評 判 分 析

国分芳宏（2022,7,18）

マーケティングやクレーム処理の支援のために、ツイッター、フェイスブックなどの記事から商品の評判分析をするプログラムを開発した。構文解析をして何がポジティブなのかネガティブなのかを評価した。使用されている用語によるネガティブ・ポジティブの判定だけでなく、共起語辞書を用いて用語同士の関係によってネガティブ・ポジティブを判定している。ネガティブ・ポジティブの評価は利用者の特に職業によって異なることがある。用語の組み合わせとその組み合わせのネガティブ・ポジティブの性質を登録できる仕組みも作った。

キーワード：意味解析、共起語辞書

Reputation Extraction.

Yoshihiro Kokubu

In order to support marketing and complaint processing, we developed a program for making the reputation analysis of articles in Twitter, Facebook, etc. First we evaluated whether individual articles were positive or negative as a whole by semantic analysis. Next we extracted points that we judged to be positive or negative across sentences. The judgment of positiveness or negativeness was based not only on terms alone but also on the relations between terms according to our thesaurus. The relations between the terms are expanded by the collocation dictionary. In order to recognize negation and make precise analyses, we treated the sequence of bound morphemes at the end of a phrase as modality.

Key Words: Semantic analysis, collocation dictionary,

1. はじめに

ツイッター、フェイスブックなどの CGM から情報を容易に集められるようになってきた。かえって量が多すぎて人手での分類は不可能に近いので、内容を調べて分類するプログラムが求められている。筆者等は集めた記事の評判が好意的であるか非好意的であるかを分析して分類するプログラムを開発した。

マーケットサポートのためには、記事単位でネガティブ・ポジティブを評価するだけでなく、複数の記事をまたいで何がポジティブで、何がネガティブかを知らせる方が効果がある。複数の記事を解析して、何が評価されているかを分類している。

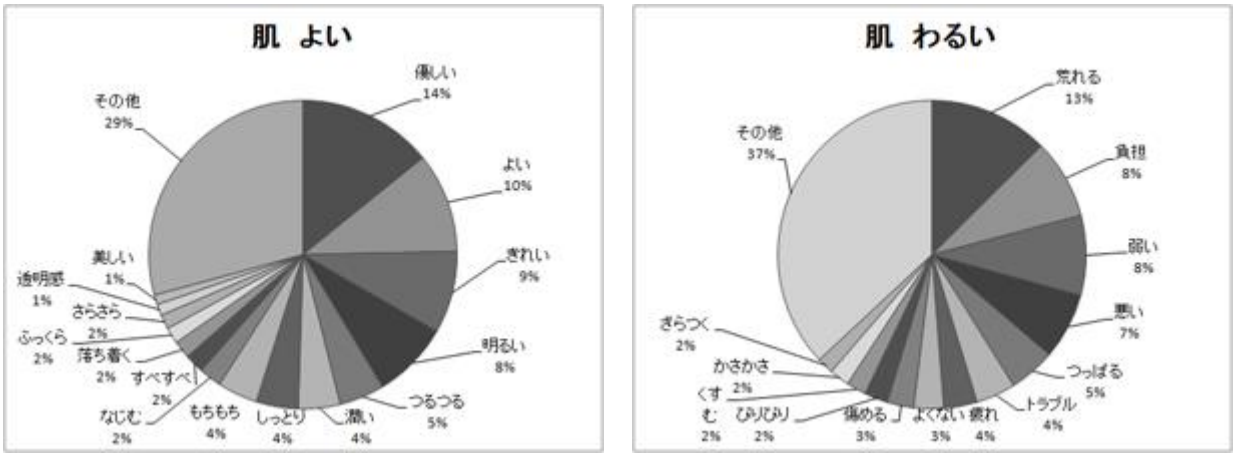

図 1　化粧品の口コミを「肌」について評価した。

説明の便宜上好意的なものを「ポジティブ」非好意的なものを「ネガティブ」と記した。また本校の筆者と区別するため、分析対象の記事を書いた人を投稿者と呼んだ。

記事中の単語でポジティブか、ネガティブかを決定しているが、単語だけではネガティブ・ポジティブが決められない場合がある。本プログラムは単語だけでなく、共起語辞書を導入して用語同士の関係も考慮してネガティブ・ポジティブを判定している。

否定によってネガティブ・ポジティブは逆転するが、否定かどうかを判定するためにモダリティーを導入した。モダリティーとは文節末尾の表 2 に示すような意味を表す部分で、本稿では（益岡）に説明してあるものを拡張して否定も含めている。文脈や投稿者の年代によってネガティブ・ポジティブの変わる用語もあるが、現在ツイッター、フェイスブックなどで使われている意味で判定した。

ネガティブ・ポジティブ判定には文脈的な検査、さらには知識ベースの支援が必要になる。ここでは、意味解析した結果の単語、連語のレベルまでで判定している。

2　ネガティブ・ポジティブの判定

2.1　単語による判定

参考文献3を用いて意味解析をした結果からネガティブ・ポジティブを判定している。まず文を構成している用語からネガティブ・ポジティブを判定する。筆者等の解析辞書には用語がネガティブ・ポジティブをつけて登録してある。CGM に出現する文には顔文字が使われているものがある。顔文字にもネガティブ・ポジティブをつけて辞書に登録してある。

例1　ネガティブ・ポジティブの性質を持つ用語の例

ポジティブ	ネガティブ	
美人	瓦礫	体言
涼しい	寒い	形容詞
生まれる	死ぬ	動詞
(^ｏ^)	(:_;)	顔文字

外来語は日本語の中で使われている意味でネガティブ・ポジティブを判定した。

ナイーブ　(無邪気) ポジティブ

日本語では無邪気なという意味で使われているが、手元の英和辞書によるとナイーブ(naive)には他に「ばか」「騙されやすい」などの意味もある。

使い方によってネガティブ・ポジティブが異なる用語がある。前後の接続関係を調べてネガティブ・ポジティブを判定した。

例2　使い方によってネガティブ・ポジティブの異なる例

さっぱりする　(動詞　ポジティブ)

さっぱりだ　(形容動詞　ネガティブ)

複合語

複合語は経験則により後ろの用語のネガティブ・ポジティブを全体のネガティブ・ポジティブにした。しかし結合するとネガティブ・ポジティブが変わる複合語がある。ネガティブ・ポジティブが変わる複合語は結合したものを一つの用語として辞書に登録した。

例3　ネガティブ・ポジティブの変わる複合語の例

リスク　カット　→　リスクカット　(ポジティブ)

リスト　カット　→　リストカット　(ネガティブ)

近代的　　（ポジティブ）
前近代的　　（ネガティブ）

連語

用語が組み合わされた結果ネガティブ・ポジティブの性質を持つ組み合わせがある。

例　気勢が　　上がる　　（ポジティブ）

陳述副詞にも注意してネガティブ・ポジティブの性質を決める。

是非　　買いたい　　（ポジティブ）
二度と　　買わない　　（ネガティブ）

一つの用語が連語の組み合わせでネガティブ・ポジティブが変わるものがある。

「甘い」という用語は多義語であるがそれぞれの意味でネガティブ・ポジティブが異なるが組み合わせによって性質が異なる。

例 13　ネガティブ・ポジティブの異なる多義語

果物が甘い　　（甘味）　　（ポジティブ）
検査が甘い　　（手ぬるい）　　（ネガティブ）

慣用句

慣用句は教訓的なものが多いためか、ネガティブ・ポジティブの性質を持つ物が多い。

活用があるので用言として登録することになる。使い方によって形が変わるので、一つの用言としてまとめた形で登録する共起語辞書に連語として登録した。

例 10　慣用句が変形する例

（彼は）機転が　利く
（彼は）機転も　利く
　　機転の　利く（男だ）

また慣用句に連体修飾格が係ることもある。慣用句を一つの用言として登録した場合、連体修飾格が受けられない。連体修飾格が係ることがある慣用句にマークをして連体修飾格が係ることができるようにした。

例 11　慣用句に連体修飾格が係る例

足を引っ張る
彼の足を引っ張る

慣用的な言い回しに含まれる動詞は普段あまり使わない活用形が見受けられる。普段使わない活用形のまま連語として登録した。

例 5　普段あまり使わない活用形

花を持たせる　　使役形
完成が待たれる　　自発形

ネガティブ・ポジティブの性質をもつ慣用句に、３つ以上の文節で成り立つものがあるが、当面２つをまとめたものを一つの自立語として辞書に登録した。

例 9　係りまたは受けが複数の文節で成り立つ例

頭の　中が	真っ白だ	「頭の中」を登録
ライバルに	後れを　取る	「遅れを取る」を登録

複文

複文では、前の単文が逆接のときは後ろの単文の値を複文全体のネガティブ・ポジティブの値にした。

単文のモダリティーを用いて単文同士の接続関係からネガティブ・ポジティブの有効性を評価して全体のネガティブ・ポジティブを判定している。ネガティブ・ポジティブの判定で不要になる単文は無視している。

例 15 モダリティーにより、ネガティブ・ポジティブが無効になる例

補強したが	負けた	ネガティブ	逆接
安ければ	買ったのに	ネガティブ	条件
買ったほうが	良い	ポジティブ	反実仮想

逆接の接続助詞の形をしていても、慣用的に逆接にならない場合があるが。このような言い回しはまとめた形で辞書に登録した。

例 16 逆接にならない例

すみませんが
申し訳ありませんが

助動詞　（ら）れる

自立する動詞によってまたその使い方によって、ネガティブ・ポジティブの値が変わる。

[]で囲まれた文字は動詞に係る名詞の意味を表す。

受身

受け身形かどうかを判定することは難しいが、受身形になると一部の動詞を除いてネガティブな意味を持つようになる。ネガティブな性質を持たない動詞は受け身形を独立した動詞として辞書に登録した。

例 4　受身形、可能形でネガティブ・ポジティブの性質が変わる例

［人］に叩かれる

［物］を叩かれる

受け身形がネガティブにならずポジティブになる受身

褒められる、癒やされる、心が洗われる、慰められる、守られる、育まれる、育てられる、与えられる、ほどこされる、たたえられる、助けられる

可能

ポジティブな値を持つ

［物］を食べられる

［物］が食べられる

尊敬

ネガティブ・ポジティブの性質を持たない。

［人］が食べられる　　尊敬形

自発

伝達や思考などの意味の動詞で受け身のかたちをとっていますがネガティブ・ポジティブの性質は持たない。引用のとに続きます。

例　と言われる、と思われる、と考えられる、と信じられる、と察せられる、と推察される、と報告される、と予想される、と偲ばれる、と見受けられる

否定

否定によってネガティブ・ポジティブは逆転するが、日本語では否定の「ない」「ません」「ず」を含んでいても否定だとは決められない。慣用句のなかに含まれている場合がある。

本プログラムでは否定かどうかを判定するために使っている。（文献 3 参照）

例 14 「ない」を含んでいても否定にならない例

	意味
調べないか	勧誘
調べるかも知れない	推量
調べなければならない	義務
調べてもらえないか	依頼
しょうがない	慣用的
すみません	間投詞
やむを得ず	接続詞

知識ベースによる判定

ネガティブ・ポジティブの判定は知識ベースを調べないと決められないことがある。

ここでは知識ベースとして次のような組み合わせを用いた。

対象　　属性　　値　　比較　　ネガティブ・ポジティブ

対象

対象を特定するときに記述する。連体修飾格になる。

例　国分の

属性

例　体重が

値

実現する値の他に用言も記述できる。

例　６０キロ

比較として次のものを用意した。値が用言のときは記述しない。

＝、＞、＜、＞＝、＜＝

増える、減る、多い、少ない、大きい、小さい、超える、超過

異常、以下、未満

ネガティブか、ポジティブかの値

P, N

4　詳細分析

記事のネガティブ・ポジティブではなく複数の記事全体を通して、ある用語に注目してどういう点がポジティブで、どういう点がネガティブかを抽出した。

4. 1　用語の標準化

同じことを表しているのにいくつもの表記方があって､記事によって異なることがある。類似した概念を整理するときの障害になるので、共起語辞書を用いてシステムで推奨する用語に標準化した。さらに厳密には意味が異なり、共起語辞書でも同義語としては扱われていないが、集計している分野だけでみれば、同義語と見て差し支えない用語は標準の用語に置き換えた。どこまでを同じ用語に置き換えてよいかは、評価する目的や分野によって異なるので人手で確認しながら作業した。

例 17 語標準化の例

お肌　乾燥肌、素肌　　→　肌

染み込む、浸透　　→　染みる

他に次の点についても標準化した。

限定用法の連語は全て叙述用法に統一した。

否定の表現は共起語辞書を参照して反義語の肯定表現に置き換えた。

例 18 否定を反義語に置き換える例

長くない　　→　　短い

4.2　集計

集計しようとする用語を共起語辞書で拡張して拡張した用語を持っている記事を収集した。構文解析して係り受けデータにしたあと、注目する用語を含む連語を抽出する。ネガティブ・ポジティブ別、係り別、受け別に整理して表にした。図１は化粧品についてのロコミを「肌」についてのネガティブ・ポジティブと評価のあった連語をまとめてグラフ化したものである。

5. 結果

ロコミランキング(http://kuchiran.jp/)に表示されている記事(14,907 件)の人手によるネガティブ・ポジティブの判定結果をもとに連語辞書を組み込んだ場合と、組み込まない場合とを比較した。連語辞書による精度の向上はあまり見られなかった。

共起語辞書付き	86.3%
共起語辞書なし	85.1%

次のようなものが誤った判定の原因になった。

.1　連語までによる評価では能力を超えていてネガティブ・ポジティブが判定できなかった。
本プログラムでは、2 つの文節の関係でネガティブ・ポジティブを判定しているが、下記のような文では判定能力の限界を超えていて人手での判定と食い違ってしまった。

例 19 人手でポシと判定した文をネガティブと判定してしまった例

「ベーコンやハムも美味しいので、行くたびに 1 キロ太る」

「ここにはその店にないものが沢山ある」

例 20 人手でネガティブと判定した文をポシと判定してしまった例

「他社の製品をお使いになることをお勧めします」

「きちんと説明してもらいたい」

2　口こみの投稿者の年代でネガティブ・ポジティブが異なる多義語がある。
文のネガティブ・ポジティブを決定する用語であるにもかかわらず共起関係で区別できない。また時とともに変わるので辞書のアップデートが必要である。

例 21 文脈でネガティブ・ポジティブが変わる多義語の例

ヤバい

凄い

クセになる

適当

結構

いい加減

悩ましい

ぞっとしない

.3　ネガティブ・ポジティブが知識によらないと決められない場合がある。

例 22 知識によらないとネガティブ・ポジティブが決められない例

体温が 40 度ある。

画面が 30 インチになった。

.4　投稿者の立場によってネガティブ・ポジティブが異なる場合がある。業種ごとにカスタマイズが必要になる。

例 23 投稿者の立場によってネガティブ・ポジティブの変わる例

円が高い

輸出産業　　ネガティブ
輸入産業　　ポジティブ

6. おわりに

今後精度を上げていくためには3文節以上の組み合わせでネガティブ・ポジティブが決定され場合の対処法を考える必要がある。まず当面は分野別のコーパスを解析して分野ごとの連語辞書を充実させていく予定である。

辞書の整備には25年かかったが、まだ十分とはいえず、今後も新語を中心に更新していく必要がある。

謝辞　口コミランキングの記事を参照させていただきました。株式会社アイスタイル並びに株式会社Hayakawaに感謝いたします。

参考文献

1.益岡隆志(2000)　命題とモダリティの境界を求めて　日本語文法の諸相　くろしお出版
2.松吉俊他,佐藤理史,宇津呂武仁(2007)　日本語機能表現辞書の編纂　自然言語処理 Vol.14 No.5, pp123-146
3.国分芳宏,梅北浩二,松下栄一,末岡隆史(2010)　シソーラスを組み込んだ意味解析システム　自然言語処理　Vol.17 No.4, pp44-57
4.国分芳宏,岡野弘行 (2010)　複数の観点で分類した自然言語処理用共起辞書　自然言語処理　Vol.17 No.1, pp247-263
5.藤村滋他,豊田正史,喜連川優(2005)　文の構造を考慮した評判抽出法　DEWS”% 6C-i8
6 寺田雄一郎,田辺俊文。小山泰男,吉村賢治,首藤公昭(2007)　日本語連語データの整備 福岡大学工学集報　第79号