

ユーザレビュー情報における感情表現の分析手法

小林 大祐[†] 井上 潮[‡]

^{† ‡} 東京電機大学工学研究科 〒101-8457 東京都千代田区神田錦町 2-2

E-mail: [†] 08gmc13@ms.dendai.ac.jp, [‡] inoue@c.dendai.ac.jp

あらまし ユーザレビュー情報は閲覧者にとって有用な情報となり得る。しかし、ユーザレビューは質がさまざまなため、レビューの有用性を判断する必要がある。ユーザレビューの数が増えると閲覧者の負担が大きくなる。そこで、閲覧者の負担を軽減することを目的として、レビュー文書の有用性の自動判別を行う手法を提案した。提案手法はレビュー内容の關係に着目した。レビュー対象商品の客観的情報である「特性」とユーザの主観情報である「態度」と、これらの相互關係にある「評価」を定義し、3つの要素の出現傾向により文章の有用性を自動的に判別する。3つの要素の自動判定には形態素情報を用いる。提案手法の有効性を検証するため、Amazon レビューを対象として、自動抽出した結果と人手による抽出を行った結果を比較した。

キーワード 情報抽出, レビュー解析, Web 情報システム

Analysis of Expression of Feeling in User Review Information

Daisuke KOBAYASHI[†] and Ushio INOUE[‡]

^{† ‡} Tokyo Denki University 2-2 Kanda-Nishiki-cho, Chiyoda-ku, Tokyo 101-8457 Japan

E-mail: [†] 08gmc13@ms.dendai.ac.jp, [‡] inoue@c.dendai.ac.jp

Keyword information extraction, web information system

1. はじめに

レビューとは、ある商品に対して意見や評価をすることである。レビューは商品の利用者（ユーザ）が書く情報なので、ここではユーザレビュー情報と呼ぶ。意見とは、自分の考えを述べることで、感想や喜怒哀楽の表現、願望や疑いの表現などである。例えば、「話題になるほどではなかったかな」や「効果を期待して継続してみたい」のような表現である。評価とは、対象物の良し悪しを判定することである。例えば、「生地が安っぽい」や「便利な一冊である」のような表現である。

ユーザレビュー情報は Web 上に多く存在する。近年、インターネットの普及により、Web を利用して情報収集する人が増えている。また、自ら情報発信をするインターネットユーザも増えている。その情報発信の例として商品のレビューが挙げられる。Amazon.com の「カスタマーレビュー」や価格.com の「ユーザーレビュー」など、レビューを扱う Web サービスを利用してレビューを発信している。また、個人の blog にレビュー情報を掲載している場合もある。

ユーザレビューには利用者の感想や意見が書かれるため、製品メーカーが提供していない情報が得られる。この情報は、製品の情報を詳しく知りたい人（閱

覧者）にとって有用である。例えば、あるノートパソコンのキーボードについて知りたかった時、メーカーサイトでは製品の売りとなる小型化・軽量化のことしか書かれていない。それに対して、レビューを見ると「キーボードに付いているポインティングデバイスの感覚が良すぎて、必要無い時にも反応してしまう」というように、新たな情報が得られる。

しかし、レビューの書き手は一般の消費者であるため、レビュー内容の信頼性の判断は閲覧者に委ねられる。レビューの数が増えると閲覧者の負担が大きくなる。そこで、レビュー情報の有用な部分を自動的に分析する手法を提案し、閲覧者の負担を軽減する。レビューの内容を「態度」・「評価」・「特性」の3つに分類し、これらの出現の組み合わせにより、レビュー情報の重要な部分を判定する。

2. 先行研究

中山らはレビューについて、感情表現に着目した分析を行い、感情表現の特性を表す構成要素を定義した[1][2]。要素は“思う”・“使いやすい”など態度や感情の表明を示す「態度」，“○○だから”・“○○なので”など態度の原因や理由を示す「理由」，“私は”・“彼が”など態度を表明した人や物を示す「主体」，“このノー

トパソコンは”・“キーボードが”など態度が向けられた人や物を示す「対象」の4つで構成される。特に「理由」が内容理解に重要である事を示している。

ユーザレビュー情報は書き手の嗜好や状況の影響により評価が変化する。そのため、書き手の感情だけでなく、感情の理由が明記されている方が読み手にとって有用となる。中山らの定義では、主観的理由と客観的理由をまとめて「理由」要素としている。主観的理由は「態度」にも「理由」にもなり得るため、有用性の判定にはさらなる分類が必要である。

3. 提案手法

本研究では、ユーザレビュー情報から有用な情報を抽出するために、レビュー内の文章に対して有用性の自動判定を行う。

3.1. 分類する要素の定義

レビューの内容を3つの要素に分解し、定義する。要素の種類は「態度」、「評価」、「特性」である。

「態度」は、ユーザが商品を使って感じたこと、喜怒哀楽や心的態度、または実際の行動などが記述される部分である。次のようなものが「態度」として挙げられる。

- 感動しました
- 読みなおそうと思います
- 地震が来たら倒れちゃわね？

「評価」は、対象物に関して自分が持つ基準を満たしているかそうでないかが記述される部分である。次のようなものが「評価」として挙げられる。

- インクの減りが早い
- コストパフォーマンスは最高
- 他のショップに比べて安い

「特性」は、商品が持っている状態や情報が記述される部分である。次のようなものが「特性」として挙げられる。

- メモリーの伝送速度が低いと映像乱れが発生する
- USBコードも付いている
- 135時間ぐらい番組が録れる

3つの要素の関係は図1のようになる。

3つの要素は理由と反応の関係において、組み合わせに制約がある。例えば「SSDを搭載しているので起

動が早いです。」は特性-評価の関係になる。特性「SSDを搭載している」は理由となり、評価「起動が早いです。」が反応となる。この理由と反応の関係は自然であるが、理由と反応の内容を逆転させてしまうと「起動が早いのでSSDを搭載している」となり、不自然な内容になってしまう。特性-態度、評価-態度でも同様のことが言える。また、態度-態度と特性-特性の関係は有用性が弱い。例えば「この本は感動しました。なので、もう一度読みなおそうと思います。」は態度-態度の関係だが、読み手にとっては本のどのような所に感動したかが分からないので、有用になりにくい。逆に、特性-特性の関係は商品の説明だけされて、書き手の思いを表す部分がないので、有用になりにくい。

従って、これら3つの要素と要素間の関係を自動判別できれば、有用なレビューを抽出できる。要素間の関係と有用性を表1に示す。

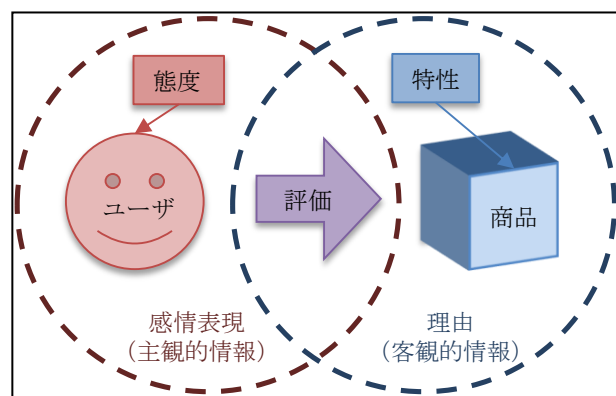


図 1 定義した要素の関係

表 1 原因-反応関係の制約

原因	反応	有用性
態度	態度	△
態度	評価	×
態度	特性	×
評価	態度	○
評価	評価	○
評価	特性	×
特性	態度	○
特性	評価	○
特性	特性	△

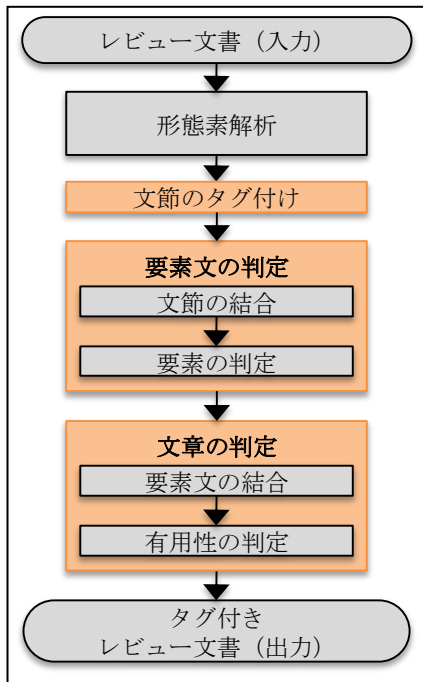


図 2 提案手法の処理の流れ

3.2. 有用文の判定方法

文章の中に要素がただ一つ存在する状態を要素文と呼ぶことにする。例えば、「SSD を搭載しているので起動が早いです。」は文章であり、「SSD を搭載しているの」と「起動が早いです。」がそれぞれ要素文となる。

要素の判定は「文節→要素文→文章」のように、語の最小単位から徐々に単位を大きくしていきながら行う。まず文節単位で要素判定に用いる情報のタグ付けを行う。次に、タグ情報を用いて文節の結合を行い、要素文の生成と要素判定を行う。さらに、要素文の要素を用いて要素文の結合を行い、文章を生成する。処理の流れを図 2 に示す。

3.2.1. 文節のタグ付け

それぞれの文節に対して、今後の処理に必要な情報を付加する。タグ付けの判定は形態素ごとに行う。タグの種類は以下の 6 種である。

- 態度タグ…人や物の態度，行動にかかわる文節
- 評価タグ…商品の評価にかかわる文節
- 特性タグ…商品の特性にかかわる文節
- 接続タグ…後の文節と意味的につながる文節
- 入れ子タグ…要素の区切りとなる文節
- 切断タグ…文の終わりとなる文節

「態度」・「評価」・「特性」タグはそれぞれ、要素文や文章の要素となる候補を表す。「接続」タグは文節を結合するときに用いる。「入れ子」は要素文の単位に区切るときに用いる。「切断」は文章の区切りを表すときに用いる。タグの判定は形態素の情報を用いる。タグ付けの判定ルールの一部を表 2 に示す。「評価」タグの判定には形態素の情報に加えて、小林ら[3]の評価値表現辞書を使用する。

3.2.2. 文節の結合と要素文の要素判定

有用性の判定で用いる要素文を作るため、文節の結合と要素文の要素の判定を行う。文節の結合と要素文の要素の判定には、前項で付けたタグの情報を用いる。文節の結合と要素判定の処理は先頭の文節から 1 つずつ行う。

結合は、「接続」・「入れ子」・「切断」のタグを用いる。「接続」タグは後ろの文節とのつながりを意味しているため、「接続」タグの付いている文節の場合は、後ろの文節と結合する。「入れ子」タグの場合は、要素文の区切りを意味している。よって、「入れ子」タグが出現したら、文節の結合を終了する。「切断」タグは文章の終わりを意味するため、「入れ子」タグと同様に、文節の結合を終了する。

文節の結合が終了した文（＝要素文）に対して、最終的な要素の判定をタグ付けする。要素の判定は文節の「態度」・「評価」・「特性」のタグを用いる。要素文の中の文節の「態度」・「評価」・「特性」のタグから 1 つ取り出し、要素文にそのタグを付加する。タグの優先順位は「特性＞評価＞態度」の順である。文節の結合と要素文の要素の判定イメージを図 3 に示す。

表 2 要素タグの判定ルール（一部）

タグ名	品詞	補足
態度	動詞	自立の動詞／仮定形、未然形、命令形である
評価	形容詞	仮定形、未然形、命令形ではない／評価表現辞書と併せて使用
特性	動詞	非自立の動詞＋直前の形態素が接続助詞である／活用形が連用タ接続である
接続	記号 助詞 副詞	読点／例) ～の、～が、～は／例) たぶん、
入れ子	助詞	次の形態素が(非自立 係助詞)ではない
切断	記号 助詞 接続詞	句点／「したがって」、「しかし」などで接続している

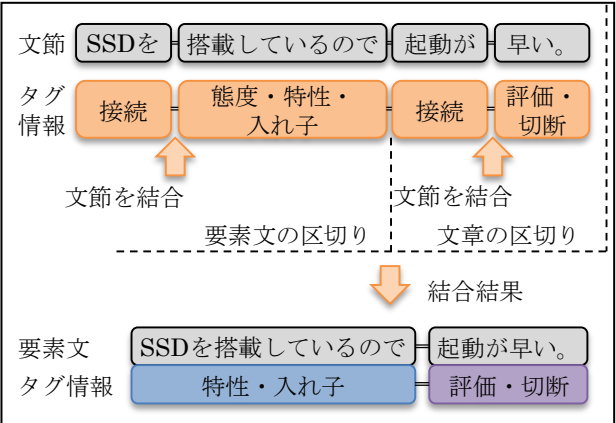


図 3 文節の結合処理

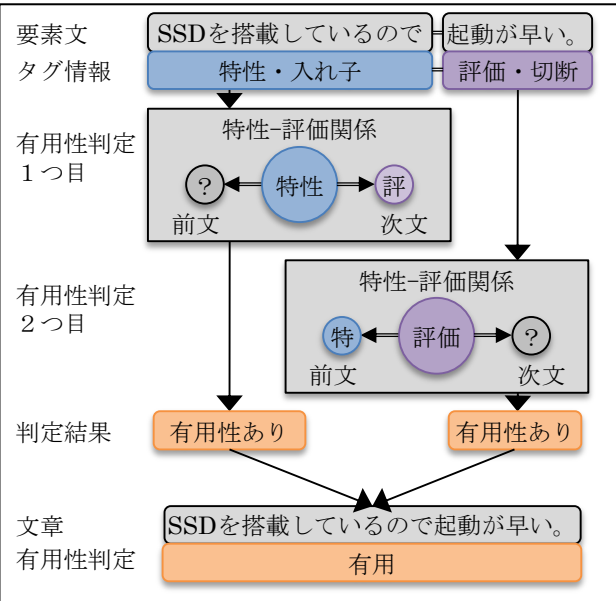


図 4 有用性の判定処理

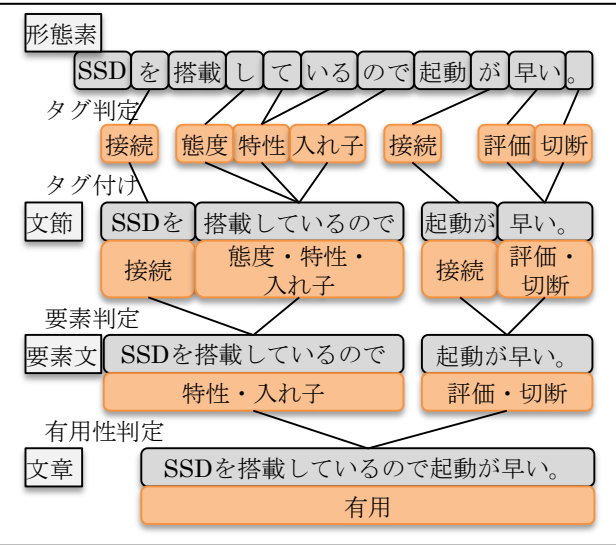


図 5 処理の例

3.2.3. 要素文の結合

有用性のあるレビューの提示は文章単位で行うため、各要素文を結合し、文章の単位にする。要素文には、「入れ子」タグ・「切断」タグの情報を用いる。

「入れ子」タグの場合は、文章のつながりがあるため、後ろの要素文と結合する。「切断」タグの場合は、文章の区切りであるため、結合を終了する。有用性の判定に用いる情報は要素文のタグ情報を用いるため、文章に対しては要素タグをつけない。

3.2.4. 文章の有用性の判定

レビューの要素文とタグ付けした情報を用いて、有用な文章の判定を行う。要素文について有用性の判定を行い、有用性があると判断された要素文を含んでいる文章を最終的な有用文とする。

要素文の有用性の判定は、前後の要素文の要素との比較を行い、有用性の判定ルールに該当する組み合わせだった場合に「有用性がある」と判定する。有用性の判定ルールは 3.1 節で示した要素間の関係の制約（表 1）に基づき作成した。文章の有用性の判定ルールを表 3 に示す。「.」は任意の要素を表す。要素文の先頭と末尾はそれぞれ前-要素文、後-要素文が存在しないので、存在する要素文のみで判定を行う。有用性の判定の処理の流れを図 4 に示す。

以上に述べた、形態素単位のタグ判定から有用文の判定までの処理の例を図 5 に示す。

表 3 有用性の判定ルール

前-要素文	現-要素文	後-要素文
評価	態度	.
特性	態度	.
評価	評価	.
特性	評価	.
.	態度	態度
.	評価	態度
.	評価	評価
.	特性	態度
.	特性	評価
.	特性	特性

入力文：
ちょいちょい映像、音声途切れてまったく使い物にならない。検索で一番上に出てきたから買った自分がバカでした。他の〇〇〇〇製品は愛用してますがね。

図 6 ユーザレビュー情報の例

出力文：
【有-評】ちょいちょい映像、音声途切れて
【有-態】まったく使い物にならない。

【有-特】検索で一番上に出てきたから
【有-評】買った自分がバカでした。

【有-態】他の〇〇〇〇製品は愛用してますがね。

図 7 提案手法による自動判定の例

4. 提案手法の実行例

提案手法による自動判別を行ったレビューの入力の例を図 6、出力の例を図 7 に示す。出力の例では、要素文の単位を一行に表示している。【有-評】は要素文についているタグで、「有用」と「評価」を意味している。文章の区切りは 1 行分の空白をあけている。形態素解析器は MeCab¹を使用し、辞書は IPA 辞書を用いた。文節の生成は、係り受け解析器 CaboCha²を用いて生成した。

5. 評価

提案手法の有効性を検証するために行った評価について述べる。提案手法では、本研究で定義した「態度」「評価」「特性」の要素を自動判別した。また、3つの要素を基にレビューの有用性の自動判別をした。よって、要素の判定、有用性の判定が正しく判定出来ているかを評価することにより、本研究の有効性を確認する。

評価用データには Amazon.co.jp のカスタマーレビューを用いた。「和書」「家電&カメラ」のカテゴリについて、カテゴリ内の売れ筋ランキングの上位からレビューが 10 件以上存在する商品を 10 件ずつ抽出した。(2009 年 1 月 17 日時点)そして、抽出した 20 件の商品についてレビューを投稿日が新しい順に 5 件ずつ取り出した。

計 100 件のレビューを用いて提案手法の評価を行う。評価方法は人手による手動判別と提案手法による機械判別を行った結果について比較を行う。レビューの比較単位は機械判別による要素文とする。

5.1. 要素の判定

要素文の要素について人手抽出と提案手法による機械抽出との比較を行う。「和書」について比較したものを表 4、「家電&カメラ」について比較したものを表 5 に示す。各表において、対角線部分が正しく抽出された要素数である。また、「和書」と「家電&カメラ」について、それぞれ精度を算出したものを表 6 に示す。

「家電&カメラ」について「特性」と「特性」の要素判定の正解数が「和書」より少なかった。「インクが溜まる」と「ストレスが溜まる」はそれぞれ特性と態度としたいが、提案手法では両方とも態度として判定してしまっていた。同じ動詞でも、係り元の名詞によって要素の種類が変わるため、動詞の係り元の名詞の種類によって判定を分ける必要がある。また評価については、「面白いと思う」という評価が「態度」にされていた。「～と思う」「～と考える」などはそれよりも前の文節の方が要素として適切である。また、評価タグの判定で用いた評価値表現辞書は、本研究で「態度」「特性」の要素とする語が少なからず含まれていた。そのため、本来「態度」要素である要素文が「評価」要素として取り出されていた。

5.2. 有用性の判定

文章の有用性について人手による抽出と提案手法による抽出との比較を行う。比較対象は要素の判定が正しかった要素文のみとする。「和書」について比較したものを表 7、「家電&カメラ」について比較したものを表 8 に示す。また、商品カテゴリ別の文章の判定の精度を表 9 に示す。

「和書」については 83%、「家電&カメラ」については 61%の精度で有用性の判定が出来ていた。

「和書」では書き手の「態度」と商品の「特性」が多く書かれていた。対して、「家電&カメラ」などは、商品の性能に対する「評価」が多く書かれていた。商品カテゴリによって、有用性の判定方法を変える必要がある。

誤判定については、「また書いて欲しい。あっさりしすぎているので。」と、書き手の反応が先に書かれ、理由が後に書かれているタイプがあった。有用性の判定ルールは、理由を先に書くタイプを想定したものだったのが誤判定の原因である。この問題の解決には、理由が記述されている文が手前の文に掛かっているものなのか、後ろの文に掛かっているものなのかを判別することが必要になる。

¹ <http://mecab.sourceforge.net/>

² <http://chasen.org/~taku/software/cabocha/>

表 4 各要素の判定数（和書）

和書		人手抽出			
		態度	評価	特性	小計
機械抽出	態度	81	17	34	132
	評価	16	55	10	81
	特性	26	19	94	139
	小計	123	91	138	352

表 5 各要素の判定数（家電&カメラ）

家電&カメラ		人手抽出			
		態度	評価	特性	小計
機械抽出	態度	31	42	12	85
	評価	10	50	7	68
	特性	32	42	12	89
	小計	73	135	34	242

表 6 機械による各要素の判定の適合率

商品カテゴリ	態度	評価	特性
和書	61%	68%	68%
家電&カメラ	36%	75%	14%

表 7 レビューの有用性の判定数（和書）

和書		人手抽出		
		有用である	有用でない	小計
機械抽出	有用である	125	26	151
	有用でない	62	17	79
	小計	187	43	230

表 8 レビューの有用性の判定数（家電&カメラ）

家電&カメラ		人手抽出		
		有用である	有用でない	小計
機械抽出	有用である	39	25	64
	有用でない	20	9	29
	小計	59	34	93

表 9 機械によるレビューの有用性判定の適合率

商品カテゴリ	有用である	有用でない
和書	83%	22%
家電&カメラ	61%	31%

$$\text{適合率} = \frac{\text{人手抽出と機械抽出が一致した個数}}{\text{機械抽出の個数}}$$

6. まとめ

本研究では、レビュー閲覧の負担低減のために、有用な文の自動判定を行う手法を提案した。要素文の前後との要素の比較を行うことで、有用性を判定した。

提案手法の評価では、Amazon.co.jp の「和書」カテゴリと「家電&カメラ」カテゴリの商品レビューを用いて、人手による手動判別と提案手法による機械判別を行った。有用性があると判定した文章の適合率は「和書」では 82%、「家電&カメラ」では 61% となり、「和書」カテゴリで有効であった。

今後の課題として、まず、要素の判定精度を上げることが挙げられる。本研究で定義した 3 つの要素は、有用性の判定に大きく関わっているため、要素の判定精度は重要である。また、今回は有用であるか、有用でないかの二択で判定を行っていたが、有用性の強さについても考える必要がある。有用性の強さとは、多くの人が有用性を感じるか、そうでないかを表すものである。有用性が強いレビューは、多くの人が有用だと思うレビューである。有用性が強いレビューを優先して提示することで、閲覧者の負担がより軽減出来る。そのためには、要素と有用性の関係を更に詳しく調べる必要がある。

参 考 文 献

- [1] 中山記男, 神門典子, “理由に着目した感情分析”, 電子情報通信学会技術研究報告 思考と言語, vol.105, no.291, pp.51-56, Sept.2005.
- [2] 中山記男, 神門典子, “レビューにおける「理由」の分析”, 情報処理学会研究報告 自然言語処理研究会報告, vol.2006, no.1, pp.81-88, Jan.2006.
- [3] 小林のぞみ, 乾健太郎, 松本裕治, “意見情報の抽出/構造化のタスク仕様に関する考察”, 情報処理学会研究報告 NL171-18, pp.111-118, Jan.2006.
- [4] 工藤拓, 山本薫, 松本裕治, “Conditional Random Fields を用いた日本語形態素解析”, 情報処理学会研究報告 自然言語処理研究会報告, 2004(47), pp. 89-96, May.2004.
- [5] 浅原正幸, 松本裕治, “ipadic version 2.7.0 ユーザーズマニュアル”, 奈良先端科学技術大学院大学情報科学研究科 自然言語処理学講座, Nov.2003.
- [6] 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, Vol.103, Num.3, pp.201-241, Nov.2006.
- [7] 天野真家, 石崎俊, 宇津呂武仁, 成田真澄, 福本淳一, “IT Text 自然言語処理”, 株式会社オーム社, Oct.2007.