

# 商品レビュー文に対する極性分析を用いた重要文章抽出

2016SC081 杉山 倅汰 2016SC090 富田 優人

指導教員:河野 浩之

## 1 はじめに

近年、レビューサイトの普及により、商品の比較が容易になり、より自分の嗜好に合うものを多くの商品の中から選択可能になっている。総務省の情報通信白書 [1] によると、インターネットショッピングサイトを対象とし、買い物をする際にレビュー文をどの程度参考にするのかを尋ねた結果、どの年代でも「かなり参考にする」、「まあ参考にする」を合わせると6割以上となり、多くのユーザがレビュー文を読むことでその商品について深く理解し、購入するかどうか判断する根拠のひとつとしていることが分かる。このようにレビュー文は重要な情報源であり、他者によって書かれた感想からは、メーカーの提供する商品の説明だけでは読み取ることのできない特徴や具体的な使用感や体験談を知ることができる。しかし、ユーザの関心の高い商品においては、コメントが膨大な量となり全て読むには時間と労力を必要とし、一部の閲覧では得られる情報に偏りが生じる恐れがある。秋山ら [3] によると、評価の高いレビュー文の分析を行った結果、ポジティブ、ネガティブ両方の印象が書かれているレビュー文は参考とされる割合が高かった。一方で、役立つと評価する数が増えるにしたがって、ポジティブな意見の割合は減少し、ネガティブな意見の割合は増加しており、レビュー文を見るユーザはネガティブな印象をより重視していた。

そこで、本研究では、レビュー文を閲覧する負担を解消するためにポジティブ重要文とネガティブ重要文の抽出システムを提案する。極性分析により商品レビュー文をポジティブ文とネガティブ文に分類する。それぞれの特徴語を基に空間ベクトルモデルを用いてコサイン類似度を測定し、各レビュー文の中からポジティブ重要文とネガティブ重要文を抽出することで、商品进行评估する意見の全体像をより効率的に提示することを目指す。

本稿は、5章で構成されており2章では極性分析を用いた商品レビューに関する先行研究、3章では極性分析を用いた重要文章抽出システムの提案、4章では実験、5章ではむすびを示す。

## 2 レビュー文に対する極性分析の先行研究

Amazon、楽天、価格.comといったレビューサイトの普及により、商品のレビューの情報を効率良く獲得するための要約や視覚化、又レビュー文を利用した推薦に関する研究が数多く行われている。今回参考にした2つの先行研究を内容、ツール・技術の観点から比較した結果を表1に示す。

吉田ら [7] はレビュー文の中から、賛否の分かれる属性は利用者の印象に強く残ったもので、商品の重要な特徴であると定義した。又、提案手法では、肯定表現や否定表現を表す評価表現を含む文をレビュー文から抽出し、評価の対象となった各属性を商品の特徴語候補とした。

平山ら [5] は、評価属性ごとにポジティブな評価数とネガティブな評価数を集計することで全体の評価を統計として提示し、根拠を伴う評価文を統計と共に提示することにより詳細な情報を読み取れるシステムを目指した。根拠を伴う評価文の抽出は、「ので」、「から」が含まれているかで判定したが、根拠を伴った評価文は、ほぼ抽出できなかった。

上拾石ら [4] は、レビュー文の中から、ユーザの商品に対する意見を表す言葉を意見モデル「対象物、属性表現、評価表現」として抽出し、評価表現と属性表現をそれぞれ出現頻度と共起率でスコアリングすることで多くの商品購入者が述べている意見を収集した。そのスコアリング結果を円グラフ表示やテーブルリスト表示をさせることでユーザの購入意思決定の支援をするシステムを提案した。しかし、レビューサイトではネガティブな意見よりポジティブな意見が圧倒的に多いため、このシステムではポジティブな意見に偏ってしまう傾向となり、ユーザが重視しているネガティブな情報を十分に得ることができない。

## 3 レビュー文に対する極性分析を用いた重要文章抽出の提案

本章では、本研究の極性分析を用いた重要文章抽出システムの提案について示す。3.1節では極性分析を用いた重要文章抽出システムの概要、3.2節では極性分析について説明する。

### 3.1 極性分析を用いた重要文章を抽出するアルゴリズムの概要

極性分析を用いた重要文章抽出のアルゴリズムのフローチャートを図1に示す。その中で示されている(1)から(7)を以下で詳細に述べる。

(1) 対象とする商品のレビュー文から本文のみを収集し、重要文の候補として1人分ずつデータベースに格納する。

(2) (1)と同じレビュー文から本文のみをまとめ、レビュー文集合として文単位に分割する。

(3) 文分割したレビュー文集合を別のデータベースに格納する。

(4) 文分割したレビュー文集合に、形態素解析と構文解析を行う。

(5) (4)を極性分析し、ポジティブ文集合、ネガティブ文集合、ニュートラル文集合に分ける。

(6) ポジティブ文集合とネガティブ文集合におけるキーワードの重要度をTF-IDFによって計算し、それらを成分とするキーワードベクトルを作成する。その際キーワードはストップワードを取り除き、名詞・動詞・形容詞・形容動詞とする。

(7) 重要文候補としてデータベースに格納しているレビュー文(1)にポジティブキーワードベクトルとネガティブキーワードベクトルのコサイン類似度を算出する。ス

表 1 極性分析の先行研究の比較

研究者	内容	ツール・技術
上拾石ら [4]	評価属性と属性表現を解析し、円グラフで視覚化	Chasen, 共起率の計算
平山ら [5]	「ので」、「から」を含む文章を根拠を伴う評価文として抽出	MeCab, CaboCha, 青木ら [2] が作成した極性反転子辞書を含めた単語極性辞書
吉田ら [7]	賛否の分かれる属性は、商品の重要な特徴と定義し、特徴語を抽出	MeCab, 日本語評価極性辞書, ナイーブベース

コアが最も高かったものを、それぞれポジティブ重要文とネガティブ重要文として提示する。

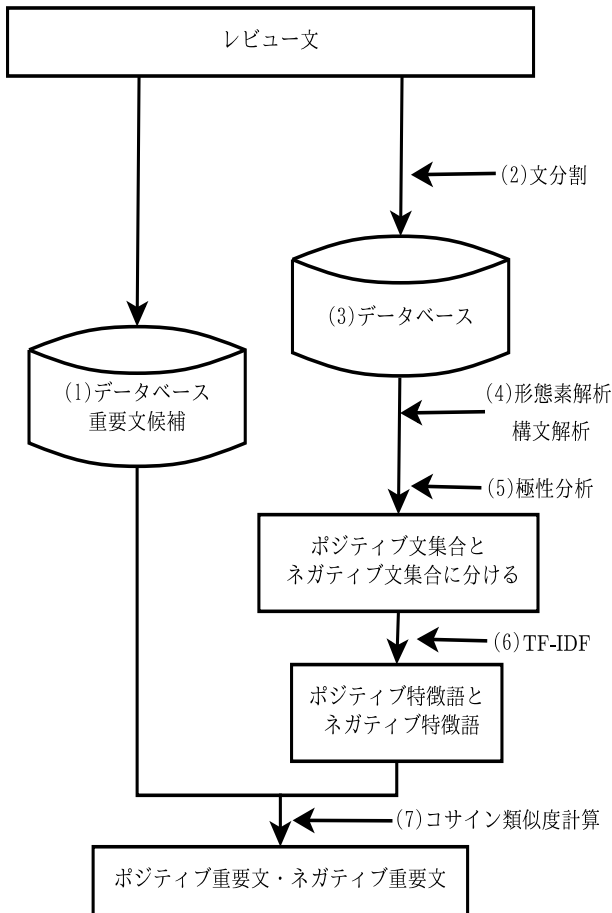


図 1 極性分析を用いた重要文章を抽出するアルゴリズム

### 3.2 極性分析

極性分析に基づくテキスト分類には、東北大学乾・岡崎教室の日本語評価極性辞書<sup>\*1</sup>の用語編と名詞編を評価辞書として利用する。評価表現は、「重い」ことが良いと評価されるものと、悪いと評価されるものがあるなど、対象となる商品によって評価の異なる表現がある。本研究で対象とするイヤフォンの場合は、「重い」がネガティブ、「軽い」がポジティブな評価となる。他にも、「(音が)切れる」など、他の商品では評価表現とならない表現がネガティブな評価表現となる。又、辞書には音質がポジティブ

表現として登録されるなど、商品の評価には適さない評価表現が含まれているためにそれらを削除する必要がある。商品の特徴を考慮し、表 2 に示す極性辞書の編集を行う。

表 2 極性辞書の編集

削除した単語	極性	追加した単語	極性
音質	ポジティブ	軽い	ポジティブ
再生	ポジティブ	フィット	ポジティブ
値段	ネガティブ	(音が)切れる	ネガティブ
価格	ネガティブ	途切れる	ネガティブ
		飛ぶ	ネガティブ
		漏れる	ネガティブ
		重い	ネガティブ

又、「ない」、「ません」、「ありません」、「にくい」は、前の評価を打ち消すので、評価表現の後にこれらの単語がある場合は、極性を反転させる。文はスコアで、+がポジティブ、-がネガティブ、0がニュートラルと分類する。スコアの計算は、辞書内の評価表現にポジティブなら1、ネガティブなら-1の極性値を付け、それにマッチすると1あるいは-1の極性値が文に加えられる。極性値の合計を文のスコアとする。例えば、「手頃な価格でトータルで満足しています」は「手頃」、「満足」がポジティブな単語と評価されるので文スコアは+2でポジティブ文に分類される。ポジティブ文、ネガティブ文のみ TF-IDF、コサイン類似度を計算する。

## 4 イヤフォンのレビュー文に対する極性分析の実験

本章では実装環境と実験の結果、考察について説明する。4.1 節では実装環境について、4.2 節では極性分析の実験と考察について、4.3 節では特徴語の結果と考察について、4.4 節では重要文抽出の結果と考察について示す。

### 4.1 実装環境

今回、実装環境は表 3 に示す。データ収集には、楽天レビューからデータを取得し、CSV ファイルに保存できるフリーソフト「らくれびゅ」を用いる。データベースには SQLite を用いる。プログラミング言語は、ライブラリが

<sup>\*1</sup> <http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary>

表 3 実装環境

OS	Windows10, Ubuntu14.04
メモリ	2GB
プログラミング言語	Python3.4.3

豊富な Python を使用する。又、形態素解析には MeCab を構文解析には CaboCha を使用する。

#### 4.2 極性分析の実験と考察

インターネット通販サイト「【楽天市場】 口コミ・レビューで人気商品を探そう！ | みんなのレビュー」より、3つのイヤホン<sup>\*2\*3\*4</sup>のレビュー文を選び、提案手法を適用した。極性分析より得られた、ポジティブ文、ネガティブ文、ニュートラル文の数を表4に示す。3商品どれもが、ポジティブ文は40%以上を占めるのに対し、ネガティブ文は20%以下と少なく、レビューにはポジティブな意見が多いことが分かる。

表 4 極性分析によって分類されたポジティブ文・ネガティブ文・ニュートラル文の数

商品	1	2	3
コメントの総数	304	411	326
文分割した総数	1056	1380	812
ポジティブ文	574(0.54)	613(0.44)	355(0.44)
ネガティブ文	144(0.14)	227(0.16)	154(0.19)
ニュートラル文	338(0.32)	540(0.40)	303(0.37)

( ) 内は、文分割した総数に対する割合を示す。

極性分析は、辞書の編集を行ったことで、図2のようにイヤホンの特性に合わせた分析が行えている。

<p>No. 162          通勤や運動時に、右左の両方ペアリングするで使用すると、なぜか音が途切れる頻度が多いたです --&gt;          positive: []          negative: [['途切れる', 32]]          文のスコア: -1</p>
--

図 2 辞書の編集により適切な極性分析ができた例

極性判定された文を1つ1つ調べたところ、正しく判定されない要因は以下の3つであった。

1. 表記ゆれ
2. 商品の評価尺度と辞書の評価尺度の不一致
3. 極性反転の課題

1の表記ゆれの例をあげると「良い」、「いい」はポジティブ表現として判定されるが「よい」は判定されない。「問題ない」はポジティブに判定されるが、「問題無い」は「無い」では極性の反転が行われず、「問題」だけが判定されネガティブになる。レビュー文は文章の自由度が高いために表記ゆれが大きな問題となる。表記ゆれをま

とめる解析器を使用するなど、手法の改善が必要であることが分かった。

2は、レビュー文では評価表現を強調するために使われることが多い「非常」や色を表すことが多いブラックがネガティブに登録されているために正しく判定されてなかった。他にも、辞書には「プレゼント」、「感度」、「方法」などがポジティブに、「部品」、「電波」、「忘れる」などがネガティブに登録され、「安い」、「弱い」、「申し分ない」が登録されていないなど、商品の評価と辞書の評価尺度の不一致が多くみられた。商品レビューの分析に特化した辞書の編集を行うことが、極性分析の精度向上につながると思われる。

3は、打ち消し語と評価表現の間に2つ以上の単語が入ると反転できず手法の改善が必要である。

#### 4.3 特徴語の結果と考察

TF-IDF で求められた上位10位の特徴語を表5, 6に示す。

表 5 ポジティブ文の特徴語上位10位

商品1	商品2	商品3
音質	音質	値段
簡単	簡単	満足
ペア	満足	音質
リング	問題	問題
満足	接続	そう
フィット	商品	感じ
ケース	使用	コード
購入	ペア	丈夫
接続	リング	購入
商品	購入	商品

表 6 ネガティブ文の特徴語上位10位

商品1	商品2	商品3
残念	残念	断線
充電	こと	購入
勝手	使用	コード
接続	部分	そう
こと	購入	音量
使用	漏れ	よう
ケース	電源	音質
ペア	感じ	もの
大変	接続	使用
リング	コード	片方

「音質」、「充電」、「接続」、「断線」などイヤホンに重要な機能を表す単語が選ばれている。「ペアリング」が「ペア」と「リング」に分割されており、MeCabに辞書を追加する又は名詞と名詞は1つにまとめるなど、改善を図る必要がある。

\*2[https://review.rakuten.co.jp/item/1/337900\\_10000222/1.1/](https://review.rakuten.co.jp/item/1/337900_10000222/1.1/)

\*3[https://review.rakuten.co.jp/item/1/319035\\_10002281/1.1/](https://review.rakuten.co.jp/item/1/319035_10002281/1.1/)

\*4[https://review.rakuten.co.jp/item/1/263585\\_10052403/1.1/](https://review.rakuten.co.jp/item/1/263585_10052403/1.1/)

#### 4.4 重要文抽出の結果と考察

コサイン類似度で抽出されたレビュー文の一部を図 3, 4, 5 に示す. コサイン類似度は, 表 7 に示す.

表 7 各商品の重要文のコサイン類似度

商品	ポジティブ重要文	ネガティブ重要文
1	0.48	0.45
2	0.52	0.42
3	0.45	0.69

18 : 0.4837777952107474 : … ペアリングは自動で簡単に接続でき、音声でもペアリング開始、終了をお知らせしてくれました。操作方法もタッチ式で簡単でした。音質は詳しくわかりませんが、クリアな音質で高音と低音のバランスが良く満足できました。…

図 3 商品 1 の類似度 1 位のポジティブ文の抜粋

28 : 0.4491634904350512 : … 通勤や運動時に、右左の両方ペアリングして使用すると、なぜか音が途切れる頻度が多かったです。あと、僕の左耳にはフィットしないようで、よく外れてしまいます。…

図 4 商品 1 の類似度 1 位のネガティブ文の抜粋

重要文には TF-IDF で上位に選ばれたイヤホン機能のポジティブな意見やネガティブな意見が記載されている。しかし、商品 1 のネガティブ重要文では、特徴語 2 位の充電の記載がないものが類似度 1 位として抽出され、類似度が 2 番目に高かったレビュー文には充電のネガティブな意見が記載されていた。「ペアリング」が「ペア」と「リング」に分割され、それぞれに TF-IDF 値がカウントされたことが影響を与えたと考えられるが、コサイン類似度が最も高い文が必ずしも商品の特徴を最も表す結果とはならなかった。

商品 2 のネガティブ重要文 1 位は対象商品以外の記載が 686 文字中 170 文字含まれていた。レビューには、以前使っていた商品の評価や購入に至るまでの経緯や気持ちなど、対象商品以外のデータが含まれることがあるが、文章の中からこれらを削除することは難しい。選ばれた重要文のコサイン類似度は 0.42 から 0.69 である。当初は、コサイン類似度の最も高い文を重要文として提示することを目指していたが、今回の結果から商品に対する評価の全体像を 1 つのレビュー文だけで提示することは適切ではないと考える。しかし、本システムで抽出されたポジティブ重要文とネガティブ重要文を上位から順に並べて掲示することで、少数のレビュー文の閲覧でも商品のポジティブな評価とネガティブな評価を偏ることなく把握することができるかと判断する。この手法が、閲覧する負担の軽減と商品購入意思決定の支援となるものであるかは、今後アンケートを通して検証していきたい。

291 : 0.4210279369879134 : … そして最大の問題は充電するケース。右の方はまだいいのですが左のしまう方は緩いのかケースに入れて持ち歩いている間に浮いてしまいずっと充電が中途半端にされてる状態!…

図 5 商品 1 の類似度 2 位のネガティブ文の抜粋

## 5 むすび

本研究では、ユーザが多数のレビューコメントから、効率よく商品に対する評価を得るために、ポジティブ重要文とネガティブ重要文を提示する手法を提案した。対象とするイヤホンの評価情報を整理し、商品に合わせた極性分析を行うことで、特徴語として抽出されたイヤホン機能のポジティブな評価、ネガティブな評価が記載されたレビュー文を重要文として抽出できた。しかし、重要文のコサイン類似度は 0.42 から 0.69 であり、重要文のみを提示するより、重要文を上位から提示することで、少数のコメントの閲覧でも偏ることなく商品の評価を把握できると考える。今後の課題として、アンケートを通した有効性の検証、極性分析の精度向上が挙げられる。極性分析は、表記ゆれへの対策や、極性辞書のさらなる編集が精度向上につながるかと考える。又、コサイン類似度が高くても短い文よりコサイン類似度が低くても長い文の方が商品のことをより明確に示すことが出来ると考える。

## 参考文献

- [1] 総務省, “平成 28 年版情報通信白書 | 情報資産 (レビュー (口コミ等)),” <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/html/nc114230.html>, 参照, Dec, 5, 2019.
- [2] 青木 伸也, 湯本 高行, 角谷 和俊, 新居 学, 高橋 豊, “論点に対する極性に注目したニュース記事からの編集意図の抽出手法,” 第 149 回 DBS 研究会, DBS-2A, 2009.
- [3] 秋山 和寛, 熊本 忠彦, 灘本 明代, “レビューの印象に基づいた評価の高いレビューの分析,” DEIM Forum 2019, C1-5, 2019.
- [4] 上拾石 博紀, 佐藤 真, 赤石 美奈, “カスタマーレビューを用いた商品利用者の意見分析,” 第 27 回人工知能学会全国大会論文集, 1C5-3, 2013.
- [5] 平山 拓央, 湯本 高行, 新居 学, 高橋 豊, “属性評価モデルに基づく商品評価の抽出と提示,” DEIM Forum 2011, F2-5, 2011.
- [6] 柳井 孝介, 庄司 美沙, “Python で動かして学ぶ 自然言語処理入門,” pp.32-223, 株式会社 翔泳社, 東京, 2019.
- [7] 吉田 朋史, 北山 大輔, “商品推薦のための商品レビューの極性分析に基づく特徴語抽出手法,” DEIM Forum 2015, B2-1, 2015.