# Web 上のレビュー情報から ユーザが重要視する製品の特徴を抽出する手法の提案

小林 大祐<sup>†</sup> 井上 潮<sup>‡</sup>

† ‡ 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@ed.cck.dendai.ac.jp, ‡ inoue@c.dendai.ac.jp

**あらまし** 製品のレビューを書くための Web サイトが存在し、利用者も多い. レビューには製品に関する意見が書かれ、メーカーが提供する情報にはない有用な情報が存在する. 閲覧者はこれらのレビューを見て、製品の情報を得る. しかし、レビューの量が増えると閲覧者が読む文章量が増えるため、内容を把握するための負担が大きくなる. そこで、レビューの内容を自動的に分析し、製品の特徴をわかりやすく提示する方法を検討している. 本稿では、レビューから製品の特徴を抽出する手法について述べる. ユーザの意見が生まれた「理由」が、製品の重要な特徴であると考え、形態素情報・係り受け情報を用いて、レビューから「理由」の書かれている部分を抽出する.

キーワード Web 情報システム,情報抽出,情報要約

# Extract Product Features which User Emphasize from Review Information on Web

Daisuke KOBAYASHI<sup>†</sup> and Ushio INOUE<sup>‡</sup>

† ‡ Tokyo Denki University 2-2 Kanda-Nishiki-cho, Chiyoda-ku, Tokyo 101-8457 Japan E-mail: † 08gmc13@ed.cck.dendai.ac.jp, ‡ inoue@c.dendai.ac.jp

**Keyword** web information system, information extraction, information summary

#### 1. はじめに

近年のインターネットの普及により、Web を通じて 情報発信することが容易になった. 情報発信の例とし てレビューが挙げられ, 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 サービスの存在もある. また個人 の blog にもレビュー記事を掲載されている場合もあ る. このように, 多くのレビューが Web 上に存在して いる. レビューには製品の利用者(以下,ユーザ)の 感想や意見が書かれ、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(以下,閲覧者)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度が良すぎて,必要ない時にも反 応してしまう」と書かれていて、知りたかった情報が 得られる. このように、閲覧者の知りたい情報をレビ ューから得ることができる. しかし, レビューの数が 多くありすぎると文章量が多くなり,全てのレビュー を把握しようとすると閲覧者の負担が大きくなる. そ こで、レビューを自動的に分析して、内容の把握に重 要となる部分を自動的に抽出し, 閲覧者に分かりやす

く提示する方法を検討している.本稿では、レビューの「意見」や「理由」などの情報がレビューの重要な部分と考え、レビューから「意見」や「理由」などの情報を自動的に抽出する方法を述べる.

# 2. 関連研究

中山ら[1][2]は「うれしかった」「悲しかった」などの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.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」、"〇〇だから"・"〇〇なので"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、特に「理由」が内容理解に重要である事を示している.分析は手動で行っていて、課題として分析の自動化を挙げている.

製品レビューの自動分析に関するいくつかの Web サービスが公開されている. みんなの評判  $\beta$  [3]は、Amazon.co.jp[4]のカスタマーレビューの書き込みを収集し、レビュー文書から製品の評判に関する部分を自動的に抽出する Web サービスである. 抽出した評判情報をランキング化して表示するなどをしているが、製品の特徴とは違う情報が抽出されてしまうことがある.

PLURIBO[5]は Amazon.com のレビューを分析する Firefox のアドオンである. レビュー文を分析し, 肯定評価・否定評価の抽出やレビュー分の要約の生成を行う. しかし, PLURIBO は英文のみの対応で, 日本語のレビューには利用できない.

## 3. 研究目標と提案手法

本研究では、レビュー内の意見が生まれた理由を自動的に抽出することで、製品の特徴を提示することを 目標にしている.

手法としては、中山らの4つの要素を参考にし、レビューの構成要素を自動的に分析する. レビューの要素として中山らが定義した要素のうち、「態度」、「対象」、「理由」の要素を採用する. また、新たに「条件」という要素を加えて、レビューの分析を行う. 採用しなかった「主体」はレビュー情報では省略されやすい要素と考えたので、今回は考慮していない. 新たに加えた「条件」は、「初心者にとっては」や「思い入れのある人にとっては」など、ある要素について制限をかける表現のものとする. この要素を検出すると、どのような時に各要素の状態になるのかが分かり、内容を把握する助けになると考えられる.

レビューから要素を抽出するための手法の流れを 図 1 に示す.

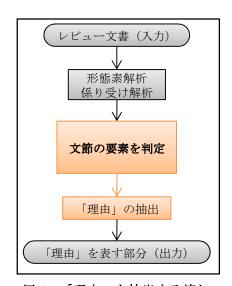


図1 「理由」を抽出する流れ

まず、レビューの文章に対して係り受け解析をする. そして、解析結果からレビューの要素の判定を行う. 要素判定の処理は係り受けの構造を末尾から先頭方向 に、係り元を辿るように行う.これは文の末尾に「態 度」を示すレビューが多く、後に述べる各要素の検出 や結合の処理が簡単になるからである.図2に分析の 順序の例を示す.

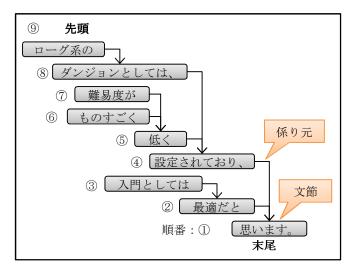


図2 文節の分析順序の例

表1 各要素の定義と例

要素	定義	例
態度	感想や意見を表す表現	最適だと思う
理由	感想や意見がうまれた理由や 根拠を表す表現	難易度が低いので
	係り先の事柄の範囲に 制限をかける表現	入門としては
対象	感想や意見の対象を表す表現	このゲームは

要素の種類としては「態度」、「条件」、「理由」、「対象」の4種類とする.これらの要素について検出・結合を行う.検出・結合処理には形態素情報と係り受け情報を用いる.表1に各要素の定義と例を示す.

## (ア) 対象・態度候補の検出と結合

係り受け解析後の情報について、「対象」や「態度」の候補を検出する.「ゲームのキャラクターが」など、複数の語で意味を表す表現があり、それが文節を跨いでいるため、一つのまとまりに結合する.この処理を「対象」と「態度」のみにしたのは、「理由」や「条件」もこれらの要素から構成されているからである.に「対象」候補の判定項目を表 2、「態度」候補の判定項目を表 3 に示す.各表に示した項目が文節内の形態素情報に存在するとき、その文節を要素として検出する.

その他の項目にある「機能語なし」については、文 節内に文法的な役割を持つ語が存在しない場合を指す、 機能語が無い状態でも形容詞や副詞の語があれば、要 素が続いている可能性が高い、対象・態度候補の検出 と結合の処理の例を図3に示す.

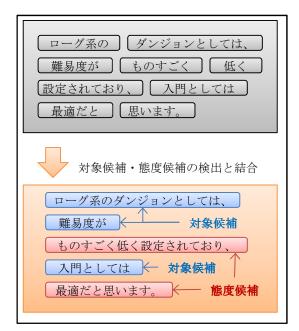


図3 対象・態度候補の検出処理の例

### (イ) 各要素の検出

(ア)の処理の後に、各要素の検出を行う. 各要素の判定項目を表 4 に示す. ただし、文節に対して複数の要素が検出されてしまうことがあるため、要素の検出時に「条件>理由>態度」という優先順位を設ける. 「対象」要素については、他の3つの要素に該当しないもののうち、「対象」候補と判定されているものを「対象」と判定する.

#### (ウ) 各要素の結合

(イ)で要素が検出された場合、係り元の文節も(イ)と同じ要素で、なおかつ同一のグループである可能性がある.このような場合は、係り元の文節と結合して、グループ化する. 各要素の結合の判定項目を表 5 に示す.

図 4 に各文節を要素のグループに分けた例を示す.この例では「難易度が、ものすごく低く設定されており、」というのを理由の要素として判定している.

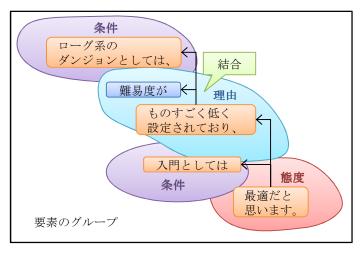


図4 各文節の要素グループの例

表 2 「対象候補」の判定項目

要素		品詞	品詞 細分類	原形	その他
		名詞	一般		
	ᄊ	名詞	自立		
按	検出	名詞	副詞可能		
対象候	щ	名詞	サ変接続		
		名詞	固有名詞		
補	<b>%+</b> :	連体詞	格助詞		
	結合	助詞	連体化		
	ı	助詞	並立助詞		

表 3 「態度候補」の判定項目

要素		品詞	品詞 細分類	原形	その他
		動詞	自立		
	検	形容詞	自立		
態度	出	名詞	形容動詞 語幹		
候補		助詞	格助詞	と	
補	結	形容詞			機能語なし
	合	名詞	副詞可能		
		副詞			機能語なし

表 4 各要素の判定項目

要素	品詞	品詞 細分類	原形	その他
	動詞	自立		
態	形容詞	自立		
度	名詞	形容動詞 語幹		
	名詞	非自立	ため	
	動詞	非自立	おる	
<b>理</b> 由	助詞	格助詞	によって	
"	助詞	接続助詞	ので	
	助詞	接続助詞	から	
	助詞	格助詞	として	
	助詞	格助詞	にとって	
     	助詞	副助詞	でも	
<del>朱</del>   件	助詞	接続助詞	と	
''	助詞	接続助詞	は	
	助詞	非自立	よう	
	名詞	非自立	とたん	

表 5 「態度」・「理由」・「条件」要素の結合項目

品詞	品詞 細分類	原形	その他
助詞	連体化	の	
助詞	格助詞	が	
助詞	格助詞	に	
助詞	格助詞	を	
助詞	副詞化	に	
助詞	格助詞	と	
助詞	並立助詞		
連体詞			機能語なし
動詞	自立		機能語なし
形容詞	自立		機能語なし
副詞	助詞類 接続		機能語なし
名詞	接尾		機能語なし
名詞	形容動詞 語幹		機能語なし

# 4. 検証と評価

提案手法によって各要素が正しく抽出できるかどうかを検証するため、Amazon.co.jpのカスタマーレビューから、レビューが 10件以上ある製品を対象に、製品を無作為に 10個選択し、各製品について最近のレビューを 10件ずつ取得した。そして、取得した合計 100件のレビューを人手で分析したものと提案手法で自動分析したものを比較した。使用したレビューの一部を図 5に示す。係り受け解析には CaboCha 0.60 pre4 を使用した。自動分析の結果の一部を図 6に示す。

2つの抽出結果について、各要素の抽出数を表 6 に

示す. 一致の判定は提案手法で自動抽出した要素が手動で抽出した要素と重なる部分が少しでもあれば一致とみなしている.

表 6 手動分析と提案手法による各要素の抽出数

要素	手動 抽出	提案 手法	一致	再現率	適合率	F値
理由	159	61	50	31%	82%	0.45
条件	166	112	33	20%	29%	0.24
態度	1094	1315	763	70%	58%	0.63
対象	1575	1187	498	32%	42%	0.36

今回,重要な要素としている「理由」について,適合率が82%となり,4つの要素中で最も高い数値となった.これは,要素判定に使用している形態素情報が文節内に含まれていれば「理由」の要素である可能性が高いことを示している.しかし,再現率は31%となった.一致しなかったものの例を表7に示す.原因として,手動の場合では「条件」要素として判定されているものがあった."環境によっては"という文節に"によって"があったため,「理由」として判定されていた.次に,提案手法の「理由」要素の判定項目に含まれないものがあった."同じ作業の繰り返しになりがちで"では「態度」要素,"気兼ねなく叩けるのが"では"気兼ね"が「対象」要素,"なく叩けるのが"が「態度」要素であった.

提案手法では、文節内に要素の判定項目が含まれているかどうかのみで行っていたことが判定ミスの原因と考えられる。文節内の形態素情報の組み合わせ、文節の組み合わせ、係り受けの状態を判定項目とすることで、精度が向上すると考えられる。

表 7 一致しなかった「理由」要素の例

一致しなかった文	原因
環境によっては	手動抽出では 条件要素
同じ作業の繰り返しになりがちで	自動抽出の 判定項目外
気兼ねなく叩けるのが	自動抽出の 判定項目外

# チョコボの不思議なダンジョン 時忘れの迷宮

### レビュー文(入力)

ダンジョンが自動生成され、思考時間がたっぷりあって考えながら進めていけるローグ系のダンジョンとしては、難易度がものすごく低く設定されており、入門としては最適だと思います。

ジョブによる味付けもいい感じですし、音楽も素晴ら しい。

ストーリーは良くあるものですが、雰囲気はとても良い

良い点は、とにかくゲームを初めてやられたような方でも楽しめるようにと、難易度が無茶苦茶下げられていること。

本格的な(になると思っている)風来のシレンまでの 練習用にはちょうど良いと思います。

悪い点は、まず死なないこと、何が何でも死なない 様にと随所に救済措置が散りばめられ過ぎていること。

ポス以外では死ぬ事がないため、緊張感がまるでない。

メインストーリーではダンジョンも短いですし。

ローグ系、メインは風来のシレンと思っていますが、 ローグ系を遊ばれたことがない方は、とりあえずつ なぎにどうぞ。

# 図 5 使用したレビューの一部

#### 分析結果(出力)

[[ダンジョンが]s 自動生成され、]a [[思考時間が]s たっぷりあって]a [考えながら]a [進めていける]a <mark>[[ローグ系のダンジョンとしては、]s ]c</mark> [機態易度が]s ものすごく低く設定されており、]r <mark>[[入門としては]s</mark> ]<mark>c</mark> [最適だと思います。]a

-{ジョブによる味付けもトs <mark>[いい]a</mark> <u>(感じですし、ls</u> <u>倍</u> 楽もトs [素晴らしい。]a

| <u>(ストーリー| は)。 [良くある] a</u> ものですが、 <u>| 雰囲気</u> | <u>は)。 [とても良い] a</u>

[良い]a 点は、とにかく<u>げームを</u>ls 初めてやられた ような<mark>[方でも]o</mark> [楽しめるようにと、**難**易度がls 無 茶苦茶下げられている]a こと。

{本格的な!s [{(Cls なると思っている]a ()風来のシレンまでの練習用!コま!s ちょうど良いと思います。

<u>[悪い]a</u> 点は、まず死なないこと、<mark>[何が何でも]c</mark> [死 なない様にと随所に。 (飲済措置が)。 散りばめられ 過ぎている]a こと。

<u>/ボス以外ではls</u> <mark>|死ぬ事がないため、]r</mark> <u>|緊張感</u> がls まるでない。

タインストーリーではb ダンジョンもb **短いです** し。Ja

<u>仰一</u>グ系、ks (メインはks <mark>((風来のks シレンと思って</mark> いますが、 la <u>[(ローグ系をks 遊ばれた]a [ことがな</u> い]a 方は、とりあえず<u>(つなぎにks</u> どうぞ。

図 6 提案手法による自動分析

#### 5. おわりに

レビューから製品の特性を自動的に分析するために、「意見」が生まれた「理由」が製品と特長を表していると考えた.本稿では、中山らの研究を参考に、「条件」「理由」「態度」「対象」の要素を定義し、レビューから要素を抽出の手法について述べた.要素の抽出には、形態素情報と係り受け情報を使用し、実際に投稿されているユーザのレビューを用いて提案手法の有効性を検証した.その結果、今回の手法では、抽出精度は良くなかった。今後は、提案手法を検討しなおし、精度向上を目指す.

# 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] みんなの評判β http://hyoban.biglobe.ne.jp/
- [4] Amazon.co.jp http://www.amazon.co.jp/
- [5] PLURIBO http://www.pluribo.com/
- [6] CaboCha http://chasen.org/~taku/software/cabocha/
- [7] MeCab http://mecab.sourceforge.net/