



## ④ ユーザ作成のコンテンツに着目した嗜好・評判抽出

# 2. テキストを用いた評判と嗜好の分析

金山 博

日本アイ・ビー・エム(株) 東京基礎研究所

本稿では、テキストを解析して、事物に対する評判や人々の嗜好を知るための技術を紹介する。単語を拾い集めるだけでなく、文の構造や文脈を解析することによって、「事物の良い点・悪い点を正確に把握する」「評価を示す語句を自動的に求める」「人々の好みの変化を推測する」といった、有用な処理を実現できるようになる。

### 人々の意見を活用するために

近年、Web上の掲示板・blogなどから、製品やサービス・企業などに対する人々の意見を知ることができるようになった。これらは、人手では読みきれない量のテキストの中に存在し、消費者が購入する製品を選ぶとき、企業が次期製品開発の計画を練るときなど、さまざまな場面で、貴重な情報となり得る。

たとえばDVDレコーダを買おうとするとき、各社製品の利用者の声を参照して、どれが一番良い製品かを入念に調べたくなる。このときに役立つ技術の1つが、文書中における語句の分布から、人々の評価の高さを推定し、5段階評価で表示したりするもので、これを**文書単位の評判分析**と呼ぶ。

しかし、製品Aは星3つ、製品Bは星2つという結果だけで、製品Aを買おうと決断できるものでもない。ダビングの速度、録画予約の操作性、または価格の安さなど、自分が重視する項目で優れている製品を選びたい。そこで、評価にまつわる個々の内容を拾い出すことができる**言及単位の評判分析**が求められてくる。各製品の総合的な評価でなく、どの部分がどういった点で良いか悪いかという意見が出力されれば、製品ごとに長所・短所を比較したり（例：機種Aは画質が良く、機種Bは操作性が良いという知識を得る）、特定の部分に関する感想を検索したり（例：機種Aのサポートについての不評の意見を読む）といった、多角的で高度な利用が可能となる。図-1は、文書単位・言及単位の評判分析の流れを図示したものである。

あるDVDレコーダについての意見の例(1)では、一文の中に、番組表に関する好評の意見と画質に関する不

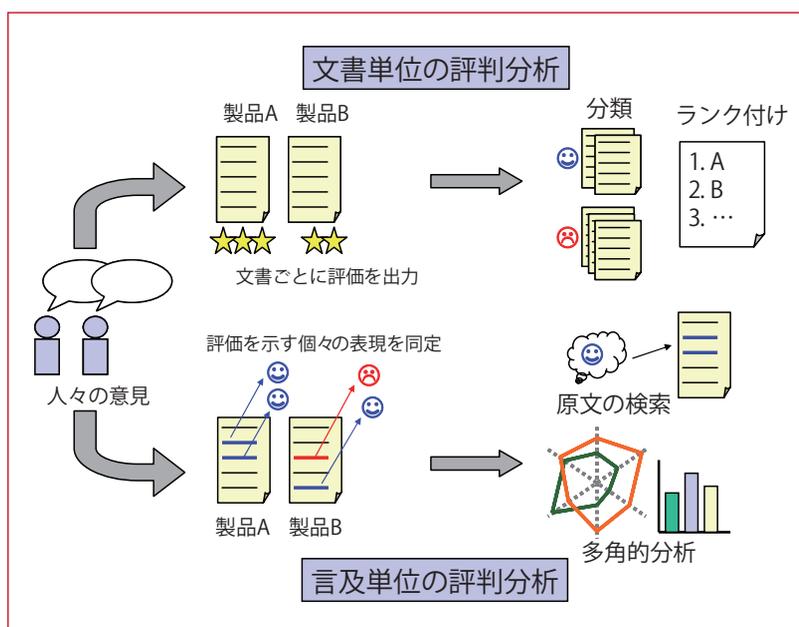


図-1 文書単位・言及単位の評判分析の比較

評の意見が含まれている。言及単位の評判分析は、これら2つの点について、好評 😊 / 不評 😞（これを**評価の極性**と呼ぶ）だけでなく、評価の内容・評価の対象を含めた、(2)のような情報を抜き出すものである。

- (1) 番組表は便利なんだけど、どうも画質が悪いね。
- (2) 😊 便利だ [番組表]
- 😞 悪い [画質]

本稿では、テキストデータを解析して、評判・嗜好にまつわる有用な知識を得るための諸手法を紹介する。いずれの手法においても、単語や句を文中から拾うだけではなく、文法的な構造を考慮し、書かれている内容をできるだけ詳しく、正しく捉えることがポイントとなる。

最初に、言及単位の評判分析において、正確に評価を拾い出すための文の解析方法について述べる。文書単位の評判分析では、文書全体における好評・不評の傾向さ

え掴めれば、多少の解析誤りがあっても影響は小さい。一方、言及単位の評判分析では、評判の検索などで原文にアクセスする機会があり、1つ1つの処理の誤りが露呈しやすいので、特に正確な処理が求められる。

次に、良いか悪いかの手がかりとなる語句の辞書を自動的に作成する手法を紹介する。より多くの意見を拾い出すためには、「😊素晴らしい」「😞醜い」といった汎用的な語だけでなく、カメラに関してなら「😊解像度が高い」、映画に関してなら「😊ドキドキする」のような分野別の語彙が必要となる。しかし、人手で辞書を作成するには多大な労力を要する。そこで、大量のテキスト(コーパス)から自動的に辞書を作成することを試みる。

さらに一歩進めて、そもそも人々が何をもちて良い・悪いと思っているかを知るための処理である**嗜好分析**について簡単に解説する。辞書の自動作成と同様に、好評・不評を述べている文脈を手がかりとして、良い・悪いと思われる性質が、時代とともにどのように移り変わっているか、男性と女性で差があるか、といった知見を得ることができる。

**正確に評価を抜き出す**

ここでは、言及単位の評判分析を目的とした、評価を正確に検出する手法<sup>1)</sup>について解説する。特に、文全体の構造を考慮することにより、局所的な語句の情報だけでは難しい処理が可能になる例を示す。

**【極性を持つ語を定める】**

評価を表す部分を見つける手がかりとして最も典型的

<sup>1)</sup>「魅力は大いに感じる」など、助詞と動詞の間に他の語句が挿入されるため、一繋がり of の句とは限らない。

	😊 好評	😞 不評
極性語	良い, 好きだ, 満足する, 面白い, 美しい, 最適だ	悪い, ひどい, 嫌う, 失望する, 醜い, つまらない, 小さい
極性句	センス-が ある, 魅力-を 感じる	頭-に くる, 被害-を 受ける

表-1 極性語句の辞書の例

なものは、良し悪しを直接表す「😊良い」「😞ひどい」などの語である。その他、好き嫌いを示す「😊満足する」や、知覚的な特性「😞醜い」などを含めて、評価の極性を示す語を**極性語**と呼ぶこととする。

また、動詞「ある」は単独では極性語とならないが、「センスがある」のように特定の名詞と助詞を従えたときは極性を持つ。このような熟語的表現を**極性句**と呼ぶこととし、「😊センス-が|ある」のように表記する<sup>☆1</sup>。極性語と極性句をまとめて極性語句、それらをまとめたデータを**極性語句の辞書**と呼び、表-1に例を示す。

**【句の構造を捉える】**

極性語句の辞書を用いて、テキスト中の表現が好評や不評を表すかどうかを判定する。その際に、用言(動詞・形容詞・形容動詞)が右端にある句の構造を調べる必要がある。以下の例文(3)・(4)の構造を解析して、辞書と比較するときの動作を、図-2で見よう。

- (3) 脇役のほうが主人公よりずっと良かったね。
- (4) 私はCGに魅力をすごく感じました。

例文(3)の内容を簡単な構造に整理するため、係り受けを求めて、副詞や終助詞などを省く。また、「脇役のほうが」という部分は、「脇役-が」としておく。すると、図-2の左側の点線枠内にある構造が作られ、極性語句の辞書と比較しやすくなる。辞書の中にある、「😊良い」

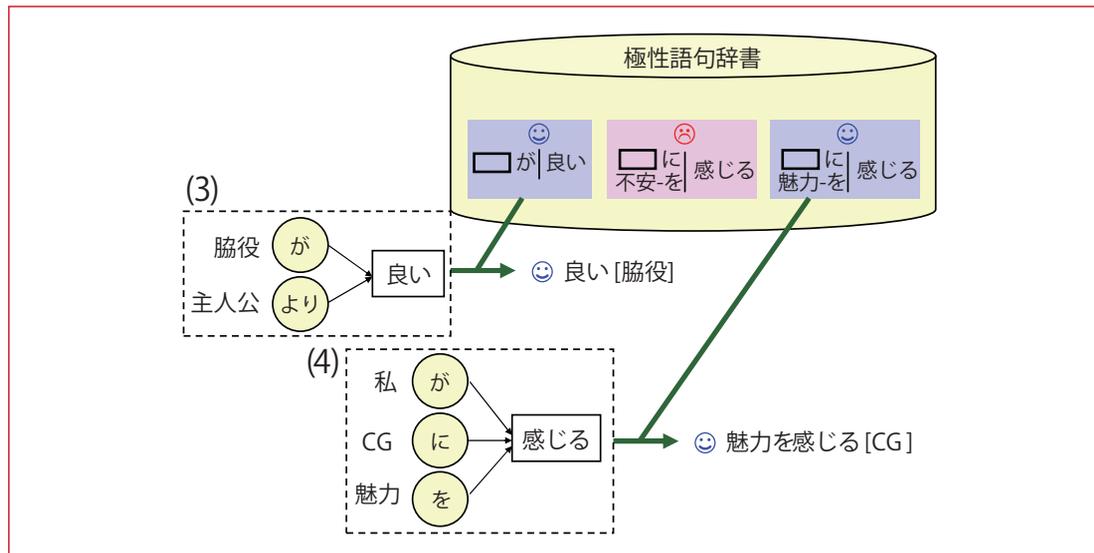


図-2 句の構造の解析

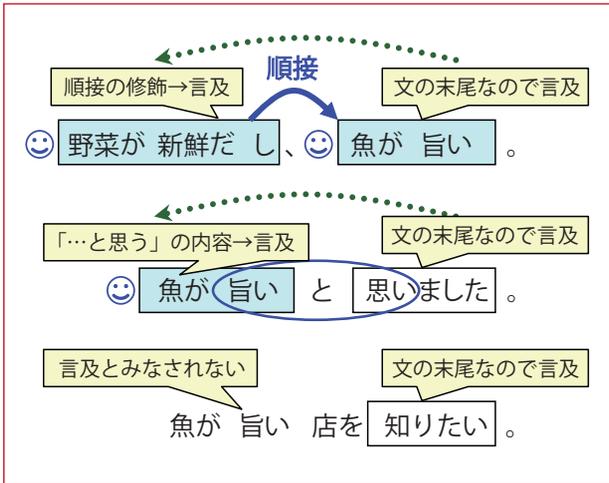


図-3 言及である部分の探索

の評価の対象は助詞「…が」の部分だという情報を用いれば、例文(3)は(「主人公」でも「ほう」でもなく)「脇役」についての好評であることが正しく求まる。

例文(4)の場合、点線枠内の構造を、辞書中の動詞「感じる」と比較する。図-2の辞書には「感じる」が2つあるが、この中で「魅力-を」という部分が一致しているものを選べば、「CG」についての好評であることが正しく求まる。なお、原文の「私は」が「私-が」に変換されているように、文中の役割に応じて助詞を変換する処理(一般に「格解析」と呼ばれる)も重要である。

このように、原文を標準的な構造に変換してから辞書と比較することで、極性と評価の対象を見通しよく判定することが可能となる。

【文の構造を捉える】

映画の分野の例文(5)・(6)は、いずれも好評の極性語句を含んでいるが、評価を表しているものではない。したがって、上記の方法による句の判定を文中の任意の場所で行うと、誤った評価を検出してしまうことになる。

(5) みんな**気に入って**くれれば見るつもりだ。

(6) 音楽が**素晴らしい**映画を知りたい。

これらの誤りを防ぐために、事実や意見を実際に述べる「**言及**」を見つけたうえで、言及に対してのみ評価の検出を行うことにする。原文を構文解析した結果をもとに、以下のような手順で、文の主節(日本語の場合は文の右端)から言及となる部分を探索する。

- 用言の言い切りになっているものは言及である。また、一部の終助詞などは外して考えるため、例文(3)・(4)の末尾も言及である。

- 探索している部分に、順接・逆接・原因などの関係に係る従属節(日本語なら左側の部分)があれば、そこから再帰的に言及を探す。たとえば、例文(1)の「番組表は便利だ」という部分も言及とみなされる。
- 探索している部分が「…と思う<sup>☆2</sup>」「…という傾向がある」「…に違いない」などの規定された表現であるとき、「…」の部分から再帰的に言及を探す。
- 探索している部分が「…ない」など、否定の表現となっているなら、「…」の部分で本来と逆の極性を持つものとして、再帰的に言及を探す。

言及を探索手順の例を図-3に示す。文の末尾から順に言及を探し、上記の規則に合う場合には言及を探索範囲が広がる。言及(枠で囲まれた部分)に対して極性辞書と比較し、相当する極性語句が見つければ、評価であると判定される。例文(5)中の条件節や、例文(6)中の連体修飾句に関しては、言及とみなされない(探索対象とならない)ため、誤った評価の検出を防ぐことができる。

なお、例文(7)では、「ギャグが面白い」という好評の表現が含まれるものの、その右側にある「思う」が否定されている。本手法では、否定された言及の中にある「ギャグが面白い」もまた否定の言及とみなすことにより、逆の極性(不評)が正しく出力される。このように、文の末尾から構造を辿ることによって、間接的な否定、二重否定なども正しく扱うことができる。

(7) ギャグが**面白い**とは全然思わなかった。

一般に、言及単位の評判分析のように、特定の基準に合致したものを出力する処理では、正確さ(出力に誤りがないか)とカバレッジ(いかに多くの出力が得られるか)はトレードオフの関係にある。先に述べた通り、言及単位の評判分析では、正確さが特に重視される。その点、この手法のように、文中の限られた場所のみを解析対象とすれば、正確な出力が得やすい。また、言及を探索するための規則を慎重に追加していくことによって、拾い出せる評価を増やすことができる。

語句の知識を得る

表-1に挙げた極性語句は、解析対象の分野によらずに普遍的に使えるものである。しかし、実際の解析のためには、扱う分野に特化した極性語句が必要となる。カメラの分野の例文(8)、映画の分野の例文(9)を見てみよう。

(8) 持ってみると**重さを**ずいぶん**感じて**しまう。

(9) ラストシーンはかなり**泣ける**。

例文(8)の「重さを感じる」はカメラの良くない性質、すなわち不評を示唆している。例文(9)の「泣ける」は、

☆2 意見を述べている文を解析する際には、「…だ」と言い切った文と、「…だと思う」という表現は同一視しても構わない。

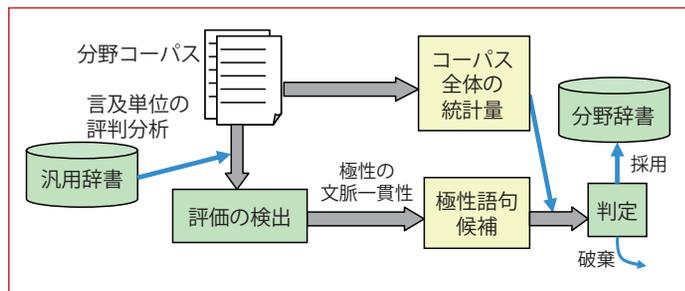


図-4 分野辞書の自動生成の流れ

他の多くの分野では不評として捉えられるであろうが、映画の分野においては好評となる。

このような分野別の知識を持っていれば、検出できる評価が増えるだけでなく、具体的な性質を知ることができるようになるため、分析の際の有用性が増す。しかし、複数語からなるものも含め、分野ごとに非常に多くの種類の極性語句を登録せねばならず、人手で行うとなると膨大な作業量となる。そこで、分野辞書の生成、すなわち、好評・不評に関連する分野依存の知識獲得を自動化したい。

ここでは、文献2)で提案されている自動的な知識獲得の手法について、簡単に解説する。図-4に示すように、まず、初期の辞書(汎用辞書)を用いて、解析対象となる分野のコーパスから評価を検出する。次に、極性の文脈一貫性という性質を用いて、新たな極性語句の候補を取り出す。そして、統計を用いて適切なものを選び出す判定をすることによって、分野辞書を自動的に作成する。

### 【極性の文脈一貫性】

テキストの中で述べられている人々の意見を観察すると、以下のような傾向があることに気付く。

- 好評→好評、不評→不評と、同じ極性の評価を連続して述べることが多い

(10) 紫色がくっきりしていて、満足しています。  
 (11) 導入部が長すぎる。見ていてがっかりした。

- 好評・不評の両方を述べるときには、その間に逆接を表す語句が挿入されやすい

(12) 役者はうまくないけど、引き込まれた。  
 (13) 画面が綺麗！ただ、ボディが大きい。

このように、同じ評価の極性を持つまとまった文脈が形成されて、逆の極性を持つ文脈の間には明示的に逆接の語句が使われるという傾向がありそうだ。この仮説(極性の文脈一貫性と呼ぶ)に基づくと、「😊満足する」「😞がっかりする」「😊うまい」「😊綺麗だ」が普遍的

な極性語句として汎用辞書に含まれていれば、文脈から、カメラの分野では「😊くっきりする」「😞ボディが大きい」、映画では「😞導入部が長い」「😊引き込まれる」が分野依存の極性語句であることを自動的に発見できるのではないかな。

### 【極性語句の候補を抜き出す】

分野辞書に追加すべき極性語句を得るために、まず、汎用辞書を用いて評価を表す言及を検出する。次に、検出された評価の文脈から極性語句の候補を抽出するのだが、「文脈とはどの範囲か」「辞書化すべき極性語句はどの部分か」を決めることが問題となる。

まずは文脈の範囲について考えよう。極性の文脈一貫性は、例文(10)・(12)のように、1つの文にある2つの言及の間だけでなく、例文(11)・(13)のように、連続する2文の間でも成り立つと考えられる。そこで、以下のようにして、汎用辞書から検出された言及(以降、既定の評価と呼ぶ)と、文内で隣接する言及、および文をまたがって隣接する言及が極性を持つと仮定する。

- 既定の評価と、同じ文の中で、順接・並列・因果関係で結ばれる言及(例:「素晴らしいし、…だ」「…なので、素晴らしい」)は、同じ極性を持つ。一方、逆接・譲歩で結ばれる言及(例:「…であるものの、素晴らしい」)は逆の極性を持つ
- 既定の評価が文の主節である場合、その前後の文の主節である言及は同じ極性を持つ。ただし、逆接の接続詞が存在する場合には逆の極性を持つ

図-5は、文脈を利用して、既定の評価と隣接する言及の極性を仮定する操作の例である。ここでは、「満足しています」という既定の評価があるときに、前後の文の主節と、文内の2つの従属節にある言及が極性を持つと仮定される。たった1つの好評の表現だけから、好評の言及2つと不評の言及2つが見つかることや、文脈の切り替わりは言及の表層上の位置だけではなく、文の構造が関係している<sup>☆3</sup>ところなどが興味深い点である。

次に、極性を持つと仮定された言及から、辞書化される単位、すなわち極性語句の候補を決める必要がある。例文(10)では、上記の方法により、「紫色がくっきりする」という言及が好評であると仮定されるが、当該分野において「緑がくっきりする」「画像がくっきりする」なども好評を表すので、それらにも共通して適用できるよう「😊くっきりする」という極性語を得たい。一方、例文(13)では、「液晶が大きい」なら好評であったり、「影響が大きい」の極性は不定だったりするため、「😞大きい」ではなく「😞ボディが大きい」という極性語を得たい。

どの単位が望ましいかはまだ判断できないので、い

☆3 「画素数が多い」と「液晶が小さい」は、出現位置は近いが、文の構造を考慮すると同じ文脈に入っていないことが分かる。

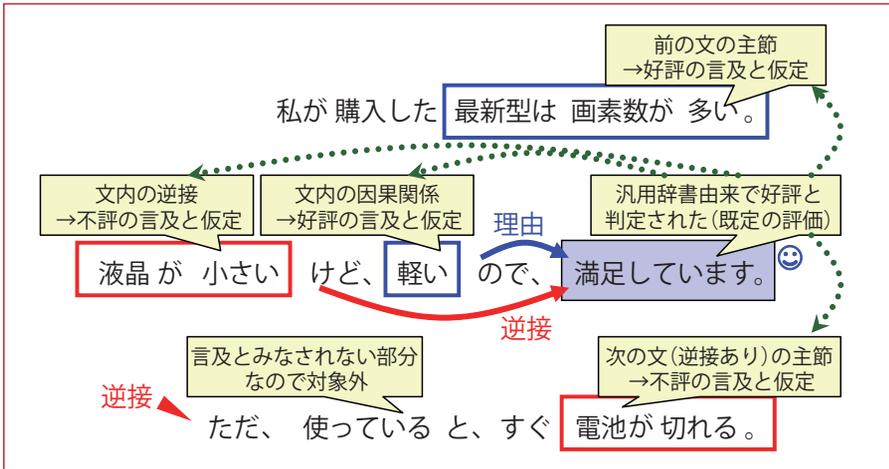


図-5 文脈を用いて隣接する言及の極性を仮定する

	全体頻度	😊 好評頻度	😞 不評頻度
しっかりする	246	54	10
小さい	3,014	226	227
ボディーが   小さい	11	4	0
木曜ーに   届く	2	0	2

表-2 デジタルカメラ分野で抽出された極性語句候補の例

ったんは極性語、極性句（状況によって複数）の双方を候補として抽出しておく。すなわち、例文(10)では「😊 くっきりする」「😊 紫色ーが | くっきりする」の両者、例文(13)では、「😞 大きい」「😞 ボディーが | 大きい」の両者が極性語句の候補となる。

【極性の文脈一貫性の不完全さ】

極性の文脈一貫性の仮定は直感的ではあるものの、常に成り立つものではない。実際には以下のように例外的な事象が多く見つかる。

- (14) 私は新機種を買った。とても良いです。
- (15) Aは好き。Bは嫌い。

(14)の例では、2文目では好評を述べているが、1文目は買ったという事実であり、評価を表してはいない。また、(15)の例のように、ぶっきらぼうに好評と不評を並べることもある。他にも、皮肉のような書き方もあり得るし、既定の評価に誤りが含まれることもある。

したがって、候補として現れた表現がすべて辞書化できるわけではなく、大量の文から得られた候補の中から統計的操作によりノイズを除去して、“本物の”極性語句を選択する必要がある。

【候補の抽出例】

表-2は、デジタルカメラの分野のコーパスから抽出された極性語句の候補の例である。「全体頻度」とは、それぞれの語句がコーパス全体で出現した頻度である。また、「好評頻度」「不評頻度」は、既定の評価と隣接する言及の中にあるために、それぞれ好評・不評であると仮定された場合の頻度である。

直感的には、好評・不評のいずれか片方の文脈のみで、かつ全体の出現に対して高い割合で現れるものが、極性語句であるといえそうである。すなわち、好評の語句と判定できるものは、次の2つの指標が高いものとなる<sup>☆4</sup>。

- 極性一致率： $\frac{\text{好評頻度}}{\text{好評頻度} + \text{不評頻度}}$
- 極性保持率： $\frac{\text{好評頻度}}{\text{全体頻度}}$

「しっかりする」は好評の極性語といえそうだが、その極性保持率は2割程度しかない。また、好評の5分の1以下の頻度とはいえ、不評の文脈でも観測されており、極性一致率は8割程度である。これらの値から、好評の極性語だとみなして辞書に加えてもよいものだろうか？

「小さい」は明らかにどちらともいえず、実際に好評・不評の頻度が均衡している。一方、「ボディーが | 小さい」は好評を表すであろう。すなわち、「小さい」がノイズとして破棄されて、「ボディーが | 小さい」が辞書に採用されれば、極性語・極性句のどちらが使われるべきかが判別できたことになる。

「木曜ーに | 届く」は、明らかに評価と関係のない表現であるが、2回の出現のうち2回が不評の文脈で観測されているため、極性一致率・極性保持率とも100%となっている。このようなノイズを除去するためには、最低出現頻度などを恣意的に決めることになるが、何回出現すれば辞書に加えてよいかは、用いるコーパスの質や量に大きく依存する。

☆4 もちろん、不評の語句の判定には、2つの式的好评頻度と不評頻度を入れ替えたものが用いられる。

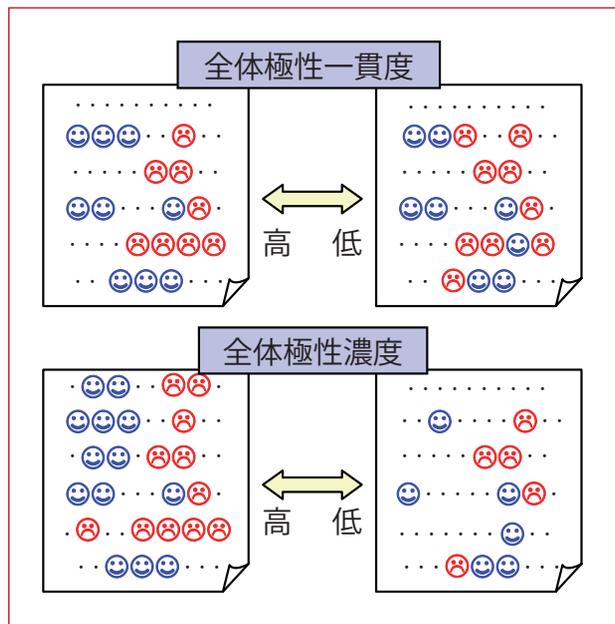


図-6 全体極性一貫度・全体極性濃度の概念図

【コーパスの特性を調べる】

そこで、用いるコーパスにおける評判の述べられ方の特性を調べるために、全体極性一貫度・全体極性濃度の2つの指標を導入する。いずれも、汎用辞書を使ってコーパス中で既定の評価を検出した状態で測定する。

全体極性一貫度は、既定の評価が隣接している場合のうち、互いの極性が「好評-好評」、「不評-不評」、「好評-（逆接）-不評」などと、仮説に矛盾しない並びとなっている割合であり、好評・不評がどれだけ自然に並び立てられているかを示す。

全体極性濃度は、既定の評価すべてに対して、隣接する言及の中に同じ極性（逆接で結ばれているときは逆の極性）を持つ既定の評価がある割合である。すなわち、どれだけ頻繁に評価を述べているコーパスであるかを示す。

図-6は、それぞれの指標が高い状態・低い状態、コーパス中の好評（😊）・不評（😞）の分布で図示したものである。ただし、簡単のために逆接の語句はないものとしている。

実際のコーパスで測定してみると、全体極性一貫度は分野により72～76%と大きな差はなかったが、全体極性濃度は7～18%と、文体や扱うテーマによって差が見られた。特に、映画分野の全体極性濃度が高かった。

【統計を用いて極性語句の判定をする】

ここまでで導入した指標を用いて、極性語句の候補を辞書に入れるべきかどうかの判定基準を定めよう。

全体極性一貫度は、用いるコーパス中における、汎用辞書に登録されている極性語句全体の極性一致率とみな

分野	😊 好評	😞 不評
デジタルカメラ	明るい ケラレ-が   ない	重さ-を   感じる 修理代-が   高い
映画	緊張感-が   ある 色っぽい	導入部-が   長い 興味-が   半減する
携帯電話	液晶-が   大きい 消費電力-が   低い	反応-が   鈍い 形状-が   違う
自動車	安定感-が   増す 振動-が   ない	曇る ドア-が   薄い

表-3 得られた極性語句の例

すことができる。同様に、全体極性濃度は、極性語句全体の極性保持率と近似できる。そこで、それぞれの極性語句の候補について、観測された全体頻度・好評頻度・不評頻度を用いて推定を行い、

- 極性一致率がコーパスの全体極性一貫度を90%の確率で超える
- 極性保持率がコーパスの全体極性濃度を90%の確率で超える

の2条件が満たされるときに分野辞書に加えて、そうでないときには破棄することとする。

これにより、恣意的な閾値を設定することなく、コーパスの性質に応じた基準を用いて、辞書に追加すべきか否かを判定し、かつ極性語・極性句のいずれを登録するかを同時に決定することが可能となる。

先ほどの例で見た各候補では、「😊 しっかりする」および「😊 ボディー-が | 小さい」が上記の条件を満たし、他「小さい」「木曜-に | 届く」は条件を満たさなかった。すなわち、ここでは正しい選択が行われている。

【得られた極性語句の例】

4つの分野に対して上記の手法を適用した。獲得された極性語句の例を表-3に示す。各分野において特有の語句が獲得されている。特に、デジタルカメラ分野の「ケラレがない」など、専門用語を含む表現も獲得できており、この手法の有効性が現れている。各分野で228～708個の語彙が得られ、新しい分野の辞書を用いて検出される評価は、92～97%の正解率を持つと測定された。

嗜好の変化や違いを分析する

前章で紹介した手法では、極性語句は分野ごとに異なるものの、1つの分野では一定であるとして、当該分野の評判分析に用いるための分野辞書を自動的に獲得した。しかし、同じ分野であっても、時代や地域・性別などによって、好き嫌いに違いが出る場合がある。以下で紹介する嗜好分析は、前章でみた極性の文脈一貫性を用いて、そのような違いを検知することを目的としており、マー

好評が増加	液晶-が   大きい, 操作-が   簡単だ
好評が減少	画質-が   よい, サイズ-が   小さい
不評が増加	三脚穴-が   ない, ノイズ-が   出る
不評が減少	起動-が   遅い, ピント-が   甘い

表-4 デジタルカメラに関する掲示板における2002年・2004年の嗜好の比較

ケティング等のために強く求められている処理である。

「良い」「嫌いだ」「美味しい」「汚い」のように良さ・悪さ自体を示す表現の場合、通常はその極性は不変である。一方、間接的に評価に繋がるような事物の性質は、評価する人の属性や時期によって、評価の傾向が変わってくることもある。たとえば、ビールに関して「苦くて美味しい」「苦くて飲みづらい」と意見が分かれるように、「苦い」という性質は良い要素・悪い要素の両方となり得る。このような嗜好の違いを捉えるのが嗜好分析の目標である。

まず、製品等に対する意見が述べられている文書を、時間軸や評価者の属性によって分類する。分割された文書群から、極性語句の自動獲得のために候補を出力したときと同様、極性の文脈一貫性を用いて、既定の評価と隣接する言及の中にある表現を取り出す。ただし、ここで規定の評価を得るときに用いる辞書は、上に例を挙げたような直接的な評価を表す語句に限定する。すると、文書群ごとに出現の傾向が異なる語句があり、それを見ると嗜好の変化や違いが浮き上がってくる。

実験として、デジタルカメラに関する掲示板のデータのうち、2002年と2004年の書き込みから、好評・不評の極性語句の候補を抽出した。その結果、2002年と2004年における出現の傾向が異なる語句があった。

表-4は、その中から主要なものを人手で選択した結果である。

表-4からは、デジタルカメラの技術の進歩に従い、

ユーザが求める性質、避けようとする性質が変化している様子が伺える。他の分野においても、人々が何を重視して商品等を選択しているかの傾向を察知できると期待できる。

### 「意味」を扱う研究

本稿では、自然言語処理技術を用いて評価や嗜好を捉える手法について述べた。これらは、表層上の単語のパターンマッチングや、位置が近い語を取り出すといった単純な処理ではなく、構文解析や格の解析といった文法に基づいた操作を必要としている。言及単位の評判分析の出力は、テキスト中から抽出する一種の論理的な意味表現であり、評価の極性という明確な軸を与えたことにより、そのまま産業分野で応用できるものである。従来は漠然としていた「意味」の扱いが現実的なものに近づいてきたといえるのではないかと。

また、嗜好分析の技術は、時代とともに刻々と変化する語句の意味を捉えるものであり、「生きている言語」を扱う研究の第一歩と位置づけられる。

### 参考文献

- 1) Kanayama, H., Nasukawa, T. and Watanabe, H. : Deeper Sentiment Analysis Using Machine Translation Technology, In Proceedings of the 20th International Conference on Computational Linguistics (COLING), pp.494-500 (2004).
- 2) Kanayama, H. and Nasukawa, T. : Fully Automatic Lexicon Expansion for Domain-oriented Sentiment Analysis, In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.355-363 (2006).

(平成19年7月23日受付)

### 金山 博

hkana@jp.ibm.com

2000年東京大学大学院理学系研究科情報科学専攻修士課程修了。同年より日本アイ・ビー・エム(株)東京基礎研究所に勤務、現在に至る。構文解析・意味解析など、自然言語処理の基礎に興味を持つ。