

商品レビュー要約のための属性-意見ペア抽出

中野 裕介¹ 湯本 高行¹ 新居 学¹ 佐藤 邦弘¹

概要: 通販サイトにおける商品レビューは、ユーザにとって商品購入を検討する際に有益な情報として利用される。しかし、レビューの投稿数が多数の場合には、それらをユーザが閲覧することは負担の大きい作業となる。そこで、本研究では、商品に対する評判を属性と意見から構成される属性-意見ペアとして定義し、ある商品における投稿レビューから、レビューの要約として属性-意見ペアの集合を抽出する手法を提案する。本手法では、レビューの各文中の文節に含まれる自立語をキーワードとし、文の係り受け構造に着目して係り元および係り先のキーワードから属性-意見ペアを抽出する。属性-意見ペアの提示により、商品のある側面に対しての記述者から意見を効率的に把握可能にする。

1. はじめに

近年、インターネットを通じての商品取引が増加しており、その一例として、通販サイトを利用したオンラインショッピングがある。Amazon.co.jp[1]のような通販サイトでは、商品についての評判情報として図1のようなカスタマーレビューが提供されている。

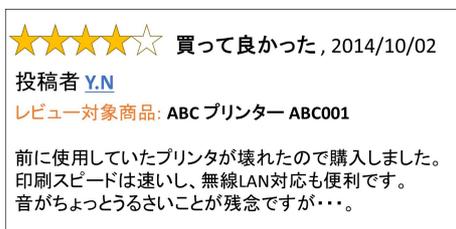


図1 複合機のカスタマーレビュー

図1は商品についての5段階評価である評点および、商品に関しての感想を述べた記事で構成されている。レビューは1つの商品につき複数投稿されており、図2のような各レビューの評点の総計結果が提供されている。



図2 レビューの総計

図1, 図2は、通販サイト利用者にとって、商品購入の意思決定の際に有益な情報として活用することができる。図2の全レビューの評点の集計結果を閲覧することにより、商品についての記述者からの評価度合いを大まかに把握することが可能となる。さらに、商品に関する詳細な評価が知りたい場合には、図1の個別のレビューを閲覧することで、具体的な感想を知ることができる。

しかし、レビューには様々な情報が含まれており、ユーザが望むような情報が記述された箇所を探索する事は、投稿レビュー数が多数に及ぶ場合、時間というコストを大きく費やす事となる。また、少数のみのレビューを閲覧した場合には、仮に記述者が商品について偏った意見の持ち主であると、閲覧者の商品に対する印象も偏ったものになってしまう恐れがある。

本稿では、レビューの要約として、レビューの文章中の商品のある側面を表す語を属性、属性に対する評価を表す語を意見とし、属性と意見のペアを抽出する手法を提案する。具体的には、「品質が良いです」といった文章から、(品質, 良い)のような語のペアの抽出を行う。レビューから得られたペア集合を集約し、要約結果として図3のように提示することにより、商品についての評判を効率的に把握することが可能であると考えられる。

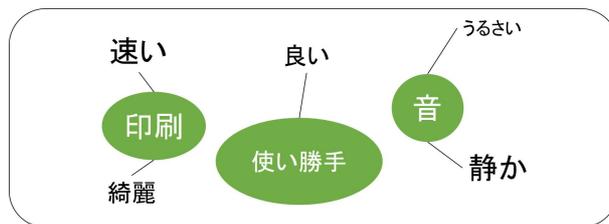


図3 ペアの集約結果

¹ 兵庫県立大学大学院工学研究科

2. 関連研究

インターネットの普及により、個人の意見をネット上に発信する機会が増加しており、このような背景に伴い、レビューや blog 等を対象とし、個人の評価をテキスト中から発見、整理および集約する手法に関する研究が活発に行われている [2].

また、商品レビューや映画レビューから、特徴を表す語における情報を集約する研究が数多く行われている [3][4][5]. これらの研究のように、商品における機能、映画における演出など、レビュー対象の特徴を表す語を特定し、文章中でその特徴の評価は肯定的か否定的かといった極性を推定し、それらについて集約結果を提示する手法が数多く用いられている.

平山ら [6] は、係り受け解析を用いて、商品の属性を抽出し、属性の極性を辞書により判定することで、商品の評判情報を抽出している. また、属性に関連のある文を提示している. 本手法では、属性と対応する意見のペアを結果として提示するため、属性に対しての評判がより直感的に把握できる.

小林ら [7] は、文書に含まれる〈対象, 属性, 評価〉の 3 つ組を評判情報と捉え、それらについて共起パターンを定義し、ブートストラップ法を用いて獲得している. 要約として抽出するモデルは本手法と類似しているが、小林らの手法は表現獲得の際に人手による辞書作成を行っており、全ての手法が機械的に行われる本手法と異なる.

本手法では、係り受け解析を用いて文章中から商品のある側面についてどういった評価が述べられているかを抽出する. 評価となる語を抽出し提示するため、極性の判定は行う必要がなく、より具体的な情報を提供することができる. また、本手法は全て機械的に行われ、人手による辞書作成などのコストがかからない点も特徴である.

3. 属性-意見ペアの抽出

商品のある側面を表す属性および、属性に対する意見となる語から構成される属性-意見ペアをレビューの文章より抽出する. 提案手法の概略を図 4 に示す. 属性-意見ペア抽出までの流れは以下の手順となる.

- (1) レビューの文章を文に分割する
- (2) 文から属性・意見となるキーワードを抽出する
- (3) キーワードを用いて候補ペアを抽出する
- (4) 候補ペアを絞り込み、属性-意見ペアを抽出する

各手順を以下の節にて詳しく記述する. また、本手法では属性-意見ペアの抽出に語の品詞情報や、文の係り受け構造を用いる. そこで、本手法は形態素解析には MeCab[8], 係り受け解析には CaboCha[9][10] を用いる.

3.1 文分割

本手法では、同文に存在する属性と意見に着目するため、レビューの文章を文単位に分割することを行う. 具体的には、一般に文末に用いられる文字 (. ! ?) や、改行で文章を区切り、文に分割する.

3.2 キーワードの抽出

文章中より、属性や意見を表すキーワードの抽出を行う. 文中のどの範囲を属性、意見とするかは様々な解釈が可能であると考えられる. 本手法では、簡潔な要約結果の提示を目的とし、意味をもつ最小単位の語として、1 文節内の自立語をキーワードとみなす. 例えば、「使い勝手が / 良い。」といった文からは“使い勝手”、“良い”をキーワードとして抽出する. 表 1 にキーワード抽出を行う品詞パターンを示す. 括弧 [] 内は形態素における品詞分類を表しており、各文節の形態素解析の結果より、表 1 の品詞パターンに合致する 1 形態素、または形態素列をキーワードとして抽出する. また、キーワードの抽出元となった品詞パターンにより、キーワードを体言か用言にタイプを分類する. キーワードのタイプは後述の属性-意見ペアの抽出条件に用いられる.

品詞パターンは 1 形態素のみならず、複数形態素からなる表現にも対応する. 具体的には、名詞の連なりは複合名詞としてまとめて抽出を行う. さらに、“大き + さ”、“使い + やすい”、“良く + ない”のような付属語を付加することでニュアンスが変化する表現も存在するため、一部の付属語は自立語とまとめて 1 つのキーワードとして抽出する.

抽出範囲を 1 文節に限定することにより、“気に / なる”、“言う / こと / 無し”といった複数文節から構成されるような表現は抽出することができないが、今回は 1 文節内の自立語で意味を持つような簡潔な表現を抽出することに限定し、複数文節からなるような複雑な表現の抽出は今後の課題とする.

表 1 キーワードの品詞パターン

品詞パターン	キーワードのタイプ
[名詞]	体言
[形容詞]	用言
[名詞のうち形容動詞語幹]	用言
[動詞]	用言
[名詞] [名詞] ...	体言
[形容詞] [名詞のうち接尾辞]	体言
[名詞 or 動詞] [形容詞 (にくい, やすい)]	用言
[形容詞] [助動詞 (ない, ぬ, ず)]	用言

3.3 候補ペアの抽出

文集合に対して係り受け解析を行い、係り受け関係にあ

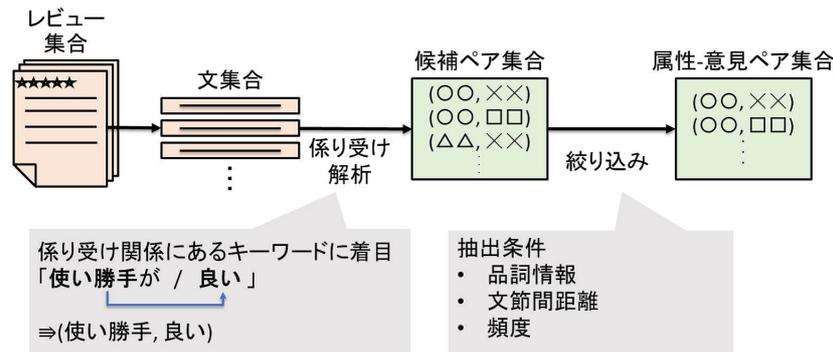


図 4 属性-意見ペアの抽出手法

る文節よりキーワードを抽出し、係り元のキーワード w_{src} 、係り先のキーワード w_{dst} を用いて (w_{src}, w_{dst}) のペアを作成し、これを属性-意見の候補ペアとして抽出する。

3.4 属性-意見ペアの抽出

候補ペア (w_{src}, w_{dst}) から属性-意見ペアを抽出する。その際、有用なペアを抽出し、不要なペアを除去するため候補を絞り込む。絞り込むための抽出条件として、以下に示す情報を用いる。

- w_{src}, w_{dst} および抽出元の文節の品詞情報
- w_{src}, w_{dst} の文節間距離
- w_{src}, w_{dst} の頻度

抽出条件をすべて満たす候補を属性-意見ペアとして抽出する。上記の抽出条件についてそれぞれを以下の項にて詳しく記述する。

3.4.1 品詞情報を用いた絞り込み

候補ペア (w_{src}, w_{dst}) に含まれるキーワード、およびキーワード抽出元の文節の品詞情報を抽出条件とする。抽出条件を以下に示す。

- (1) w_{src} が体言かつ w_{dst} が用言
- (2) w_{dst} が動詞で1形態素からなる語でない
- (3) c_{src} に格助詞か係助詞(が, は, も)の助詞を1形態素だけ含む
- (4) c_{dst} に助詞を含まない

上記条件では、文節を c とし、キーワード w_{src}, w_{dst} の抽出元の文節を c_{src}, c_{dst} と表している。条件 (1) により、表 1 のキーワードのタイプを参照し、係り元が体言、係り先が用言であるペアを抽出する。この条件により、語の組合せから意味が読み取れない候補ペアを削減する。また、条件 (2) により、動詞のような動作を表す品詞は意見として不適当であるとみなし、 w_{dst} が動詞かつ、1形態素のみからなる候補は抽出しない。さらに、条件 (3) と条件 (4) は文節の助詞に着目した抽出条件となっており、「 w_{src} (が, は, も) w_{dst} 」といった簡潔に書かれた文から作成されたペアのみを抽出する。

3.4.2 文節間距離を用いた絞り込み

文章中の語は、文節が近いほどそれらの意味の結びつきは強く、反対に遠いほど意味の結びつきは弱いと考えられる。よって、候補ペア (w_{src}, w_{dst}) の w_{src} と w_{dst} の文節間距離を $d(w_{src}, w_{dst})$ とし、しきい値 th_d を設定し、以下の式を満たす候補のみを抽出する。

$$d(w_{src}, w_{dst}) \leq th_d \quad (1)$$

ここで、文節間距離 $d(w_{src}, w_{dst})$ は、キーワード w_{src}, w_{dst} の文節のインデックス番号 i_{src}, i_{dst} を用いて、 $i_{dst} - i_{src}$ により算出する。よって、隣接する文節の文節間距離は 1 となる。

3.4.3 頻度を用いた絞り込み

文書集合で多く出現する語ほど、その文書集合において一般的で重要である表現であると考えられる。よって、候補ペア (w_{src}, w_{dst}) において、属性候補である w_{src} の商品における投稿レビューの集合 R の DF 値を $DF(w_{src}, R)$ とし、しきい値 th_f を設定し以下の式を満たす候補のみを抽出する。

$$DF(w_{src}, R) \geq th_f \quad (2)$$

4. 実験

提案手法の有効性を検証するための評価実験を行った。具体的には、人手により (属性, 意見) ペアの正解データを作成し、提案手法により機械的に抽出した属性-意見ペア集合との比較を行う事で評価を行った。

実験データとしては、Amazon.co.jp における商品より以下の 3 商品を選択し、それぞれの全投稿レビューから無作為に抽出した 100 件のレビューを用いた。

- CASIO 掛け時計 アナログ IQ-126-5JF アナログタイプ
- Canon インクジェット複合機 PIXUS MG3230
- TWINBIRD サイクロンスティック製クリーナー TC-E117W

4.1 正解データ

被験者 1 名にレビューを閲覧させ、属性-意見ペアを抽出させた。具体的な手順としては、文に分割され、CaboCha により文節区切りのガイドが付いたレビューに対して、以下の条件を満たすように属性-意見ペアを抽出させた。

- 属性と意見は 1 文節内より選択する。
- 同文に出現する属性と意見からペアを作成する

上記の条件に当てはまらないような、複数文節からなる表現や、対応する属性と意見が異なる文に存在するケース、属性が省略されているケースは本手法では考慮しておらず、今回は正解データから除外している。

また、属性-意見ペアの抽出後、被験者が抽出した属性に対して、同義の属性どうしを集約させた。具体的には、“音”と“動作音”、“大きさ”と“サイズ”のように文中で同義である属性どうしに対してタグ付けを行わせ、概念毎の属性集合を作成させた。その後、各属性集合より、商品購入を検討する際に、重要であると感じる属性集合を最大 10 個選択させた。被験者が選択した属性集合に含まれる属性を持つ正解ペアを重要なペアとした。抽出された正解ペアの数を表 2 に示す。

表 2 正解ペアの数

商品	全ペア	重要な属性を含むペア
掛け時計	126	49
複合機	116	43
掃除機	97	71
平均	113	54

4.2 ペア抽出の有効性検証実験

提案手法により抽出されたペア集合 N をテストデータとし、被験者による正解ペア集合 C の正解データとの比較を行い、提案手法の評価を行った。評価には以下の式より算出した再現率 (R)、適合率 (P)、F 値 (F) を用いた。

$$R = \frac{|N \cap C|}{|C|} \quad (3)$$

$$P = \frac{|N \cap C|}{|N|} \quad (4)$$

$$F = \frac{2 * R * P}{R + P} \quad (5)$$

4.2.1 係り受けによる候補ペア抽出の有効性検証

本手法では、候補ペアの抽出を係り受け関係にあるキーワードの組合せに限定している。そこで、本手法による候補ペア抽出と、1 文における全てのキーワードを組み合わせて候補ペアを抽出する手法について再現率と適合率を算出し、比較を行った。結果を表 3 に示す。表 3 では、全てのキーワードの組合せペアに対する抽出数 $\#_{dep}$ 、再現率 R_{dep} 、適合率 P_{dep} とし、係り受け関係にあるキーワードから抽出し

た候補ペアに対する抽出数 $\#_{dep}$ 、再現率 R_{dep} 、適合率 P_{dep} として示している。ここでは、候補ペアがどの程度の正解ペアを網羅しているかに焦点を当て、主に再現率に着目する。 R_{all} の平均の 0.831 と比較し、 R_{dep} は 0.595 であることから、候補ペアを係り受け関係にある語からの抽出に限定することにより、再現率は 0.236 低下している。しかし、 $\#_{dep}$ 、 $\#_{all}$ より、全キーワードの組合せにより抽出される候補ペアの数は、係り受け関係にあるキーワードの組合せより抽出される候補ペアの数よりも 10 倍以上多いため、係り受け関係に抽出を制限することは妥当であるといえる。

また、 R_{all} から R_{dep} の差分より、係り受け関係にない正解ペアは 23.6% 存在することがわかった。この問題に対しては、2 文節間の直接的な係り受け関係だけでなく、係り受け先の係り受け先を参照するといった、間接的に係り受け関係にある語からペアを抽出するといった手法により再現率の改善を試みる。

さらに、 R_{all} は 0.831 であることから、全キーワードの組合せでも 16.9% の正解ペアを抽出できていないということがわかった。抽出漏れのペアには、「見やすい / 文字盤です。」「上質な / デザインです。」のような意見の後に属性が出現する文章から作成される正解ペアがみられた。本手法では係り元を属性、係り先を意見とみなしているため、これらの文からは (見やすい, 文字盤) というように属性と意見が逆順のペアが作成されており、正解とみなされていない。よって、用言から体言に係る場合は、係り元を意見、係り先を属性とみなして属性-意見ペアを抽出するような処理が必要である。他には、「コスト最高。」のように同じ文節内の属性と意見から作成される正解ペアは抽出できておらず、再現率低下の原因となっていた。

表 3 候補ペアの抽出結果

商品	$\#_{dep}$	R_{dep}	P_{dep}	$\#_{all}$	R_{all}	P_{all}
掛け時計	1637	0.595	0.045	18659	0.873	0.005
複合機	1556	0.546	0.034	21580	0.824	0.003
掃除機	2256	0.637	0.032	29785	0.793	0.003
平均	1816	0.595	0.037	23341	0.831	0.004

4.2.2 属性-意見ペア抽出の有効性検証

まず、予備実験により、文節間距離のしきい値 th_d 、頻度のしきい値 th_f のしきい値を決定した。手順は、始めに文節間距離、頻度のいずれかの抽出条件のみにより候補ペアを絞り込み、 th_d 、 th_f の値を変化させながら、評価値を算出した。その結果より、しきい値を変化させた範囲内で、最も F 値が高くなったしきい値を採用した。 th_d 、 th_f の値を変化させたときの各評価値の結果を図 5、図 6 に示す。ここの評価値は、実験商品における評価値の平均を示している。図 5、図 6 より、F 値が最大となるときのしきい値は $th_d = 1$ 、 $th_f = 8$ と求めた。

続いて、3.4.1 項における品詞情報を用いた絞り込み条

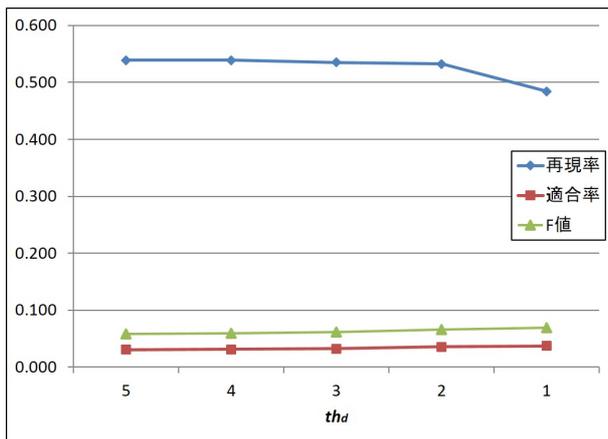


図 5 文節間距離を条件としたペア抽出結果

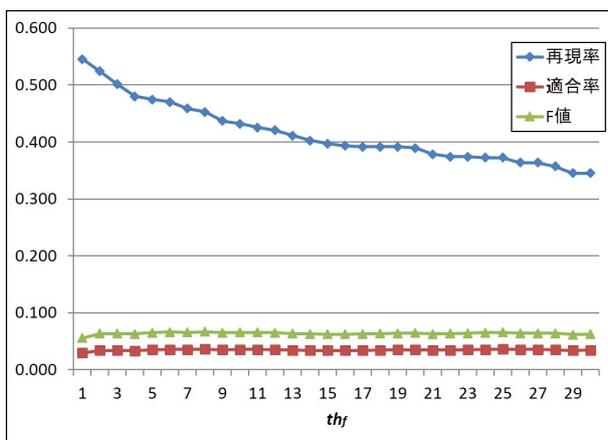


図 6 頻度を条件としたペア抽出結果

件に加え、予備実験により求めた文節間距離のしきい値 $th_d = 1$ 、頻度のしきい値 $th_f = 8$ を絞り込み条件として属性-意見ペアの抽出を行った。各商品において抽出された属性-意見ペア集合を表 4、評価値を表 5 に示す。ここで、表 4 のペアは意見を終止形に変換し、重複するペアを集約した結果を示している。

表 5 ペア抽出の有効性検証結果

商品	R	P	F
掛け時計	0.158	0.740	0.261
複合機	0.175	0.500	0.259
掃除機	0.172	0.434	0.246
平均	0.168	0.558	0.255

表 5 の結果より、いずれの商品においても再現率は低く、平均で 0.168 という値となった。抽出できなかった正解ペアについて調査すると、品詞情報による絞り込みにより、抽出されていない正解ペアが存在していた。代表的なものとして、「音がする。」、「吸引力が落ちる。」から作成される(音, する), (吸引力, 落ちる) ような、意見が動詞となるペアが比較的多く見られた。その他の絞り込み条件についても、条件に当てはまらず抽出されていない正解ペアが存在しており、抽出漏れした正解ペアについて分析を行い、抽出

条件を見直す必要がある。

また、適合率は商品においてばらつきがあるが、これは各商品でレビューの文章に出現する表現が異なることが原因であると考えられる。掛け時計と比べ、複合機や掃除機は特徴が多く、キーワードの抽出パターンや、品詞条件による絞り込みが不十分であり多様な表現に対応できなかったため、適合率が低下したと考えられる。具体的には、以下のような抽出に失敗しているペアがみられた。下線部分が属性と意見、括弧内は作成された属性-意見ペアを表している。

- Data の印刷ならば 問題 は 無い。(問題, 無い)
- 部屋 が 狭い というのもあると思いますが～(部屋, 狭い)
- 評価 が よかった ので購入しました。(評価, よい)

1つ目の例は、“問題は無い”という部分が1つの意見として抽出されるべきであるが、本手法では属性・意見の抽出範囲を1文節に限定しているため、(問題, 無い)と誤ったペアで抽出されているという問題である。よって、意見部分の抽出範囲を複数文節に拡張する必要がある。

2つ目と3つ目の例は、記述者の商品の評価に直接関係がない文章からペアが抽出されていることが問題となっている。これらのペアの抽出を防ぐため、係り受け関係にある語が含まれる文節内の品詞情報だけでなく、さらに周辺の形態素の情報を用いたルールを新たに設定する必要がある。

4.2.3 要約としてのペア抽出結果の評価

4.1 節における正解データにおいて、被験者が抽出したペアの総数の平均は 113 ペアであり、全てのペアを提示した場合には、数が多いため閲覧者の負担になると考えられる。よって、抽出できたペアのうち、より購入検討者にとって有用なペアを選定し提示する必要がある。ここでは、表 2 における重要な属性を含むペアを用いて、本手法で有用なペアがどれほど抽出できているかを再現率により算出した。また、重要な属性集合の被覆率 Cov として、被験者が選択した複数の重要な属性集合のうち、正解した抽出ペアの属性を含む重要な属性集合の割合を算出した。結果を表 6 に示す。

表 6 要約としての評価結果

商品	R	Cov
掛け時計	0.183	(2/7=) 0.285
複合機	0.186	(4/7=) 0.571
掃除機	0.169	(4/6=) 0.667
平均	0.179	0.508

表 6 の結果より、再現率はいずれの商品においても低い値となった。また、被覆率においても平均 0.508 と低い結果となった。頻度により候補ペアを絞り込むことで、有用な属性を含むペアが抽出されていると予想していたが、同義語どうしの頻度を足し合わせるといった工夫を行っていないため、“音”と“動作音”、“サイズ”と“大きさ”のよう

表 4 属性-意見ペア抽出結果

掛け時計	複合機	掃除機
(アナログ, 好き)	(インク, 四角い)	(アタッチメント, しょぼい)
(サイズ, コンパクト)	(インク, 少ない)	(コード, 短い)
(デザイン, よい)	(インストールソフト, 多い)	(ごみ捨て, 簡単)
(デザイン, 素敵)	(ソフト, メイン)	(サイクロン, 良くない)
(音, ない)	(デザイン, 台無し)	(サイズ, 小さい)
(音, 静か)	(プリンター, ありがたい)	(ただお値段, 安い)
(何年, 故障)	(プリント, 遅い)	(たより, 軽い)
(外枠, 濃い)	(印刷, きれい)	(デザイン, いい)
(質感, 美しい)	(印刷, メイン)	(テスト, 欲しい)
(書体, シンプル)	(印刷, 早い)	(ところ, 強い)
(消費, 激しい)	(印刷機能, 十分)	(やる気, ない)
(値段, 安い)	(印刷品質, 良く)	(音, 嫌い)
(値段, 手頃)	(印字, きれい)	(吸引, 十分)
(品, 良い)	(音, 大きい)	(吸引力, 強い)
(部屋, 狭い)	(価格, 安い)	(吸引力, 申し分)
(部分, 小さい)	(確認, 必要)	(恐れ, ない)
(文字盤, 小さい)	(機械音, うるさい)	(構造, シンプル)
(木製, 良い)	(機能, 十分)	(仕方, ない)
(枠, 大きい分)	(使い心地, いい)	(持ち運び, 楽)
	(紙詰まり, ない)	(収納, コンパクト)
	(質, 十分)	(収納, ない)
	(手間, 馬鹿)	(収納, 楽)
	(修理, 必要)	(場所, 多い)
	(性能, 低い)	(人, 多い)
	(接続, 便利)	(性能, 凄い)
	(設定, 簡単)	(掃除, 必要)
	(大きさ, いい)	(値段, 安い)
	(値段, 安い)	(注意, 必要)
	(調子, 悪い)	(能力, 高い)
	(反応, 遅い)	(分解, 簡単)
	(評価, よい)	(分解掃除, 簡単)
	(殆ど差, ない)	(目詰まり, 早い)
	(無線ラン, よい)	(予定, ない)
	(無線通信, 遅い)	(量, 凄い)
	(問題, 無い)	(力, 十分)

に同義で異なる文字列で出現する語に対しては頻度が分散するといった問題がある。よって、同義語を多く含む属性集合を被験者が重要と選択した場合、頻度による絞り込みが有効に作用していないことが再現率および被覆率の低下を招いたと考えられる。この問題については、WordNet[11]のような同義語辞書の言語資源を活用し同義語を集約することで、有用なペアが頻度によって上位にランキングされるように改善を行う。

5. おわりに

本稿では、商品レビューの要約として属性-意見ペアを抽出する手法を提案した。係り受け関係にあるキーワードから候補ペアを作成し、品詞情報、文節間距離、頻度を抽出条件としてペアの抽出を行った。

ペア抽出の有効性検証の結果、再現率の平均は 0.168、適

合率の平均は 0.558、F 値の平均は 0.255 という結果となった。再現率より、直接係り受け関係にない属性と意見の組合せからなるペアや、意見に動詞を含むペアの抽出が課題であることがわかった。また、適合率より、複数文節からなる意見の抽出失敗、商品の評価と直接関係ない文からのペア抽出について手法の改善が必要であることがわかった。抽出漏れのペア、抽出に失敗したペアについて分析を行い、ペア抽出のルールの見直し、および追加を行い F 値の向上を目指す。

また、重要な属性を含むペアにおける再現率の平均は 0.179、重要な属性の被覆率は 0.508 という結果となった。これについては、同義語を集約し頻度の算出を行うことで、有用な属性を含むペアが上位にランキングされるように処理を改善し、ユーザが重要と考える有用なペアの抽出精度の向上を目指す。

謝辞 本研究の一部は、平成 26 年度科研費若手研究 (B) 「情報の詳細関係に基づく Web ページの組織化」 (課題番号: 24700097) によるものである。ここに記して謝意を表すものとします。

参考文献

- [1] Amazon.co.jp, <http://www.amazon.co.jp/>
- [2] 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.13, No.3, pp.201-241, (2006).
- [3] Minqing Hu, Bing Liu: Mining and Summarizing Customer Reviews, In Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 22-25, (2004).
- [4] Binq Liu, Minqing Hu, Junsheng Cheng: Opinion Observer: Analyzing and Comparing Opinions on the Web, In Proceedings of International Conference on World Wide Web (WWW), 342-351, (2005).
- [5] Li Zhuang, Feng Jing, Xiao-Yan Zhu: Movie review mining and summarization, In Proceedings of ACM International Conference on Information and Knowledge Management (CIKM), pp.43-50 (2006).
- [6] 平山拓央, 湯本高行, 新居学, 佐藤邦弘: 語の共起と極性に基づく商品レビュー閲覧システム, 情報処理学会研究報告, 2012-DBS-155, 1-9, (2012).
- [7] 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, 2002-NL-154, pp.77-84, (2003).
- [8] MeCab, 入手先 (<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>)
- [9] CaboCha, 入手先 (<https://code.google.com/p/cabocho/>)
- [10] 工藤 拓, 松本 裕治: チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, 43(6), pp.1834-1842, (2002).
- [11] 日本語 WordNet, <http://nlpwww.nict.go.jp/wn-ja/>