

属性評価モデルに基づく商品評価の抽出と提示

平山 拓央[†] 湯本 高行^{††} 新居 学^{††} 高橋 豊^{††}

[†] 兵庫県立大学工学部 〒671-2280 兵庫県姫路市書写 2167

^{††} 兵庫県立大学大学院工学研究科 〒671-2280 兵庫県姫路市書写 2167

E-mail: [†]eo07n083@steng.u-hyogo.ac.jp, ^{††}{yumoto,nii,takahasi}@eng.u-hyogo.ac.jp

あらまし 通販サイトに投稿されているレビューは、商品に関する詳細な情報が記述されており、商品の購入を考えるユーザにとって有益な情報である。しかし、レビューの数が多い場合にはすべてを参照することは負担のかかる作業である。また、一部のレビューのみを閲覧した場合には、商品に対して偏った印象を受けてしまう場合がある。そこで本稿では、商品レビューから評価を抽出し、ユーザが容易にレビュー集合全体の評価を把握できる提示を行う手法を提案する。本手法ではレビュー中の文章の中で、商品の機能や特徴（評価属性）に対する評価が記述されている評価文に着目し、属性評価モデルに基づき、評価文から評価属性に対する評価を抽出する。また、レビュー集合から抽出した評価を表形式に可視化することで、商品に対するレビュー集合全体の評価を容易に把握可能とする。さらに、根拠を伴った評価文により商品に関する補足的な情報を提示する。

キーワード 属性評価モデル、評価の根拠

Extraction and Presentation of Product Reputation by Attribute Evaluation Model

Takuo HIRAYAMA[†], Takayuki YUMOTO^{††}, Manabu NII^{††}, and Yutaka TAKAHASHI^{††}

[†] School of Engineering, University of Hyogo 2167 Shosha, Himeji, Hyogo, 671-2280, Japan

^{††} Graduate School of Engineering, University of Hyogo 2167 Shosha, Himeji, Hyogo, 671-2280, Japan

E-mail: [†]eo07n083@steng.u-hyogo.ac.jp, ^{††}{yumoto,nii,takahasi}@eng.u-hyogo.ac.jp

1. はじめに

近年、インターネットを利用した商品取引の機会が増加している。それに伴い、通販サイトでは商品に関する評判情報を提供している。評判情報には、図1のような商品に対する段階評価の統計結果がある。これは、商品の購入者が商品を段階で評価したものを集計した結果である。図1では、全体的に5や4の評価数が多く、比較的良好な商品であることがわかる。このように、統計結果は全体の評価が可視化されており、商品の良し悪しを容易に把握することができる。



図1 商品評価の統計

★★★★☆ 印刷音が軽くてうるさめです。2010/1/21

[レビューをすべて見る](#)

レビュー対象商品: EPSON Colorio インクジェット複合機 EP-802A 有線・無線LAN標準搭載インク (Personal Computers)

PM-A890からの買い替えです。

印刷スピードが早く、写真も高画質で、大変使い勝手がいいです。

デザインも上部がスッキリとフラットなのが気に入っています。

(用紙置場と化してしまうので注意が必要ですが...)

給紙がトレイ型のもホコリが入らなさそうで良いです。

ただ...印刷の音が軽くてうるさいのが気になります。

"静か"とレビューされている方もいらっしゃるのですが、個体差?があるのかもしれませんが、家にきたのはそれで、そこだけ残念でした。

図2 商品に関するレビュー

また、図2のような購入者の評価や感想が記述されているレビューが提供されている。図2では、「印刷スピード」や「使い勝手」について評価が記述されていることが読み取れる。このように、レビューは商品の機能や特徴について評価が記述されており、商品に関する詳しい情報を読み取ることができる。

しかし、図1のような統計結果からは全体の評価はわかるが、商品の機能等に対する評価は読み取れない。そこで、機能等に

対する評価を知りたい場合はレビューを読む必要がある。しかし、レビューはある程度の文章量を有するテキストであり、レビュー数が多い場合には全てを閲覧することは負担のかかる作業となる。また、多数あるレビューの中で、少数のみを閲覧するだけでは、商品に対して偏った印象を受けてしまう場合がある。このように、通販サイトで提供されている評判情報から、商品に関する全体の詳細な評価を把握することは困難である。

本稿では、このようなレビュー数の多い商品を対象とし、ユーザが大量のレビューを閲覧することなく、容易に商品の機能等に対する全体の評価を把握できる提示を行う手法を提案する。本提案手法では、商品の機能や特徴を評価属性とし、レビュー中の評価属性に関して記述されている文（評価文）から評価を抽出する。評価は、「評価属性について良く（ポジティブな）、もしくは悪く（ネガティブな）評価しているのか」と「評価文に根拠が含まれているか」の2つの観点に基づき抽出する。前者の観点により商品の評価属性毎に全体のポジティブな評価数とネガティブな評価数を算出し、後者によりユーザにとって有用な評価文を抽出する。また、本手法では抽出した商品評価を表形式に可視化することにより、評価属性に対するレビュー集合全体の評価を容易に把握可能とする。

2. 関連研究

谷本ら [1] は特定の商品に関係する評価表現辞書を作成した後、その辞書を用いて商品に関するレビューテキストを「デザイン」等の評価属性ごとに点数化することで評価を可視化し、ユーザの商品の評価を把握しやすくする手法を提案している。しかし、この手法では商品の可視化の際には必ず作成した評価表現辞書を用いる必要がある。またレビューテキストひとつに関して点数化をするため、全体の評価は算出していない。本研究の提案手法では、商品の評価属性に対する全体の評価を算出するため、レビュー集合全体の評判情報を容易に把握することができる。また、辞書をあらかじめ必要としないため、汎用的な処理が可能である。

小林ら [2] は対象名、属性表現、評価表現の共起パターンに基づく、属性、評価表現の半自動的収集方法を提案している。共起パターンは「属性表現」が/は/も/に/を「評価表現」というようなテンプレート表現をなすものとしている。本手法では、日本語係り受け解析器の CaboCha^(注1)を用いた係り受け解析を用いることで、商品の評価の基準となる評価属性、評価属性に対する評価表現を取得する。

小倉ら [3] は、ある対象に対する評価情報に付与されている評価に着目し、レビュー文の一般閲覧者から得た評価をもとに支持度と参考度を算出することによって、有益な評判情報の抽出を行う手法を提案している。しかし、一般閲覧者の中には、商品に対して熱狂的な支持者や批判的な意見を持つ者が存在し、このような支持には偏りが生まれてしまう場合がある。このような場合には、取得してきた情報がユーザにとって有益とならない可能性がある。提案手法では、記述されている評価の支持

度は考慮せず、有用であるかを文章の根拠から判定するため、偏りの少ない評価の取得が可能であると考えられる。

3. 商品評価のモデル化と提示

本手法では、対象とする商品に関するレビュー集合から、商品評価を抽出しユーザに提示する。そこで、レビューは商品の機能や特徴について評価が記述されている評価文とそれ以外の文で構成されていると捉える。この観点から、レビュー中の評価文の特徴を読み取ることにより、レビュー集合から商品評価を抽出し、ユーザに提示を行う。

本章では、評価文、属性評価モデル、評価の提示について詳しく述べる。

3.1 評価文

評価文は、商品の機能や特徴について評価が記述されており、時に評価を証明する根拠を伴う文である。複合機に関するレビューの評価文の例を以下に示す。

- 動作音がガチャガチャうるさいです。
- なんとと言っても無線 LAN 機能があるので、置き場所は選ばないまでもいいけど使い勝手はとても良い。

上記の2文は、ある複合機に関するレビュー中の評価文である。1番目の文からは「動作音」が「うるさい」と記述しており、「動作音」に対してネガティブな評価を表現していることがわかる。2番目の文からは、複合機の「使い勝手」が「良い」と評価しており、こちらは「使い勝手」に対してポジティブな評価を表現していることがわかる。また、評価の根拠として「無線 LAN 機能がある」と記述されており、「使い勝手」の良し悪しに関する補足的な情報を読み取ることができる。評価の抽出には、この評価文の特徴を読み取ることにより行う。

3.2 属性評価モデル

評価文から評価属性に対する評価を抽出するため、本手法では次に示す2つの観点を考える。

(1) 評価属性に対して良く、もしくは悪く評価しているのか。

(2) 評価文中に根拠が含まれているか。

1つ目の観点は、評価文が評価属性に対して良く、もしくは悪く評価しているかを読み取るためである。また2つ目の観点は、評価文自体がユーザにとって有用な情報となるかを判定するためである。

また、3.1節で述べた評価文の特徴を踏まえ、評価を構成する要素を以下のように定義する。

- 評価属性 … 商品の具体的な機能や特徴といった、評価の基準となる語。
- 評価極性 … 評価属性に対して良く、もしくは悪く評価しているのかを表す極性。
- 根拠 … 評価属性に対する評価を証明するための根拠にあたる句。

3.1節の例において、それぞれ「動作音」と「使い勝手」が評価属性であり、評価属性に対する「うるさい」と「良い」という評価の極性が評価極性となる。また、2番目の評価文の「無線 LAN 機能があるので」は、「使い勝手」が「良い」事を証明

(注1): <http://chasen.org/~taku/software/cabocha/>

する根拠に相当する。

以上の定義を踏まえ、評価文中の評価属性に対する評価を属性評価モデル eva としてモデル化する。

$$eva = (at, va, re) \quad (1)$$

ここで、 at は評価属性、 va は評価極性、 re は評価文の根拠(ない場合は ϕ) である。本手法では、属性評価モデルに基づき、評価文から評価属性に対する評価の抽出を行う。図 3 に評価の抽出イメージを示す。

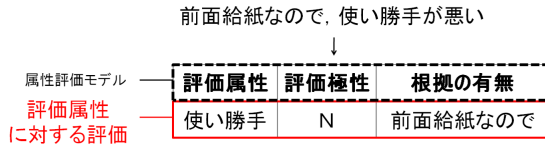


図 3 評価属性に対する評価の抽出

3.3 商品評価の提示

本手法では、3.2 節で示した属性評価モデルに基づき、レビュー集合から抽出した評価を提示する。具体的には、評価属性ごとにポジティブな評価数、ネガティブな評価数を集計することで、全体の評価を統計として提示する。これにより、レビューを全て閲覧することなく、商品の機能や特徴に関する評価を容易に把握することができる。レビュー集合全体の評価の統計を提示するため、先入観を持つことなく評価が可能である。また、評価に根拠を伴う評価文を、上述した統計と共に提示することで、より詳細な情報を読み取ることが可能である。図 4 に商品評価の提示イメージの例を示す。図 4 から、「使い勝手」が全体的に悪い評価の傾向にあることが一目でわかる。また、評価が悪い傾向として、根拠を伴った評価文から「前面給紙」が影響していることが読み取れる。



図 4 商品評価の提示

4. 評価属性に対する評価の抽出

本章では、対象とする商品のレビュー集合から評価文を抽出し、属性評価モデルに基づき評価を抽出する手法について述べる。対象商品 p に対するレビュー集合 R_{all} は、対象商品 p に対する各レビュー a が書いたレビュー R_a の和集合として表される。

$$R_{all}(p) = \bigcup_a R_a(p) \quad (2)$$

本手法では、(2) 式に示すレビュー集合から評価文を収集し、評価を抽出する。評価の抽出までの一連の流れを図 5 に示す。それぞれに記載されている数字は処理の順番である。また、本手法における単位文とは、文中の 1 文(文章の初めから文末記号まで)のことをさす。

評価文を抽出する処理には、単位文が評価文であるかどうかの判定の足掛かりとして、評価属性とその評価属性に対する評価の語を用いる。そのため、評価属性と評価表現を単位文から取得し、評価属性の集約処理の後に評価文の抽出を行う。それぞれの処理に関しては、4.1 節から 4.5 節で述べる。

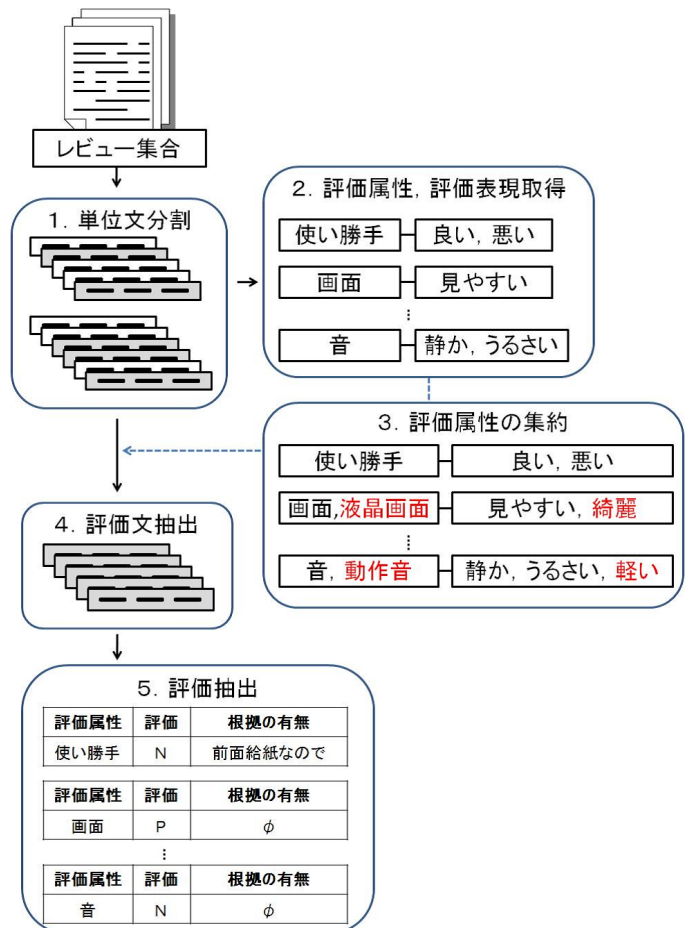


図 5 処理の流れ

図 5 の処理では、文章の単語や品詞の特徴を読み取るため、形態素解析を用いる。本手法では、形態素解析によって得られた語の品詞から、評価属性や評価表現を発見する。形態素解析

には MeCab^(注2)を利用して行う。

4.1 単位文分割

本節では、レビューテキストを単位文に分割する処理について述べる。レビューテキストは対象商品に関する評価をするという以外に、記述に規則がほとんど存在しない。そのため、文末に用いられる記号として句点(。)以外にも、感嘆符(!)や疑問符(?)といった記号が用いられる。そこで、以下に示す文末に用いられる記号を基準として区切ること、レビューテキストを単位文に分解する。

「。」「!」「?」「!」「?」

また、レビューの分割処理の概略図を図6に示す。この処理をレビュー集合すべてのレビューテキストに対して行い、レビュー集合を単位文に分割する。

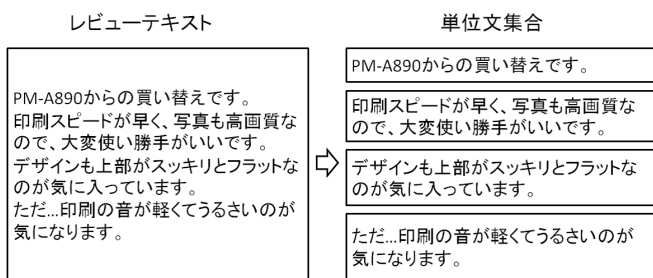


図6 単位文分割

4.2 評価属性、評価表現取得

レビュー中の評価属性に対する意見や評価の記述には、「音が静か」や「使い勝手がいい」といったように、評価属性と評価は文中で主語と述語という修飾関係にあることが多い。そこで、本手法では日本語係り受け解析器の CaboCha を用いて、単位文の係り受けを解析し、評価属性と評価表現を取得する。評価表現とは、評価属性に対する評価を表現する語である。例えば、「使い勝手」に対する「良い」という語が評価表現にあたる。

本手法における、評価表現に用いられる語の品詞を以下に示す。

- 形容詞
- 名詞, 形容動詞語幹
- 名詞, ナイ形容詞語幹
- 動詞(ただし、係り先の次の語が形容詞の場合のみ)

したがって、単位文の係り受け解析を行った際に、上記の品詞に係っている語を評価表現、係り元となっている名詞を評価属性として取得する。なお、「名詞, 形容動詞語幹」に関しては MeCab の解析で出力される形態素の品詞であり、形容動詞を意味するが、本手法では「名詞, 形容動詞語幹」という名詞として扱う。図7に評価属性と評価表現の取得の例を示す。

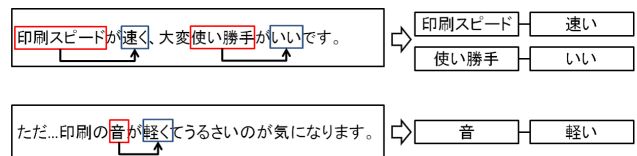


図7 係り受け解析による評価属性、評価表現の取得

4.3 評価属性の集約

4.2節により取得した評価属性の中には、同義語のような評価属性同士の関係が存在する。例えば、複合機の評価属性として「音」、「動作音」がある場合、それぞれ表記は違うが、複合機と同じ特徴を意味している。このような評価属性同士の関係に対応するために、本手法では評価属性の集約を行う。

レビューにおける評価属性は、同じ機能や特徴を示すような語同士の表記は類似することが多く、違う場合には基本的に表記は異なる。そこで、評価属性の文字列を比較することで、評価属性の類似を判定し、集約を行う。基準となる評価属性と比較対象の評価属性の類似は、次の処理によって判定する。

(1) 評価属性の分解

比較対象となる評価属性を形態素に分解する。

(2) 評価属性の類似判定

基準となる評価属性に、分解した評価属性の形態素が含まれているかを判定する。

また、上記の処理の概要を図8に示す。

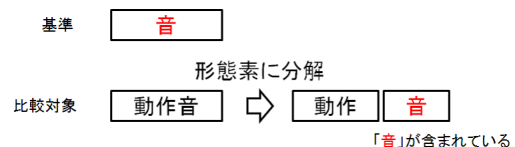


図8 評価属性の類似判定

この処理により、基準となる評価属性に比較対象の形態素が一語でも含まれている場合、双方の評価属性を集約する。集約を行うことで、上述の2つの評価属性が「動作音、音」という語の集合から成り立つ評価属性となり、それぞれの評価属性に対応する評価表現も同様にして集約される。なお、出力の際には評価属性を構成する語のうち最も出現頻度の高い語を用いる。

4.4 評価文抽出

評価文は、評価属性に対する評価が記述されている単位文である。そこで、4.2節で取得した評価属性と評価表現の対応関係が、一文中で確認できる単位文を評価文として抽出する。

ここで、単位文 ul を次のように表現する。

$$ul = (N, A, V) \quad (3)$$

N, A, V はそれぞれ単位文中に出現した名詞の集合、形容詞の集合、動詞の集合である。 N に取得した評価属性が存在し、 A か V のどちらかに評価属性に対応する評価表現が確認できる場合、注目している単位文は評価属性に対する評価文とみなす。また、単位文中に複数の評価属性と評価表現の対応関係が確認できる場合、注目している単位文は、それぞれの評価属性

(注2): <http://mecab.sourceforge.net/>

に対する評価文とする。

例えば、「使い勝手が以前のものと比べて良いと思います。」という単位文の場合、名詞は「使い勝手、以前、もの」、形容詞は「良い」、動詞は「比べる、思う」となる。4.2節で述べた評価属性と評価表現の取得において、「使い勝手」という評価属性と、それに対応する評価表現として「良い、悪い」が取得できた場合、この単位文は「使い勝手」に関する評価文と判定し抽出する。

4.5 評価抽出

本節では、3.2節で述べた属性評価モデルに基づき、評価文から評価属性に対する評価を抽出する手法について述べる。

4.5.1 評価の極性判定

4.4節で抽出した評価文が、評価属性に対して良く、もしくは悪く評価しているのかを判定する。本手法では、評価文の評価属性に対する評価極性を判定する。

単語は、それ自体がポジティブ(+), ネガティブ(-)といった極性を持つものや、極性を持たないニュートラルなものも存在する。また、極性を持つ単語を修飾することで、単語の極性を変化させる極性反転子がある。評価にも上記のような極性を持つ語が用いられることが多い。

青木ら[4]は、日本語評価極性辞書(名詞編)[5]と単語感情極性対応表[6]に自身が作成した極性反転子辞書を含めた単語極性辞書を用いて、ニュース記事中の見解を述べた見解文の極性を判定し、発信者の意図を抽出する手法を提案している。本手法では、この単語極性辞書を評価の極性を判定するために用いる。

ここで、上記の極性判定を行うためには、評価文の評価を述べている語を特定する必要がある。本手法では、評価文の抽出の際に取得した評価属性と評価表現の対応関係を用いたが、抽出の際に確認できた評価表現が、評価文の評価とは限らない場合がある。例えば、

「音が多少うるさいですが、いまは快適です。」

という文では、「音」という評価属性に対する「うるさい」という評価表現の対応関係を確認することで、評価文を抽出したが、評価は「快適」と述べられている。このように、評価属性に対する評価表現は評価文中の評価とならない場合がある。

そこで、文章では一般的に結論が末尾に書かれることが多いことを考慮し、文末から最も近い極性を持つ語を評価属性に対する極性の基準とする。本手法では、極性の基準に続く極性反転子の数を確認することで、評価文の評価極性を判定する。評価文 el の評価極性 $Polarity(el)$ は、極性の基準となる語 va_{el} の極性と極性反転子の数 i_{cnt} に対し、式4のように表す。

$$Polarity(el) = \begin{cases} + & va_{el}が+で i_{cnt}が偶数 \\ - & va_{el}が+で i_{cnt}が奇数 \\ - & va_{el}が-で i_{cnt}が偶数 \\ + & va_{el}が-で i_{cnt}が奇数 \end{cases} \quad (4)$$

+, - はそれぞれポジティブ, ネガティブである。

4.5.2 評価文の有用性判定

評価文が根拠を伴っているかどうかを判定し、有用であるか

を判定する。文中で根拠を表現する際には、多くの場合根拠に続く語が存在する。したがって、根拠は以下のような構文パターンで表わす事ができる。

$$\text{根拠} = \langle \text{根拠部} \rangle + \langle \text{ので} \mid \text{から} \rangle \quad (5)$$

根拠部は、根拠の内容となる部分である。本手法では、上記の根拠の構文パターンを考慮し、評価文中に「ので」や「から」が含まれているかにより、根拠の有無を判定する。

「無線LAN機能があるので、使い勝手は良い。」という評価文は、次のように形態素ごとに分解される。

無線/LAN/機能/が/ある/ので/, /使い勝手/は/良い/。

この時、形態素の中に「ので」が含まれているため、この評価文は根拠を含む。

5. 実験

本章では、評価属性の取得、評価属性の集約、評価文の抽出、評価文の有用性判定に対する実験結果と考察を述べる。ただし、5.1節の評価属性の取得実験以降の実験に関しては、5.1節の結果を考慮し、商品の評価属性として正しい語に対してのみ実験を行った。なお評価実験には、以下の amazon.co.jp の商品に関するレビューを用いた。

- ドラゴンクエスト IX
- EPSON Colorio インクジェット複合機 EP-802A
- 1Q84 BOOK 1
- Apple iPod nano 4GB ホワイト

5.1 評価属性の取得実験

4.2節の評価属性の取得手法によって取得された評価属性のうち、出現頻度が上位10件のものを表1に示す。ただし、それぞれの評価属性の出現頻度は、4.3節に基づき集約された語の出現頻度を含めた合計である。表記されている評価属性は、集約された評価属性の中で最も出現頻度が高いものである。

表1 評価属性

ドラゴンクエスト	複合機 EP-802A	1Q84 BOOK 1	iPod nano
ストーリー	インク	作品	音質
ゲーム	音	本	傷
ドラクエ	写真	小説	容量
クエスト	プリンター	世界	nano
テンポ	色	文章	ケース
敵	速度	作者	液晶
要素	スピード	言葉	価格
ボス	画質	文体	画面
妖精	使い勝手	格	カラーボディ
レベル上げ	画面	心	本体

ゲームソフトの「ドラゴンクエスト IX」では「ストーリー」、
「クエスト」、「セーブデータ」といったゲームの内容やシステム面の語が評価属性として取得できた。また「EPSON Colorio インクジェット複合機 EP-802A」では「音」や「スピード」、
「Apple iPod nano 4GB ホワイト」では「音質」や「容量」等

のそれぞれ商品に関する性能が評価属性として取得できた。小説の "1Q84 BOOK 1" では「話自体」「文章」といった小説の内容に関する評価属性を取得できた。

ここで、それぞれの商品の評価属性で適切でないものを次に示す

- ドラゴンクエスト IX
ゲーム, ドラクエ, 要素
- EPSON Colorio インクジェット複合機 EP-802A
プリンター
- 1Q84 BOOK 1
作品, 本, 小説, 作者, 格, 心
- Apple iPod nano 4GB ホワイト
nano, ケース, 価格, 本体

適切でない語には「ドラクエ」「プリンター」「小説」「作品」「本体」のようにそれぞれ商品自体を指す表現が取得されてしまった。また「ゲーム」「小説」「本」のような商品自体の上位概念にあたる語が出現頻度の上位として取得された。これらの語は商品の評価の基準となる部分ではないため、評価属性としては不適切である。そこで、このような評価属性に対しては、商品との概念の関係から適切であるかどうかを判定することにより解決できると考える。

また、「要素」のように語自体が汎用性を帯びているものや、「ケース」という商品のオプションとなる語が評価属性として取得された。前者は、前につく語によって様々な意味に変わるため、評価属性としては適切でないと考えられる。後者に対しても、商品とは別のものであるため、適切でない。このような語は概念の比較からは判断できず、出現頻度も高いため排除は難しい。そのため、現在の評価属性の取得に関する手法を改善する必要があると考える。

5.2 評価属性の集約実験

4.3 節で述べた評価属性の集約の結果を表 2 に示す。表 2 の集約数は、基準となる評価属性に対して、集約の対象となる語の数である。基準は 5.1 節で示した評価属性である。また、適合率 pre は集約数 sum_num に対する正解率 crt_num より次の式で算出した。

$$pre = sum_num / crt_num \quad (6)$$

"1Q84 BOOK 1" に関しては、「世界」と「別世界」という組み合わせのみしか集約が行われなかったため、表には記さないものとする。ただし、表 2 の全体には、上記の組み合わせを考慮した結果を示す。

表 2 評価属性の集約結果

対象	集約数	正解数	適合率
ドラゴンクエスト	29	19	0.655
複合機 EP-802A	10	7	0.700
iPod nano 4GB ホワイト	4	1	0.250
全体	44	27	0.614

"ドラゴンクエスト IX" と "EPSON Colorio インクジェット複合機 EP-802A" では、それぞれ 0.6 以上の適合率となっ

たが、"Apple iPod nano 4GB ホワイト" では 4 つの集約の組み合わせの内、正解は 1 つしか確認できず、0.25 の適合率となった。

また、"EPSON Colorio インクジェット複合機 EP-802A" の複合機では「速度」と「スピード」が別々の評価属性として取得されたが、それぞれ印刷のスピードを意味しているため、この 2 語は集約されるべきである。しかし、現在の手法では文字列の比較でのみ評価属性同士の類似を確認しているため、このような語の関係には対応できない。そこで、今後は概念辞書を用いた語の関係の比較を集約処理に用いることを検討している。

5.3 評価文の抽出実験

4.4 節で述べた評価文抽出の結果を表 3 に示す。ただし、適合率 (緩) は分割した単位文中に別の属性に対する評価が記述されている場合でも正解と扱った場合である。適合率 (厳) は注目している評価属性に対してのみ評価がされている単位文を正解とした場合である。

表 3 評価属性

対象	評価文数	適合率 (緩)	適合率 (厳)
ドラゴンクエスト	1201	0.605	0.403
複合機 EP-802A	102	0.796	0.437
1Q84 BOOK 1	62	0.661	0.403
iPod nano 4GB ホワイト	108	0.610	0.410
全体	1471	0.623	0.390

取得した評価文のうち、1 文中に複数の評価属性に対する評価が記述されていることが多く、全体的な適合率 (厳) は 0.39 となった。これは、単位文分割を文末記号を基準として行っているため、一文中に複数の評価属性に対する評価が記述されている場合に対応できないためであると考えられる。また、文末記号がないレビューが存在するため、そのようなレビューに対しては現在の単位文分割では対応できない。そこで、読点 (「、」) のような文の途中の区切りや、改行コード、箇条書きに用いる記号 (「・」や「」) も基準に加えた単位文分割を行い、評価属性と接続表現を基準とした単位文の再接合、再分解による評価文の抽出を考案する。

5.4 評価文の有用性判定実験

4.5.2 節で述べた根拠の有無判定手法を用いて、実際に抽出した評価文から根拠を伴った評価文を取得したところ、次のような評価文が取得できた。

- ドラゴンクエスト IX
「あと、無駄に動くので戦闘のテンポが悪くなってます。」
- EPSON Colorio インクジェット複合機 EP-802A
「なんと言っても無線 LAN 機能があるので、置き場所は選ばないまでもいいけど使い勝手はとても良い。」
「特に黒色インクのタンクが他の色と同じ大きさなので直ぐに無くなってしまいます。」

"ドラゴンクエスト IX" の評価文では「テンポ」に対する「悪い」という評価に「無駄に動く」といった根拠を伴っていることがわかる。他の評価文も同様に、評価属性に対する評価に根拠が伴っており、評価の補足的な情報を読み取ることができる。

しかし、抽出した根拠のある評価文のほとんどが「クエストや通信が楽しいので寄り道ばかりしてます」のように、評価属性に対する評価を根拠とするような文章が多く、上記のような評価を証明する根拠を伴った評価文はほぼ抽出することができなかった。そこで、今後は評価属性と評価に用いられている語の関係から、評価文の有用性を判定する。

6. おわりに

本稿では、対象とする商品のレビューを閲覧する負担を解消するために、容易な商品評価の把握を可能とする提示手法を提案した。提案手法では、レビューから商品の機能や性能について評価が記述されている評価文に着目し、評価文から評価を抽出するために、評価属性、評価極性、根拠という要素を定義し、抽出される評価を属性評価モデルとしてモデル化した。そして、レビュー集合から評価を抽出するための各処理に関して述べた。

また、評価属性の取得、評価属性の集約、評価文の抽出、評価文の有用性判定に関する実験結果と考察を述べた。評価属性の取得実験では、取得された評価属性には商品と同じか、上位の概念である語が含まれていたため、概念の比較が必要であることが判明した。評価属性の集約実験に関しては、語の類似のみで判定するのではなく、概念の比較を用いることで、「速度」と「スピード」のような組み合わせを集約できるよう改善する必要があることが分かった。評価文の抽出実験では、現段階の単位文分割では、一文中に複数の評価属性に対する評価が記述された文章には対応できないため、分割の手法を改善し、抽出の際の手法を改善する必要があることがわかった。また、根拠を伴った評価文を実際に抽出したところ、評価を証明する根拠を有する評価文は非常に少数であったため、今後は語の関係から有用性を判定する手法を検討する。今後の課題は、取得した評価文の極性判定を行う手法の実装と、抽出した評価の可視化を行うことである。

文 献

- [1] 谷本融紀, 太田学, 評価表現辞書を用いた評判情報の可視化, Webとデータベースに関するフォーラム (WebDB Forum2009), 1B-1. 2009.
- [2] 小林のぞみ, 乾健太郎, 松本祐治, 立石健二, 福島俊一. テキストマイニングによる評価表現の収集, 情報処理学会研究報告, 2003-NL-154, pp. 77-84, 2003.
- [3] 小倉達矢, 穴戸開, 今藤紀子, 山口実靖, 浅谷耕一, レビューサイトにおける良質なレビューの特性とそれを考慮した評判情報の抽出に関する一考察. DEWS2008, B8-5. 2008.
- [4] 青木伸也, 湯本高行, 角谷和俊, 新居学, 高橋豊. 論点に対する極性に注目したニュース記事からの編集意図の抽出手法, 第149回DBS研究会, DBS-2A. 2009.
- [5] 日本語評価極性辞書, <http://cl.naist.jp/~inui/research/EM/sentiment-lexicon.html>.
- [6] 単語感情極性対応表, <http://www.lr.pi.titech.ac.jp/~takamura/pndic-ja.html>.