

商品の Consumer Generated Media 上での 評判解析

国分 芳宏・梅北 浩二・松下 栄一
株式会社言語工学研究所 (2011, 5, 28-2014, 2, 8)

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で商品の評判分析をするプログラムを開発した。まず意味解析で個々の記事がポジかネガかを評価した。次に複数の記事をまたいでポジと評価した点、ネガと評価した点とを抽出した。用語単独によるポジネガの判定だけでなく、シソーラスを用いて用語同士の関係によってポジネガを判定している。その用語同士の関係もシソーラスで拡張している。解析を正確にするためと、否定を認識するために文節の後ろの付属語の並びをまとめてモダリティーとして整理した。

キーワード：意味解析、シソーラス、連語、多義語、モダリティー

Reputation Extraction Across Consumer Generated Media.

Yoshihiro Kokubu, Kouji Umekita, Eiichi Mathushita

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 thesaurus.

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, thesaurus, collocation, a polysemous word, modality

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、記事単位でポジネガを評価するだけでなく、複数の記事をまたいで何がポジで、何がネガ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。

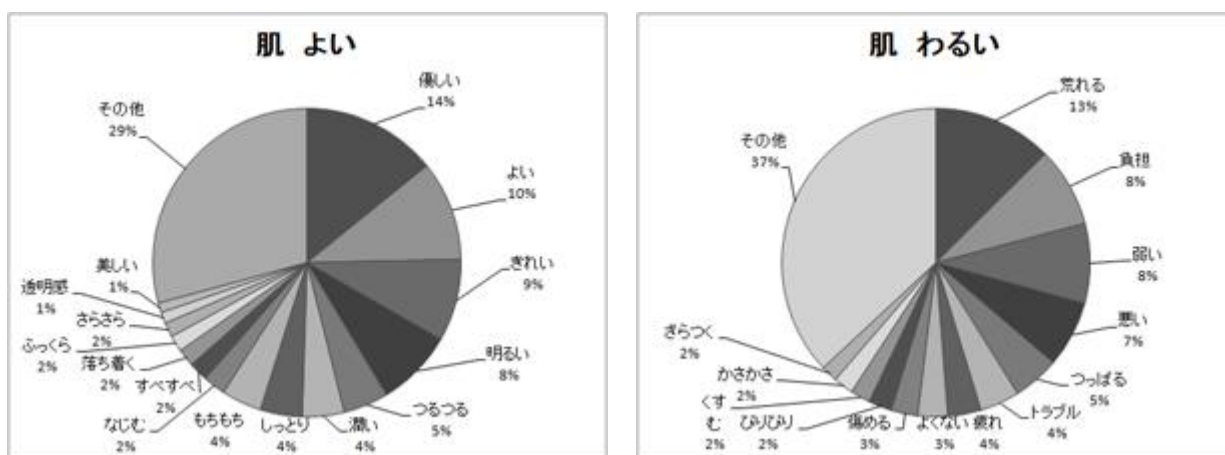


図1 化粧品の口コミを「肌」について評価した。

説明の便宜上好意的なものを「ポジ」、非好意的なものを「ネガ」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か、ネガかを決定しているが、単語だけではポジネガ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、シソーラスを導入して用語同士の関係も考慮してポジネガを判定している。

否定によってポジネガは逆転するが、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ポジネガ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。

（藤村滋他）ではポジネガを文節の n-gram で判定しているが、本プログラムでは意味解析をして連語で判定している。

（松吉俊他 2007）には、単独の機能表現が集めてあるが本プログラムでは複数の機能表現が組み合わせられると、単独の機能語とは異なった意味を持つことがあることに注目して、

連続した形で管理している。

2 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の連語をベースにしてポジネガを判定している。まず連語に含まれている用語自体の性質からポジネガを判定する。解析辞書には用語がポジネガ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り、投稿者の感情を推測できる。よく使われる顔文字はポジネガをつけて辞書に登録してある。

例1 ポジネガの性質を持つ用語の例

ポジ	ネガ	
美人	瓦礫	体言
涼しい	寒い	用言
(^o^)	(:_;)	顔文字

外来語は日本語の中で使われている意味でポジネガを判定した。

ナイーブ (無邪気) ポジ

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。

使い方によってポジネガが異なる用語がある。前後の接続関係を調べてポジネガを判定した。

例2 使い方によってポジネガの異なる例

さっぱりする ポジ
さっぱりだ ネガ

2.1 複合語

ポジネガが矛盾する組み合わせの場合は経験則により後ろの用語のポジネガを全体のポジネガにした。しかし結合するとポジネガが変わる複合語もある。ポジネガ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 ポジネガが変わる複合語の例

リスク カット → リスクカット ポジ
リスト カット → リストカット ネガ

2.2 受身形、可能形

動詞の受身形や可能形に対応するための語形の変化は語幹に活用語尾、助動詞を付けて派生させていた。しかし受身形、可能形になるとポジネガが変わる動詞がある。受身形はネガの意味、可能形はポジの意味を持つ動詞が多かったが一般性がないため、ポジネガが変わる動詞は派生した形を独立した動詞として辞書に登録した。

- 例 4 受身形、可能形でポジネガの変わる例
 褒められる (褒めるの受身形：ポジ)
 売れる (売るの可能形：ポジ)

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。

- 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

3. 連語による判定

用語単独ではポジネガが決まらず、連語になるとポジネガが決まる組御あわせがある。

そこで次の目的でシソーラスを導入した。(文献 4 参照)

- ・連語になると単独のときとポジネガが変わるものを判定する。
- ・意味解析で係り先の曖昧性を減らす。
- ・多義語の意味を特定する。

全ての連語を登録するわけにはいかなかったが、よく使われる連語だけでなく、組み合わせられた結果ポジネガの変わる連語をシソーラスに登録した。連語によるポジネガの評価は、用語単独での評価より優先した。

連語はシソーラスの一部として登録してある。「係り」と「受け」の用語と、間に挟まれる「格助詞」、および「ポジネガ」の情報が記入してある。「係り」、「受け」用語はシソーラスでそれぞれの同義語、狭義語を拡張できる。固有名詞を扱えるようにするために係りの用語として体言の意味を指定することができる。また接尾辞を指定することもできる。(参考文献 2 を参照)

- 例 6 動詞「行く」に係る「係り」の例

駅	に	行く	
集合場所	に	行く	複合語は後ろの用語の意味を使う
(名古屋)	に	行く	固有名詞名古屋は場所意味を持つ
(支給) 所	に	行く	「所」場所の意味の接尾辞を指定

シソーラスの連語には格助詞も登録されている。キーになる文の助詞を外してシソーラスを引いた後で、キーにある助詞を拡張して許される格助詞に置き換える。

場所	へ	行く	
場所	に	行く	→ 場所 に 行く
場所	まで	行く	

格助詞が省略された場合や、副助詞や係助詞に置き換わっていた場合もプログラムで拡張している。

限定用法は叙述用法にしてシソーラスを引く

用言の使い方には限定用法と叙述用法がある。文中に「長い寿命」（限定用法）と書いてある連語で、「寿命が長い」（叙述用法）というシソーラスを引いても検索できない。文中の限定用法の連語は叙述用法にしてシソーラスを引く。

用言が動詞の場合は、叙述用法に直すときも、自動詞か他動詞かによって置き換える格助詞が異なる。

例 7 自動詞、他動詞で置き換える助詞が変わる例

限定用法		叙述用法
食べたリンゴ	→	リンゴを食べた（他動詞）
落ちたリンゴ	→	リンゴが落ちた（自動詞）
赤くしたリンゴ	→	リンゴを赤くする（形容詞の他動詞用法）
赤くなったリンゴ	→	リンゴが赤くなる（形容詞の自動詞用法）

処理する文が受動態の場合には係りと受けの間に挟まれる格助詞が変わるのでプログラムで置き換えている。動詞が自動詞か、他動詞で置き換える格助詞が変わる。

例 8 能動態、受動態で格助詞が変わる例

能動態		受動態
田中 が 広める（他動詞）		田中 に 広められる
噂 を 広める（他動詞）		噂 が 広められる
父 が 死ぬ（自動詞）		父 に 死なれる

ポジネガの性質をもつ連語に、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

3. 1 慣用句

慣用句は教訓的なものが多いためか、ポジネガ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が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ことができない。分解してシソーラス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	ポジ
(彼は) 機転も 利く	ポジ
機転の 利く (男だ)	ポジ

また係りの体言に連体修飾格に係ることもある。慣用句を一つの動詞として登録した場合、連体修飾格が受けられない。これも慣用句を分解してシソーラスに登録した理由である。

例 11 慣用句に連体修飾格に係る例

(彼の) 足を引っ張る (彼が) 足を引っ張る。

3.2 ポジネガの判定

連語の組み合わせでポジネガの変わる組み合わせがある。

例 12 連語の組み合わせでポジネガの性質が発生する例

寿命が延びる ポジ

寿命が短い ネガ

「寿命」「延びる」「短い」などの用語は単独ではポジネガの性質を持たないが連語関係になるとポジネガの性質を持つ。

連語の組み合わせで多義語の意味を区別してポジネガを決定できるものがある。

「甘い」という用語は多義語であるがそれぞれの意味でポジネガが異なるが組み合わせによって意味が決定できる。

例 13 ポジネガの異なる多義語

果物が甘い (甘味、スイート) ポジ

検査が甘い (手ぬるい) ネガ

3.3 否定の判定

否定によってポジネガは逆転するが、日本語では「ない」「ません」と書いてあっても否定だとは決められない。モダリティーを用いて否定かどうかを決めている。

文節の自立語の後ろに続く付属語の並びをモダリティーと呼んでいる。付属語には助詞、助動詞だけでなく、いわゆる機能動詞、形式名詞などを含んでいる。会話で使われる崩れた言い回しも含んでいるため総数は 137 万行に及んでいる。本プログラムでは否定を検出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
飲まないか	勧誘
飲むかも知れない	推量
飲まなければならない	義務
飲んでもらえないか	依頼
しょうがない	慣用的

3.4 複文

複文では、構成する短文のポジネガの値を文全体の値にしている。前後の短文でポジネガの値が異なるときは、経験的に後ろの短文の値を複文全体の値にした。

短文のモダリティーを用いて短文同士の接続関係からポジネガの有効性を評価して全体

のポジネガを判定している。ポジネガの判定で不要になる短文は無視している。

例 15 複文でモダリティーによりポジネガが無効になる例

連続する短文の例

是非	買いたい	ポジ	
二度と	買わない	ネガ	
強いが	負けた	逆接	無視
良ければ	買ったのに	条件	無視
買ったほうが	良い	反実仮想	無視

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

4 CGM での詳細分析

記事のポジネガ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で、どういう点がネガかを抽出した。

4.1 用語の標準化

同じことを表しているのにいくつかの表記方があって、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、シソーラス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、シソーラス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。

否定の表現はシソーラス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シソーラス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。ポジネガ別、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口コミを「肌」についてのポジ、ネガ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) の人手によるポジネガ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。

シソーラス付き	86.3%
シソーラスなし	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ポジネガが判定できなかった。

本プログラムでは、2つの文節の関係でポジネガを判定しているが、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 人手でポジと判定した文をネガ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 人手でネガと判定した文をポジ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ポジネガが異なる多義語がある。

文のポジネガ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 文脈でポジネガ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 ポジネガが知識によらないと決められない場合がある。

例 22 知識によらないとポジネガが決められない例

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ポジネガ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ポジネガの変わる例

円が高い

輸出産業	ネガ
輸入産業	ポジ

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ポジネガ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 ロ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シソーラス 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