

ポジティブ語・ネガティブ語抽出による 不当なレビュー評価の是正方法に関する一考察

廣瀬研究室 4年 C1172313 佐藤利紀

概要

商品レビューには記述者の人為的なミスによって評価値とレビューの記述内容が食い違っている場合がある。これはレビュー数が少ない商品の評価値に大きく作用し、適切な評価が反映されず、消費者の購買意思決定や企業の商品ブランドに強い影響を及ぼす。そこで本研究では、その商品カテゴリの高評価のレビュー群に現れる「ポジティブ語」、低評価のレビュー群に現れる「ネガティブ語」を抽出し、評価値とレビュー文に食い違いのある不当なレビューの検出を図る。そして不当なレビュー評価の是正を目指す。

1 はじめに

近年、著しい情報技術の発展によって、AIによる作業の効率化がなされている。しかし、どんな作業にも人手が加わり、そこに人為的なミスが生じる。その一例が商品レビューである。商品レビューはECサイトにおいて、その商品に対する評価を評価値とレビュー文で表したものである。商品レビューの集計や整理は自動で行われるが、商品レビュー自体を記述や投稿は人手で行われる。そのため、評価値とレビュー文の内容が食い違う可能性がある。たとえば、本の商品レビューにおいて、評価値が低いにもかかわらず「面白い」というレビュー文が書かれているなどの不一致が考えられる。これは寄せられたレビューが少ない商品においては問題となる。ECサイト Amazon¹における商品の評価値は寄せられた商品レビューの評価値の加重平均となっている [1]。これはレビューの多寡によって商品レビュー1件あたりの評価の重みが異なるという。そのため、寄せられたレビューが少ない商品の場合、商品レビュー1件がその商品の評価値に及ぼす影響が大きいと考えられる。その商品レビューの中に評価値と文章が不一致のレビューがあった場合、商品の評価値が正当に反映されず、消費者の購買意思決定の妨げや企業の商品ブランド低下などにつながる恐れがある。

そこで本研究では商品カテゴリごとの商品レビューをテキストマイニングしこの問題の解決を図る。具体的には、評価値ごとの主要な単語を見出し、それらの単語を評価値によってその商品カテゴリのネガティブ語とポジティブ語に分類する。これを用い高評価のレビューに含まれるネガティブ語もしくは低評価レビューに含まれるポジティブ語を見出す手法を提案する。そして不当なレビューが混在した商品の評価の是正を目指す。

2 テキストマイニングと関連研究

テキストマイニングとは形式化されていない大量のテキストデータを分析して有用な情報を抽出することである [2]。商品レビューのような大量のテキストデータの分析にこのテキストマイニングの技術が用いられる [3]。商品レビュー分析に使われるテキストマイニングでは、主に形態素解析²によって単語の出現頻度の偏りなどから意見の傾向を可視化する研究がなされている [4]。

商品レビューの分析に関する研究はいくつかある。河野ら [5]の研究では対象、属性、評価表現の単語とその係り受け情報を用いて評価情報を抽出している。那須川ら [6]の研究では「けれども」や「しかし」といった接続表現の周辺に出現する単語を抽出し、好評もしくは不評の極性を付与している。

しかし、河野らの研究では対象、属性、評価表現が完全に揃った文を対象としており、不完全な日本語の表現は想定されていない。たとえば「パソコンの価格は安い」など対象「パソコン」、属性「価格」、評価表現「安い」が揃った文は対象とされているが、「これは安い」といった対象と属性が含まれていない文はこの方法では抽出できない。一方、那須川らの研究では人的過誤による好評、不評の解釈については想定されていない。そこで本研究ではレビューに含まれる単語に焦点をあて、その商品カテゴリ内における高評価、低評価それぞれの特徴となる単語の抽出を試みる。また、不当なレビューに着目した分析、考察を行うことで人為的なミスによって生じる不正確な商品評価の是正を目指す。

¹<https://www.amazon.co.jp/>

²品詞情報や文法などをもとに文章を単語に分解すること。

3 商品レビューの分析

3.1 研究手順

まず EC サイトの商品レビューを統計分析ソフト R³ を用いて収集する。本研究では評価値と商品カテゴリのフィルタリングが可能な EC サイト Amazon のレビューを使用する。評価値だけでなく商品化カテゴリごとにも分類する理由は商品カテゴリごとに評価の指標となる単語の偏りが異なると考えたためである。

次に学習用データと評価用データについて述べる。学習用データはレビュー文に含まれる主要な単語の抽出方法を構築するためのレビュー群である。一方、評価用データは構築した抽出方法を評価するためのレビュー群である。収集したレビューをあらかじめこの 2 つに分類することで後の検証に使用する。

次に主要な単語の抽出方法について述べる。まずレビューの評価値から学習用データの商品レビューを商品カテゴリごとに高評価レビュー、低評価レビューに分類し形態素解析ソフト RMeCab⁴ を用いて形態素解析を行う。次に高評価レビュー、低評価レビューそれぞれの単語の出現頻度の高い順に順位をつける。ここで高評価、低評価における頻出単語の順位の差を求める。本研究では、この差が大きいほどその商品カテゴリにおける高評価、低評価のレビューの主要な単語とみなし、高評価レビューにおける主要な単語を「ポジティブ語」、低評価レビューにおける主要な単語を「ネガティブ語」と定義する。

最後に評価用データから学習用データで求めたポジティブ語・ネガティブ語を抽出し、その結果の妥当性と評価是正の可能性について考察する。

3.2 商品レビューの収集

2020 年 8 月 25 日時点の Amazon における商品カテゴリ「本⁵」と商品カテゴリ「ノート PC」それぞれの商品レビュー合計 1600 件を R で収集した。Amazon ではレビューワーが 5 段階評価でその商品の評価値を付与する。そこで本研究ではこの評価値が最高の 5 である商品レビューを高評価レビュー、1 もしくは 2 である商品レビューを低評価レビューとみなした。1 商品に対して高評価レビュー 50 件、低評価レビュー 50 件集めるものとし、これを商品カテゴリ「本」においては 10 商品分、商品カテゴリ「ノート PC」においては 6 商品分のレビューを収集した。商品の選定については Amazon 上でその商品カテゴリの売れ筋ランキング上位から、高評価レビュー、低評価レビューを 50 件ずつ集めることが可能な商品とした。この 1600 件の商品レビューの内、半分を学習用データ、残り半分を評価用データとして活用する (表 1)。

表 1: 収集した商品レビューの件数の内訳

商品カテゴリ	データ項目	高評価レビュー	低評価レビュー	合計
本	学習用データ	250	250	500
	評価用データ	250	250	500
ノート PC	学習用データ	150	150	300
	評価用データ	150	150	300
	合計	800	800	1600

3.3 ポジティブ語・ネガティブ語の抽出

次にポジティブ語・ネガティブ語の抽出について述べる。まず、RMeCab を用いて商品カテゴリおよび高評価、低評価ごとの学習用データの商品レビューから単語の出現頻度、品詞を求めた。さらに単語ごとの高評価、低評価における出現頻度の順位を求め、その順位差を算出した。この順位差を求める式は以下のとおりである。

Hd : 高評価レビューの順位差

Ld : 低評価レビューの順位差

Hr : 当該単語の高評価レビューでの出現頻度順位

Lr : 当該単語の低評価レビューでの出現頻度順位

$$Hd = Lr - Hr$$

$$Ld = Hr - Lr$$

前記の式を用い各学習データにおける単語の順位差を算出した。その結果、記号、助詞、助動詞の単語が頻出し、いずれの単語も出現頻度の順位差の絶対値が小さい値となった。そのため本研究では記号、助詞、助動詞はポジティブ語・ネガティブ語になりにくい単語とみなし、これらの品詞を除外しポジティブ語・ネガティブ語の抽出を試みる。

出現頻度の順位差でポジティブ語・ネガティブ語の抽出を試みるにあたり、単語の有効性の検証のため出現頻度の閾値を設ける必要がある。しかし商品カテゴリごとの学習データのサンプル数が異なるため出現頻度から閾値を求めることはできない。そのため単語の出現頻度を単語の総出現数で割った出現率を求める⁶。本研究ではこの出現率 0.001 を閾値とし、出現率がこの数値以上の単語を有効な単語とみなす。出現率 0.001 以上でフィルタリングし、順位差の降順に並び替えた上位 20 単語を表 2、表 3、表 4、表 5 に示す⁷。

本研究ではこの表 2 で示した 20 単語を商品カテゴリ「本」のポジティブ語、表 3 で示した 20 単語を商品カテゴリ「本」のネガティブ語とする。同様に表 4 で示した 20 単語を商品カテゴリ「ノート PC」のポジティブ語、表 5 で示した 20 単語を商品カテゴリ「ノート PC」のネガティブ語とする。

³統計解析に適したフリーソフトウェア。

⁴R からテキストやファイルを指定し MeCab に解析させ、その結果を R で標準的なデータ形式に変換して出力するインターフェイス。

⁵「ビジネス・経済」のジャンル。

⁶小数第六位を四捨五入。

⁷“N/A”は高評価、低評価いずれかのレビュー群に出現しなかったことを表す。

表 2: 商品カテゴリ「本」の高評価レビューにおける順位差上位 20 単語

単語	品詞	品詞細分類	出現頻度	出現率	順位	順位差
○	名詞	サ変接続	24	0.00146	96	N/A
分断	名詞	サ変接続	20	0.00122	127	N/A
気づく	動詞	自立	39	0.00238	55	964
見方	名詞	一般	23	0.00140	100	919
教える	動詞	自立	17	0.00104	149	870
本能	名詞	一般	97	0.00591	13	695
+	名詞	サ変接続	54	0.00329	31	677
相手	名詞	一般	43	0.00262	48	660
物事	名詞	一般	21	0.00128	118	590
対人	名詞	一般	19	0.00116	135	573
上司	名詞	一般	17	0.00104	149	559
ひとつ	名詞	副詞可能	17	0.00104	149	559
基づく	動詞	自立	27	0.00164	85	481
すぐ	副詞	助詞類接続	23	0.00140	100	359
事実	名詞	副詞可能	44	0.00268	46	332
悩み	名詞	一般	24	0.00146	96	282
原因	名詞	一般	25	0.00152	92	196
思い込み	名詞	一般	60	0.00366	23	187
勧める	動詞	自立	19	0.00116	135	191
他者	名詞	一般	56	0.00341	28	182

表 3: 商品カテゴリ「本」の低評価レビューにおける順位差上位 20 単語

単語	品詞	品詞細分類	出現頻度	出現率	順位	順位差
君	名詞	代名詞	16	0.00112	133	N/A
ページ	名詞	一般	17	0.00119	119	1513
こういう	連体詞	*	15	0.00105	147	1485
人達	名詞	一般	17	0.00119	119	995
使える	動詞	自立	17	0.00119	119	710
彼	名詞	代名詞	17	0.00119	119	710
哲学	名詞	一般	20	0.00140	91	570
なぜ	副詞	助詞類接続	17	0.00119	119	542
設定	名詞	サ変接続	15	0.00105	147	406
アクセス	名詞	サ変接続	22	0.00154	77	400
日本語	名詞	一般	20	0.00140	91	386
印象	名詞	一般	19	0.00133	102	375
チンパンジー	名詞	一般	15	0.00105	147	330
当たり前	名詞	形容動詞語幹	19	0.00133	102	318
語る	動詞	自立	18	0.00126	109	311
無駄	名詞	形容動詞語幹	18	0.00126	109	311
こんな	連体詞	*	18	0.00126	109	311
筆者	名詞	一般	24	0.00168	71	291
トラウマ	名詞	一般	23	0.00161	73	289
感じ	名詞	一般	23	0.00161	73	289

3.4 妥当性の検証

次に選定したポジティブ語・ネガティブ語の妥当性を評価用データを使い検証する。検証方法としてウィルコクソンの符号付き順位検定⁸を使用する [7, 8]。学習用データにおける変数 x と評価用データにおける変数 y との差から、その差の有意性を求める。変数 x, y については、各データにおけるポジティブ語・ネガティブ語の出現率を使用する。これは収集したレビューの件数とレビュー内に含まれる単語の総出現数が異なるためだ。そのため評価用データの形態素解析を行い、単語の出現率を求めた。母集団分布の中央値を θ としたとき、各レビューの検定において共通する仮説は以下のとおりである。

帰無仮説 (H_0): 学習用データ (x) と評価用データ (y) との間に単語の出現率の差はない ($H_0: \theta = 0$)。

⁸2 変数 x および y に関する n 組のデータ $(x_1, y_1), \dots, (x_n, y_n)$ が与えられたとき、それらの差を $z_i = y_i - x_i (i = 1, \dots, n)$ とし、 z_i の母集団分布 $f(z)$ の中央値 θ が 0 かどうかを調べる手法。

表 4: 商品カテゴリ「ノート PC」の高評価レビューにおける順位差上位 20 単語

単語	品詞	品詞細分類	出現頻度	出現率	順位	順位差
自分	名詞	一般	22	0.00175	99	N/A
順	名詞	固有名詞	22	0.00175	99	N/A
ありがとう	感動詞	*	18	0.00144	132	N/A
笑	名詞	固有名詞	16	0.00128	144	N/A
心配	名詞	サ変接続	15	0.00120	152	N/A
最高	名詞	一般	14	0.00112	163	N/A
耐久	名詞	一般	14	0.00112	163	N/A
バージョン	名詞	一般	13	0.00104	173	N/A
機能	名詞	サ変接続	20	0.00159	114	700
買い物	名詞	サ変接続	18	0.00144	132	682
ちよつと	副詞	助詞類接続	17	0.00136	139	675
嬉しい	形容詞	自立	15	0.00120	152	662
速い	形容詞	自立	15	0.00120	152	662
メモリー	名詞	一般	15	0.00120	152	662
2020	名詞	数	15	0.00120	152	662
多少	副詞	助詞類接続	13	0.00104	173	641
安心	名詞	サ変接続	13	0.00104	173	641
軽い	形容詞	自立	24	0.00191	86	439
和	名詞	固有名詞	23	0.00183	92	433
大変	名詞	形容動詞語幹	21	0.00167	104	421

表 5: 商品カテゴリ「ノート PC」の低評価レビューにおける順位差上位 20 単語

単語	品詞	品詞細分類	出現頻度	出現率	順位	順位差
返金	名詞	サ変接続	12	0.00190	76	N/A
使い物	名詞	一般	11	0.00174	86	N/A
落ちる	動詞	自立	7	0.00111	156	N/A
広告	名詞	サ変接続	7	0.00111	156	N/A
表記	名詞	サ変接続	8	0.00127	132	1129
失う	動詞	自立	7	0.00111	156	1105
詐欺	名詞	サ変接続	11	0.00174	86	764
最悪	名詞	一般	11	0.00174	86	764
型番	名詞	一般	10	0.00158	104	746
回答	名詞	サ変接続	7	0.00111	156	694
電話	名詞	サ変接続	7	0.00111	156	694
自体	名詞	一般	7	0.00111	156	694
アマゾン	名詞	一般	13	0.00206	71	568
Amazon	名詞	一般	10	0.00158	104	535
仕方	名詞	ナイ形容詞語幹	9	0.00143	116	523
安物	名詞	一般	8	0.00127	132	507
不良	名詞	形容動詞語幹	8	0.00127	132	507
再度	副詞	助詞類接続	7	0.00111	156	483
販売元	名詞	一般	7	0.00111	156	483
ダメ	名詞	形容動詞語幹	7	0.00111	156	483

対立仮説 (H_1): 学習用データ (x) と評価用データ (y) との間に単語の出現率の差はある ($H_1: \theta \neq 0$)。

本研究では有意水準 0.05 とし仮説検定を行う。R を使い検定統計量および P 値を算出した結果を表 6 に示す⁹。

表 6: ウィルコクソンの符号付き順位検定の結果

商品カテゴリ	レビューの分類	検定統計量	P 値	中央値
本	高評価レビュー	206	0.000013	0.001163
	低評価レビュー	186	0.001383	0.000623
ノート PC	高評価レビュー	184.5	0.001755	0.000475
	低評価レビュー	203	0.000034	0.000598

表 6 からいずれのレビュー群においても P 値が有意水準 0.05 以下であるため帰無仮説が棄却され、対立仮説が採択される。つまり学習用データと評価用データとの間におけるネガティブ語の出現率の差は有意な働きを持つと考えられる。

⁹小数第七位を四捨五入。

4 考察

ウィルコクソンの順位付き符号検定を行った結果、P値がいずれも有意水準の0.05以下であり、妥当性が低い抽出となった。これは学習用データから選定したポジティブ語・ネガティブ語が評価用データでは出現頻度が低かったためであると考えられる。つまり商品レビューを収集した段階で学習用データと評価用データとの間において頻出単語が大きく偏っていた可能性がある。そのためサンプル数を増やし、各データ間の頻出単語の偏りを緩和させることが必要であると考えられる。

また不当なレビュー評価の是正可能性について、ネガティブ語が高評価レビューにどのような文脈で含まれているのか分析した結果、大きく3つのパターンがあると考えられる。1つ目は文脈によってポジティブ・ネガティブの極性が変化するパターンである。その一例として本研究では「“こういう”方法があったのか、と衝撃を受けました」、「“再度”購入しました」などの文章がみられ、ネガティブ語がポジティブな表現としてレビュー文に含まれていた。そのため文章全体においてその単語がどのような文脈で用いられているのかをポジティブ語・ネガティブ語の抽出に組み込み、両方の極性を持つ単語を除外するなどの処理が必要であると考えられる。2つ目は評価対象の商品とは無関係の文脈に現れるパターンである。その一例として本研究では「“日本語”教師という仕事柄」、「前のものが“ダメ”になり買い替えました」などの文章がみられ、自分の身の上話や以前購入していた商品に対する言及などレビューのエピソードに関する文脈にネガティブ語が出現していた。そのため言及している対象が明確な場合、その対象が評価対象の商品とどの程度関連性が高いかを考慮する必要がある。また、身の上話のような評価対象の商品と関連性の低い文章を除外するため、単語の共起関係を分析する必要があると考える。3つ目は逆説表現の文脈に現れるパターンである。その一例として本研究では「PC“自体”古いとはいえ」、「質感は“安物”感が半端ないですが」など「が」や「とはいえ」といった逆説表現の接続詞がネガティブ語とともに含まれている文章が確認された。そのためポジティブ語・ネガティブ語の抽出において、このような文脈の可能性も考慮する必要がある。しかしこの逆説表現の接続詞に焦点を当てることで高評価レビューに含まれるネガティブな表現もしくは低評価レビューに含まれるポジティブ表現が抽出が可能となり、不当なレビュー発見につながる。一方、ポジティブ語が低評価レビューにどのような文脈で含まれているのか分析した結果、前述のパターンに加え、否定表現の文脈に現れるパターンがみられた。たとえば「“気づ”きもなく」、「お“勧め”できません」などポジティブ語の後に「ない」、「できない」といった否定表現が付随し、ポジティブ語がネガティブな表現となる文章がみられた。そのため文章内のポジティブ語・ネガティブ語の周囲における否定表現の有無を調べる必要がある。

5 結論

本研究では不当なレビュー評価是正を目指し、単語の出現頻度の順位差を利用したポジティブ語・ネガティブ語の抽出を試みた。そしてウィルコクソンの符号付き順位検定を行い、その商品カテゴリ内のポジティブ語・ネガティブ語が妥当なものであるかを明らかにした。これにより商品レビューのテキストマイニングによるポジティブ語・ネガティブ語抽出の可能性を見出した。

本研究では商品レビュー最大250件を学習用データとしポジティブ語・ネガティブ語の抽出を行ったが、評価用データと異なる抽出結果となり、各データにおける出現単語の偏りがみられた。そのためサンプル数を増やし、抽出するポジティブ語・ネガティブ語の妥当性を高める必要がある。また実際に不当なレビューを抽出するためには、レビュー文内における単語同士の関係性を分析する必要がある。複数のテキストマイニングの手法を併用し、ポジティブ語・ネガティブ語の抽出および抽出単語の検証をすることが今後の課題としてあげられる。そして、本提案手法により不当なレビューの検出が容易になり、評価の是正を行う際の一助とすることが期待できる。

参考文献

- [1] Ben Fox Rubln,cnet. “<https://www.cnet.com/news/amazon-updates-customer-reviews-with-new-machine-learning-platform/>”.2015
- [2] 石田基広.Rによるテキストマイニング入門 第2版, 森北出版,2017
- [3] 那須川哲哉. テキストマイニングを使う技術/作る技術:基礎技術と適用事例から導く本質と活用法, 東京電気大学出版局,2006
- [4] 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,13(3),pp.201-241,2006
- [5] 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,3Q-06,pp.535-536,2017
- [6] 那須川哲哉, 金澤博. 文脈一貫性を利用した極性付評価表現の語彙獲得. 情報処理学会研究報告自然言語処理,pp.109-116,2004
- [7] 岩崎学. 統計的データ解析入門 ノンパラメトリック法, 東京図書,2006
- [8] 村上秀俊. 統計解析スタンダード ノンパラメトリック法, 朝倉書店,2015