

要望表現の抽出と整理

金山 博 那須川 哲哉
日本アイ・ピー・エム株式会社 東京基礎研究所
{hkana,nasukawa}@jp.ibm.com

1 はじめに

テキストマイニング [3] の一形態として、アンケートに対する回答・web 掲示板における発言といった文書群から製品やサービス等に関する評判を示す記述を抽出する「評判分析」「Sentiment Analysis」の技術への注目が高まっている [4, 5, 2]。特に、好評・不評の意見（以下評判表現と呼ぶ）の抽出・分類・整理は、対象物の利用者・提供者の双方に有益な情報を提供する。例えば、ある製品の購入を検討する消費者は、候補となる製品の長所・短所を比較して、嗜好に合った製品を探すことができる。また、あるサービスを提供している企業にとって、自社と他社の評判の違いを把握することは不可欠である。

我々は文書中の評判表現を正規化された命題の形に変換するシステムを作成している [1]。例えば、携帯電話に関する記述 (1) を、意図フラグ (好評・不評) と命題 (主に、用言および格) の組で表す構造 (2) に変換する。この構造を意図フレームと呼ぶ。

- (1) まあ、液晶はなかなかきれいだと思うよ。
- (2) [好評] 綺麗だ (液晶)

評判表現は、対象物についての評価を表し、テキストデータ上で述べられた製品やサービスの善し悪しを数値的に比較する際の手がかりとなる [11, 7]。

一方、評判分析の対象となるテキストには、(3) や (4) のような要望を示す記述も見受けられる。これらは評判表現とは異なり、現状の評価を述べるものではないため、好評・不評の頻度の集計の際には除外すべきものである。

- (3) 液晶をきれいにしてください。
- (4) 液晶が綺麗だと嬉しいんだけど。

このような要望表現は、将来の製品・サービスに求められているものを直接的に知ることができる、非常に有益な情報である。しかし、評判表現のように、特定の用言句で表現されず、「～してほしい」「～たい」等のモダリティの文末表現によって意図を示すことが多いため、内容の自動的な整理が難しい。特に、自由回答アンケート上では多くの要望が言及され、大塚らがその同定方法について研究している [8, 10] が、要望されている事柄の整理の自動化には至っていない。

そこで、従来は評判情報を主体に扱ってきた評判分析システムにおける要望表現の扱いを見直し、新しい「評判・要望分析システム」を実現した。本稿では、要望表現を整理するために好適な意図フレームの設計と、その構築方法について論ずる。提案する要望表現の意図フレームは、要望されている内容を命題で表し、モダリティを「要望」「要求」のフラグを用いて簡潔に表したものである。例えば、上記の (3)(4) の要望表現を (5) と表す。

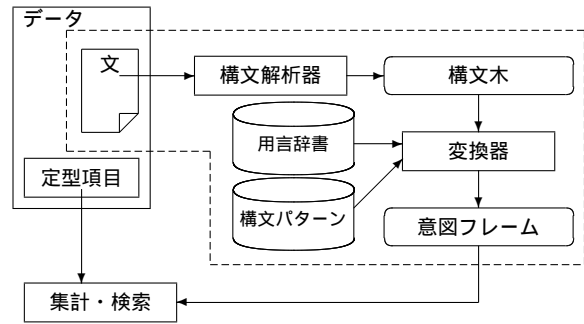


図 1: 評判・要望分析の処理の流れ。本研究のスコープは点線の枠内の、文中の表現を意図フレームに変換する部分。

- (5) [要望] 綺麗だ (液晶)

意図フレームの再設計と言語解析部分の見直しによって、求められている事柄の発見を支援するために必要な、「大規模なデータから要望表現を高精度で抽出すること」「多様な形式で書かれた同種の事象をまとめ上げること」が可能になったことを実験結果によって示す。

2 評判・要望分析の基本方針

本研究で構築する評判・要望分析システムの処理の流れを図 1 に示す。図中左下の「集計・検索」が利用時の機能であり、主に次の二つの目的がある。

1. 評判・要望の絶対数・相対数の集計。対象物の全体的な好評・不評の割合（例：A 社の携帯電話の評判は $x\%$ が好評）を知るだけでなく、好評・不評の根拠となる具体的な表現の頻度（例：「音質が悪い」という記述が N 件）を知りたい。
2. 評判・要望の検索。あるキーワードに関する評判や要望の一覧を表示し、深く調べる必要がある項目について原文を参照したい。例えば、「字が大きい」という要望を見た時に、「高齢の親に画面を見せるために字を大きくしてほしい」という具体的な記述を含む原文を検索する操作。

上記の 1 番目の用途のためには、好評・不評の総数を数えるだけでなく、同じ意図を示す記述を一つの構造に正規化した上で集計できる必要がある。また、2 番目の用途のように、個々の情報を参照する場面が想定され、その際に提示された例が明らかな誤りであると、システ

ムの信頼性が著しく損なわれるため、再現率よりも適合率を重視するのが妥当である。

本稿では、図1の点線枠内、すなわち、以上の目的に適うような意図の表現形式(意図フレーム)を生成する処理に焦点を当てる。なお、より詳しい解析のためには、アンケートや掲示板のタイトルなど、テキストに付随する定型項目や、文脈上のキーワードを利用する必要があるが、それらは意図フレーム構築とは独立の処理で行うこととし、議論の対象外とする。

3 意図フレームへの変換

本研究にて用いる意図フレームは、意図フラグ([好評][要望]など)と命題の組である。命題は、一般的には任意の部分木であるが、通常は「述語(格1, 格2, ...)」の形で表す¹。例として、文集合(6)に含まれる二つの評判表現は、意図フレーム(7)(8)へと変換される。

- (6) 音質のほうは気に入ったものの、
デザインが素敵だとはとても思えない。
- (7) [不評] 素敵だ +ない (デザイン)
- (8) [好評] 好む (音質)

上記の意図フレームを出力するにあたって、評判表現を表す語句を抽出するだけでなく、いくつかの統語的な変換がなされている。我々はこの変換を、日本語の文から意図フレーム集合への翻訳と捉えて、機械翻訳と共通の処理を行っている[1]。以下では、評判の意図フレームを生成する際に用いる用言辞書と、構文パターンを用いた構文木上でのトップダウンの解析について解説する。

3.1 評判分析のための用言辞書

用言辞書は、変換元の格フレームと変換先の意図フレームの組である。図2は、動詞句「～を気に入る」を好評の意図フレームに変換する辞書項目であり、「気に入る」の述語を「好む」に置き換えることで、命題部分が「入る(音質, 気)」となって「気」が格として扱われるのを防いでいる。また、格となる名詞の意味属性等をもとに好評・不評の属性を変えたり、格解析をする²機構も備えている。用言辞書は、機械翻訳の辞書をもとに人手で作成したが、我々は分野に依存する語彙を自動獲得する手法も考案している[9]。

3.2 構文パターンとトップダウンの解析

再現率よりも適合率を重視するため、文内の任意の場所からフレーズを抽出するのではなく、原則として各文の主辞からのみ意図フレームを生成する。従って、「音質が気に入ると仮定する。」という記述中の「音質が気に入る」は意図フレームに変換されない。

但し、主辞からトップダウンに構文パターンでカバーされる場合は、子となる部分木の探索をする。上記(6)

¹格助詞は命題形式での出力時には表示しないが、内部的には区別しており、文生成の際に用いられる。

²例えば、(6)の副助詞「は」の句が、図2のヲ格を埋める。

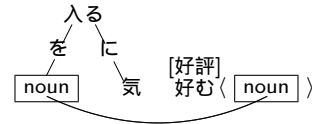


図2: 「Nを気に入る」から好評の意図フレーム「好む(N)」を得るための用言辞書の項目の例。左辺が入力とマッチする部分木、右辺が意図フレームの断片で、弧は対応関係を表す。なお、「気に入る」の主体となるヲ格は、通常は評価の対象となり得ないため、意図フレームの中にも含まない。

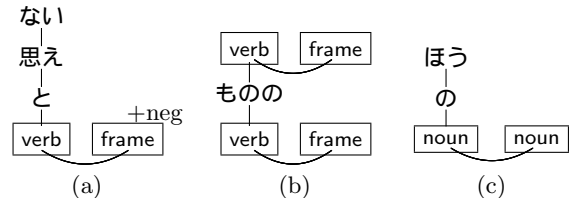


図3: 構文パターンの例。(a)(b)は主辞以外の用言を適切に抽出するためのもので、(c)は命題中の格となる語句の情報量を高める役割を持つ。[verb]は用言、[frame]は他のパターンで構築された意図フレームの断片。

の場合、主辞の「～と-思え-ない」は、図3(a)の構文パターンとマッチし、「と」の下の部分木から得られる評判表現の極性を反転したものが出力される。同様に、接続詞「ものの」を扱うためのパターン(b)も登録されているため、「音質のほうは気に入った」の部分が探索の対象となる。

このほか、図3(c)の構文パターンを用いて、「気に入る」のヲ格に入る名詞句の主辞である形式名詞「ほう」が意図フレームに現れるのを避けている。

3,762 エントリの用言辞書と131の構文パターンを用いて、Webの掲示板テキストをから評判表現を抽出したところ、再現率は43%にとどまっているが、89%という高い適合率が得られている[1]。

4 要望表現の意図フレーム

要望表現とみなす言語現象について、社内システムに関するアンケート、Web上で収集した掲示板のデータに頻出する要望表現や、モダリティ表現に関する文献[6]等を参考に、以下のタイプを考えた。

- A: 要望を示す用言 「～が欲しい」「～を期待する」など、それ自身が要望を示す用言による記述。
- B: 依頼・命令・願望 「～して下さい」「～しろ」など、読み手に対して直接的に働きかける表現。また、「～て欲しい」「～てもらいたい」のように、他者の動作を希望する表現。
- C: 当然・条件付好評 「～べきだ」のほか、「～ば嬉しい」「～と望ましい」のように、望ましい状態や動作を示す表現。

これらを表す構造として、評判分析で用いていた従来型の意図フレームについて述べたのち、要望分析に適した改良型意図フレームを提案する。

4.1 従来型意図フレーム

上記の分類のうちAは、評判表現と同様に、用言+格の命題の形式で意図フレームを構築できる。例えば、文(9)は意図フレーム(10)で表現される。

- (9) メモ帳機能が欲しい。
- (10) [要望] 欲しい (メモ帳-機能)

一方、B・Cは、要望表現と同定する根拠がモダリティの部分にある。従って、意図フレームの命題の部分に「要望」という意図をもたせるべく、1節の例(3)(4)に対する意図フレームを設計すると、以下のような複雑な構造にせざるを得ず、まとめ上げに不向きである。

- (11) [要望] 下さい (綺麗に-する (液晶))
- (12) [要望] 望む (綺麗だ (液晶))

4.2 要望解析に向けた改良型意図フレーム

構造を簡略化して扱いやすくするために、抽出・整理された要望表現を利用する側、すなわち対象となる製品やサービスの提供者の立場から整理し直す。そのために、命題部分を望まれている事物に統一し、以下の[要望][欲求][所望対象]という3種の意図フラグを導入する。

要望 「～て欲しい」「～て下さい」「～てもらいたい」のような読み手への依頼のほか、「～べきだ」「～と望ましい」「～ことを期待する」等、望ましい状況を記述するものをすべて[要望]とする。

- (13) 説明書と用語を統一してもらいたいです。
[要望] 統一する (説明書, 用語)
- (14) 字がもっと大きいとありがたい。
[要望] 大きい (字)

欲求 「～たい」に加えて、「～できるようにして下さい」など、書き手がしたいことを示す表現に[欲求]のフラグを付ける。

- (15) 色を変更できるようにして頂けますか。
[欲求] 変更する (色)

所望対象 要望の対象が名詞的に表現できるものを[所望対象]としてまとめる。その際、連体修飾する属性があれば付与する。上記(9)もこれに当たる。

- (16) ガイドがあると良いと思います。
[所望対象] ガイド
- (17) 高い解像度を期待しています。
[所望対象] 解像度 (高い)

命題部分を大幅に簡素化し、内容のみを表すようにしてあるが、[要望]と[欲求]を区別することによって、命題で表現される事柄が書き手が欲する動作なのか否かの情報が失われない³。これは「提供者がなすべきこと」「利用者がしたいこと」「利用者が欲しいもの」という、製品開発等の際に考えるべき視点に即した分類である。さらに、意図フレームから直感的な言語形式を生成することもできる。例えば、上記(15)の場合、「色を変更したい」という文を生成できる。

³ 「写真を撮って欲しい」と「写真を撮りたい」が区別される。

[所望対象]について、「ある(ガイド)」「高い(解像度)」のように、用言による命題形式でも表すことができるが、名詞的に表現する方が簡潔であり、望まれているものを把握しやすくなると考え、可能な限りは[所望対象]の意図フレームに変換することとした。

なお、「～を避けて欲しい」「～まい」などに関しては、負の要望表現として命題の部分否定形とする。

- (18) 絵文字を表示するのはやめて下さい。
[要望] 表示する +ない (絵文字)

4.3 その他の改良

評判表現の抽出の際には、ほとんどの場合時制の違いを無視していたが、要望表現の場合には、助動詞「た」の有無が意味を持つことがあるため、以下のような場合には区別するようにした。

- (19) 説明に時間を割いて欲しい。
[要望] 時間を割く (説明)
- (20) 説明に時間を割いて欲しかった。
[不評] 時間を割く +ない (説明)

また、用言によって、要望表現か、単なる(掲示板の他の読み手などへの)示唆かが異なる場合がある⁴ため、意志性の有無の属性⁵を用いて分類するようにした。

- (21) Aの機能を使うと便利です。φ
- (22) 光学ズームが付いていると便利です。
[要望] 付く (光学ズーム)

5 実験

5.1 要望表現抽出の精度

4.1節で述べた従来型の意図フレームへの変換器と、4.2～4.3節の拡張を施した改良型意図フレームへの変換器を実装し、社内システムに関するアンケートのデータ3,905文からの要望表現の抽出の性能を、以下の3つの観点で評価した。

適合率 抽出された要望表現のうち、アンケートの読み手にとって意義のあるような書き手の要望を表したものである割合。

再現率 アンケートの文章を見て人手で付けた正解のうち、正しく抽出された要望表現の割合。ただし、本研究では再現率を重視しないことと、人間によっても判定が難しいものが多い[8]ことから、200文のみで評価した。

可解率 正しく同定された要望表現のうち、意図フレームの命題・名詞句を見て、求められているものを推測することができるものの割合⁶。

⁴ アンケートの回答と異なり、掲示板の発言など複数の読み手を想定しているテキストでは、「～下さい」などの表現の多くは要望表現とならないので、構文パターンの切り替えが必要である。

⁵ 意志性の有無は、「用言+ために」が目的と原因のいずれを表すかという基準とほぼ一致する。

⁶ 「形が悪いので改善してください。」から、[要望] (改善する) を出力しても、何を改善すればよいかを把握できない。このような場合が「可解でない」と判定する例である。

	従来型	改良型
適合率	96.4% (486/504)	97.4% (517/531)
再現率	48% (20/42)	60% (25/42)
可解率	39.7% (200/504)	49.3% (262/531)

表 1: 改良前後における要望表現の適合率・再現率・可解率の評価。

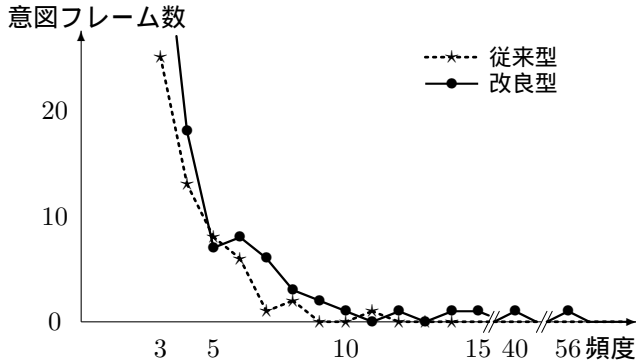


図 4: 改良前後の意図フレームの頻度別分布。

結果を表 1 に示す。97.4% と高い適合率が得られており、実用上は問題無いレベルである。誤って要望表現と判定した例は、「～と言いたい。」という意見を [欲求] としたり、形態素解析の誤りによって命令文と判定されるような、些細なもののみであった。

再現率は 60% と、評判表現抽出の場合を上回っている。要望表現だと判定できなかった例は、「～ではなかろうか。」などの未登録のパターンの下に要望が記述されていたり、「～を求む」といった未登録の用言によるものなどであるが、いずれも辞書整備により解決される。

可解率が大幅に向上されたのは、要望内容を示す命題や、修飾語を含む所望対象を用いて表現したことにより、意図フレームに含まれる情報量が増えたことによる。

5.2 まとめ上げの効果

図 4 は、改良前後の意図フレームの頻度分布である。従来型では表現形式が異なるために別の意図フレームとして表現されていたものが、改良後には同じ命題を持つものとして同一視されているため、頻度の高い意図フレームが増加している。従って、頻繁に言及されている要望の把握が容易になるといえる。

また、38,904 文中の [要望] 4,360 件のうち 230 件については、命題部分が一致する [好評] の意図フレームが同じ文集合の解析結果中に見つかり、[不評] と一致する例は無かった。一方、[欲求] の 662 件うち [好評] として出力されるものと一致する例は「改善したい。」のような例が 3 件あるだけであり、多くは好評の要素となるものとは異なっていることがわかる。このような関連性から、[要望] の命題のうち、[好評] として辞書登録していないものから、将来的に高評価を与えるきっかけとなる新しい観点を発見できると思われる。

6 まとめ

テキスト中の要望表現を抽出し、簡素化された意図フレームに変換することにより、求められている事柄を適切にまとめ上げる手法を提案した。我々の手法では、構文パターンの適用、用言辞書の場合分けなどにおいて柔軟性があるため、これらをより洗練させ、テキストの種類や対象分野に即した辞書を整備（機械学習による自動構築を含む）することにより、多彩な形式の評判・要望表現を、情報量を増やした形で、より正確に抽出することが期待される。

要望表現と評判表現を共通の命題構造に変換することにより、物事の性質を現状と理想形との間で比較する土台ができた。また、本研究は、評判・要望の把握という目的を定めた上で、同じ意味を持つ複数の言語現象を同一視することを目指すものであり、実用的な意味解析に関する研究と位置づけることができ、今後広く発展していく見込みがあると考えている。

参考文献

- [1] Hiroshi Kanayama, Tetsuya Nasukawa, and Hideo Watanabe. Deeper sentiment analysis using machine translation technology. In *Proc. 20th COLING*, pp. 494–500, 2004.
- [2] Satoshi Morinaga, Kenji Yamanishi, Kenji Tateishi, and Toshikazu Fukushima. Mining product reputations on the web. In *Proc. of the 8th ACM SIGKDD Conference*, 2002.
- [3] Tetsuya Nasukawa and Tooru Nagano. Text analysis and knowledge mining system. *IBM Systems Journal*, Vol. 40, No. 4, pp. 967–984, 2001.
- [4] Tetsuya Nasukawa and Jeonghee Yi. Sentiment analysis: Capturing favorability using natural language processing. In *Proc. of the Second International Conferences on Knowledge Capture*, pp. 70–77, 2003.
- [5] Jeonghee Yi, Tetsuya Nasukawa, Razvan Bunescu, and Wayne Niblack. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of the Third IEEE International Conference on Data Mining*, pp. 427–434, 2003.
- [6] 益岡隆志. モダリティの文法. くろしお出版, 1991.
- [7] 乾孝司, 乾健太郎, 松本裕治. 出来事の望ましさ判定を目的とした語彙知識獲得. 言語処理学会第 10 回年次大会発表論文集, March 2004.
- [8] 大塚(乾) 裕子, 伊佐原均. アンケート回答に現れる要求意図の認定に関する分析. 言語処理学会第 10 回年次大会発表論文集, March 2004.
- [9] 那須川哲哉, 金山博. 文脈一貫性を利用した極性付評価表現の語彙獲得. 情報処理学会第 162 回自然言語処理研究会, July 2004.
- [10] 内山将夫, 大塚(乾) 裕子, 伊佐原均. 自由回答アンケートにおける要求内容の分析. 言語処理学会第 10 回年次大会発表論文集, March 2004.
- [11] 立石健二, 福島俊一, 小林のぞみ, 高橋哲朗, 藤田篤, 乾健太郎, 松本裕治. Web 文書集合からの意見情報抽出と着眼点に基づく要約生成. 情報処理学会 163 回自然言語処理研究会, September 2004.