポジウム*, 2013.
- [5] 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*, 12(3):203–222, 2005.
- [6] 尚学図書・言語研究所 (編). *擬音語・擬態語の読本*. 小学館, 1991.
- [7] 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. In *言語処理学会第 14 回年次大会発表論文集*, pages 584–587, 2008.
- [8] 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. In *言語処理学会第 19 回年次大会発表論文集*, pages 810–813, 2013.
- [9] 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*, 2010(15):11–18, 2010.

<sup>8</sup>[http://www.ai-gakkai.or.jp/pse\\_201501-1/](http://www.ai-gakkai.or.jp/pse_201501-1/)