

属性-評価ペアを単位とした評判情報の要約

菊池 悠太^{1,a)} 高村 大也^{1,b)} 奥村 学^{1,c)}

概要: 本稿では, レビュー文書中におけるレビュアーの意見を要約する枠組みについての提案を行う. ある製品に対するレビュー文書集合から, 意見を表す評価表現を抽出し, それらを代表する評価表現の部分集合を選択して提示することにより, 要約とする. 今回は, 意見を表す単位として属性語, 評価語のペア(以下, 属性-評価ペア)を利用する. 構文木の係り受け構造から属性-評価ペアを抽出し, 抽出したペアの中から, 全体を被覆するペアの部分集合を選択する問題を, 施設配置問題によりモデル化する.

キーワード: 評判要約, 文書要約, 施設配置問題, 評判抽出

Generating a Review Summary as a Set of Aspect-Evaluation Pairs

YUTA KIKUCHI^{1,a)} HIROYA TAKAMURA^{1,b)} MANABU OKUMURA^{1,c)}

Abstract: We propose a method for opinion summarization of product reviews. Our method generates a summary by identifying evaluative expressions in the reviews on a product, and selecting representative ones from the identified expressions. As evaluative expressions, we focus on pairs of an aspect and its evaluation, which are identified out of dependency parse trees. We employ the facility location summarization model to select representative aspect-evaluation pairs.

Keywords: opinion summarization, text summarization, facility location problem, opinion mining

1. はじめに

ウェブ上には多くの意見や評判が存在しており, その数はウェブの普及や発達に伴って増加し続けている. 中でもそのような情報が多く集まる場所として, 通信販売サイトが挙げられる. それらのサイトでは, ユーザは購入した製品に対し, その製品を実際に利用した際に感じた意見や感想を, レビューという形で残すことができる. これらのレビューは, 他のユーザにとっては購買意思決定のための有用な参考意見となる. また企業にとっても, 顧客の意見を自社製品の改善へと活かすための有用な情報源となり得る. しかし, 製品によってはそのレビューの数は膨大なものになり, それらすべてを読んで把握することは負担のか

かる作業となる. その上, 複数の製品についてそれらのレビュー内容を比較しようとするのは困難である. 本研究では, 評判情報が多く含まれており, なおかつ評判情報の対象が絞りやすい通信販売サイト中のレビュー文書を対象に, 評判情報の要約について取り組む.

製品に対する評判情報という観点で, 重要となる情報のうち最も基本的なものは, その製品の“どの側面”が, “どのような評価”を得ているのかという情報であると考えられる. このような, 評価の軸となりうる, 製品の特徴を表す重要な側面を示す語を属性語, その属性語に対する具体的な評価を表す語を評価語と呼ぶ. 本研究では, この属性語と評価語を対にして抽出したものを属性-評価ペアと呼び, 要約を行う単位とする.

例えば掃除機に対する“吸引力が強い”という表現から, “吸引力”という属性に対して, “強い”という評価がされていることがわかる. それぞれを属性語, 評価語として“吸引力-強い”という形で抽出したものが, 属性-評価ペアで

¹ 東京工業大学
Tokyo Institute of Technology

a) kikuchi@lr.pi.titech.ac.jp

b) oku@lr.pi.titech.ac.jp

c) takamura@pi.titech.ac.jp

ある。

一般的に、レビュー文書を参照する目的は、文中の評判情報の把握であり、必ずしも文全体を読む必要はない。例えば、「この掃除機は音は大きいですが、掃除機はうるさいものと割り切って特に気にしてませんし、私自身はもう慣れちゃいました。」という文において、この文の内容は実質的には“音が大きい”という表現で代替可能であり、この情報さえあれば、この文の後半にあるレビュアーの個人的な立場表明やエピソードは必ずしも読む必要はない。そのため、評判情報の要約という観点では、文そのものではなく、属性-評価ペアのような評価表現を単位として扱うことで、レビューに特化した、評判情報の要約の生成が可能になる。

我々の提案する要約手法は、まず与えられたある製品に対するレビュー文書クラスタから、評価表現として、属性-評価ペアを抽出する。次に、抽出した全ての属性-評価ペアの集合から、ペア集合全体を内容的に被覆するペアの部分集合を選択することにより、要約を生成する。

ペアの抽出には、事前に定義した抽出規則を用いる。これは、評価表現の特徴から、品詞情報と係り受け関係に基づいた規則を設定した。

ペアの選択については、整数計画問題の一種である施設配置問題としてモデル化を行う。施設配置問題は、全ての顧客がいずれかの施設へと割り当てられるという設定において、顧客全員の利得（移動距離などのコスト）が最適となるような施設の配置を決める最適化問題である。このとき、顧客そのものを、他の顧客が割り当てられる施設として扱うような問題を考えることもできる。すなわち、施設配置問題としてモデル化することにより、施設として選択されたペア集合が、入力されたペア集合全体を内容的に被覆するようなモデルを考えることができる。

この問題を解くことにより、施設として選択されたペア集合を代表ペアとして提示することで、入力されたペア集合の要約とみなすことができる。また、ペアがレビュー文書における重要な評価表現を表していることから、代表ペアは与えられたレビュー文書全体の評判情報を要約したものであるとみなすことができる。

2. 関連研究

従来の文書要約タスクでは、主にニュース記事を対象とした研究が主流であったが、近年のウェブの急速な発達により、ウェブ上のあらゆる文書が爆発的に増加し、様々なデータが要約の対象となっている。評判情報も例外ではなく、これまで多くの研究がなされている。

2.1 文書要約

文書要約とは、与えられた文書クラスタから、その内容を簡潔に表す短い文書を要約として提示する研究タスクで

ある。文書クラスタは単一の文書からなる場合と、複数文書からなる場合とがある。レビュー文書を扱う評判要約では、一つの製品に対して複数のレビュアーが各々独立してレビュー記事を投稿するため、必然的に複数文書要約の問題を取り扱うことになる。

文書要約に対するアプローチは二種類に大別できる。文書中の重要な要素をそのままの形で抽出する extractive な要約手法と、必ずしも文書中の表現を使わずに、入力文書クラスタの内容を表す要約文書を独自に生成する abstractive な要約手法である。従来の文書要約では extractive な手法が主流である。多くの研究は要素として文を単位とし、文に対してスコアを定義することで、重要な文集合を選択して要約として提示する。特に複数文書要約の場合は、冗長性や文の提示順序について考慮することも重要な課題となる。

高村らは、要約において重要となる、文間での情報の被覆性を実現する手段として、施設配置問題による複数文書要約のモデル化を行った [1]。本稿の提案モデルは彼らのモデルを参考としており、選択単位を文ではなく属性-評価ペアとしている。

以降は、従来の文書要約手法に沿ったアプローチで評判要約という問題に取り組んだ研究の例を挙げる。また、評判情報という、要約対象の特性を活かして、これまでの文書要約とは大きく異なるアプローチで評判情報の要約に取り組む研究も近年盛んであるため、いくつか例を挙げる。

2.2 従来の文書要約手法の評判情報への適用

上述の通り、従来の文書要約の分野では、文の重要度を定義し、重要な文を選択することにより、要約を作成するというアプローチが盛んに行われている。このような文選択のアプローチを意見や評判文書の要約に適用する場合、意見や評判に特化した特徴を利用することになる。

西川らは、文の重要度として、意見情報の表現単位である concept とその重みを定義した。そして、要約として選択した文集合の持つ concept によって、文書全体の concept を被覆するよう最大被覆問題として評判情報の要約をモデル化した。彼らは巡回セールスマン問題をモデルに組み込むことで、文の一貫性も同時に最適化することに取り組んでいる [2]。

嶋田らは、レビューに定義された複数の評価軸と、それらに対してレビュアーが付けた評点をを用いて、各評価軸のバランスを取った要約を行った。要約の冗長性を避けるために、文のクラスタリングを行い、各クラスタの代表文を、関連するいずれかの評価軸に割り当てることで、冗長性と評価軸のバランスとを考慮した要約の生成に取り組んだ [3]。

いずれの研究も、文選択というアプローチは従来の文書要約と同様だが、レビュー特有の情報を取り入れることで、

レビューの要約生成に取り組んでいる。

本研究では、文を単位として用いずに、属性-評価ペアを単位とした要約に取り組む。その利点は、評判情報という観点で最も重要となる情報を直接的に扱うことで、無駄な情報を省いた要約の生成が期待できる点である。

2.3 属性語をベースとした評判要約

近年、従来の文書要約とは大きく異なるアプローチで評判情報の要約を生成しようとする研究も盛んに行われている。Aspect-based opinion summarization[4]と呼ばれるこのタスクでは、与えられた文書の集合に対し、従来の文書要約に見られたような短くまとめられた自然文を出力するのではなく、より構造的な形式で評判情報を扱う。例えば製品を対象とした場合、その製品を評価する際の着眼点として相応しいであろう特徴や性質を表す属性語を同定し、各属性語についてどのような言及がなされているかを、感情分析などの技術を用いて推定する。そして、最終的には各属性語についての意見の分布を、推定した感情極性に基づいて集約し、要約として提示する。

本タスクは大きく3つのサブタスクに分けることができる。文書中から、その製品、あるいは製品カテゴリにとって、要約対象として重要となる属性語を同定する Aspect identification、同定した属性語に対して、レビュー文書中で言及されている評価表現の感情極性を推定する Sentiment prediction、そしてそれらの結果から、各属性語についての評価をまとめて提示する Summary generation である。

2.3.1 Aspect identification

本タスクでは、入力文書中で要約対象となっているものの中で、どのトピックが要約されるべき重要な情報であるかを同定するタスクである。例えば iPod を対象とした場合は、“音質”、“電池容量”などが考えられる。このような、製品の特徴や性質を具体的に表す語を属性語と呼び、これらを文書中から同定する。

Luらは、短いコメントが多くなされるデータセットに注目した。そのようなデータにおいては、評価表現が名詞と形容詞の係り受け関係によって表されることが多いことに着目し、構文木からそのような関係になっている名詞と形容詞のペアを抽出している。彼らはその後、名詞について、どのような形容詞に修飾されているかについてクラスタリングを行うことで、 k 個の重要な属性語を同定している [5]。

2.3.2 Sentiment prediction

本タスクは、同定した各属性語について、文書中でどのような意見を述べられているかを予測するタスクである。多くのアプローチでは、属性語に関連した文について、評価の内容がポジティブかネガティブかの感情極性を予測する。

Luらは、レビュアーがレビュー単位で付けた評点の集

合から、抽出した属性語の評点を推定した [5]。

Huらは、評価語を形容詞に限定し、WordNet のリンク構造を用いて、seed 語との関係を見ることで、評価語の極性を決定した [6]。

2.3.3 Summary generation

前述の二つのサブタスクで収集した、属性語の集合と、それに対する評価極性から要約を作成する。要約の形式としてはいくつかの方法が試されている。

上記の結果を直接的に扱う要約形式としては、同定した各属性語についての評価極性を、統計的にまとめるものが挙げられる。

Huらは、属性語毎にポジティブな意見とネガティブな意見の数を示し、その元となった文を同時に参照可能な形で提示した。ただし、それらの文の選択的な提示はしておらず、ここでいう要約とはあくまで属性語毎の評価の感情極性の分布を意味する [7]。

このような形式は、実際にどのような言及がなされているかということよりも、その製品に対する評価を可視化し、その製品に対する評価全体の概観を把握することが目的である。いわばデータマイニング的な立場で、データの構造化に取り組んでいると言える。

実際にレビュアーがどのような言及をしているか、レビュー文書の一部を選択的に提示する研究も存在するが、属性語とその評点について、なぜそのような点数になったかを示すことが目的であり、システム中で評価極性による分類を行なっていることに変わりはない [5]。

対して本研究では、2.2 節における文書要約のように自然言語で記述された要約文書の生成を目的としており、要約について異なる立場を取っている。

データを構造化し、評価の可視化を行うことと、自然言語による要約文書を生成することはそれぞれ異なる利点が存在する。

前者は、その製品に対しての人々の意見の概観を把握するには非常に有用である。しかし、意見の詳細を理解するためには、実際の文書を参照する必要性については避けられない。

例えば、ある属性語について感情極性が拮抗していた場合、実際の文書を読んで、購入者がどのような面で満足しており、どのような面を不満に感じられているのかを見比べる必要がある。また、いくつかの製品間での評価を比較したい時などは、微妙に形の異なるグラフを見比べることよりも、実際にどのような特徴があるのかについてまとめられた文書を読むほうが有用であると考えられる。

3. 属性-評価ペア

本研究で要約の単位として用いる、属性-評価ペアについて説明する。属性-評価ペアとは、レビュー文中から、属性語と評価語を対にして抽出したもので、レビュー文書にお

ける評価表現を構造的に扱う単位としては、最も基本的なものであると考えられる。意見抽出タスクを扱う多くの研究では、評価表現の単位を対象、属性、評価の三つ組として扱っている [8][9]。本研究では、ある製品に対してのレビュー文書を対象とすることから、三つ組における“対象”を“その製品そのもの”とみなし、属性語と評価語のペアを処理の単位としている。本研究の主眼は、評価表現の抽出ではなく、そのような表現を単位とすることの、評判要約タスクにおける有効性であり、従来研究で行われているような複雑な構造化や抽出手法は、今回は考えないこととする。

また、複雑な構造化表現を単位として扱わないことで、特にレビュー文書においては、以下で説明するような単純な抽出規則でも、抽出されるものの多くは我々が抽出目的としている属性-評価ペアに該当する。そのため、今回はこのようなアプローチを選択した。Aspect-based opinion summarization における先行研究で多く行われている、属性語を同定してから、それらに対応する評価表現を収集するというアプローチを取っていないのも同様の理由である。

3.1 属性-評価表現の抽出規則

本研究では、構文解析器の出力した係り受け構造と品詞情報から、事前に定義した規則を用いて属性-評価ペアを抽出する。属性語と評価語はそれぞれ別の規則で判定され、属性語と判定された文節が、評価語と判定された文節に係っている場合にのみ、属性-評価ペアと判定する。構文解析器には CaboCha^{*1} を用いており、以降の品詞体系はその出力に基づいている。

操作の手順は以下の通りである。まず構文解析を行った文を受け取り、文頭から順に文節を参照していく。参照している文節が、格助詞の“が”あるいは係助詞で終わっている場合、属性語候補の抽出規則を適用する。

3.2 属性語の抽出規則

文節の末尾が“が”あるいは係助詞である文節に対して、その前の語が属性語であるかどうかの判定を行う。文節末から文節頭方向へ順に語を参照し、品詞が“名詞”で、品詞細分類が“サ変接続”、“一般”、“固有名詞”、“接尾”のうちいずれかである間、その単語を属性語候補へ含める。ここで、もしも品詞細分類が“接尾”であった場合は、次の単語は品詞にかかわらずに無条件で属性候補語に含める。

文節頭に到達するか、上記の品詞以外の品詞であった場合に、その時点での属性語候補を保持したまま、係り先の文節へと移り、評価後の判定を行う。

3.3 評価語の抽出規則

属性語候補を抽出した文節の係り先文節について、以下の規則により評価語の判定を行う。ここでは、基本的に二種類の判別法を用意している。

一つ目の規則は、文節中における、“すぎる”、“やすい”、“っぽい”、“的”、“にくい”、“づらい”という表現がある場合は、その形態素から文節頭までの形態素を全て評価語として含める。ただし、途中で記号などのノイズが含まれていた場合は、解析誤りである可能性が高いので、その文節から評価語を抽出しないこととする。上記の表現は、品詞による判定ではなく、各形態素の見出し語と一致するかを直接比較している。これは、我々が MeCab^{*2} による形態素解析を行う際に使用している IPA 辞書中の品詞分類を見てみると、上記の表現と同じ品詞分類である語の中に、評価語としてはふさわしくないと判断できる表現も含まれており、品詞による判定が難しいと判断したためである。

次に、前述の表現が含まれなかった文節での、評価語の判定方法について説明する。まず、品詞が“形容詞”で品詞細分類が“自立語”、あるいは品詞が“名詞”で品詞細分類が“形容動詞語幹”である形態素を探し、その形態素の位置が（接頭語を除いた）文節頭であれば、評価語候補に含める。続いて、文節末方向へと形態素を参照していき、品詞が“名詞”で品詞細分類が“非自立”、品詞が“動詞”で、品詞細分類が“非自立”あるいは“接尾”が続く限り評価語候補へと追加する。

最終的に、評価語候補が存在していれば、属性語候補と共に、属性-評価ペアとして抽出する。適切な評価語候補が抽出出来なかった場合は、属性語候補は削除する。

上記の手順でレビュー文書中の各文からペアを抽出していくが、単純な抽出規則のため、取ったペアの中にはノイズも少なからず含まれる。そこで、抽出したペアのうち、属性語が N 回以上出現していないペアは、ノイズとして削除する（今回は $N = 3$ とした）。

与えられたレビュー文書中の全ての文について、以上の抽出規則を適用し、結果として出力されたペア集合を、続く処理である施設配置問題によるペア選択モデルへの入力とする。

4. 施設配置問題による属性-評価ペアの選択

本研究が目的とするタスクを、施設配置問題によりモデル化する。すなわち、あるレビュー文書クラスから、3節における規則により抽出されたペア集合として与え、そこから集合全体を被覆するようなペアの部分集合を選択するようなモデル設計を行う。

*1 <http://code.google.com/p/cabocha/>

*2 <http://code.google.com/p/mecab/>

4.1 施設配置問題

施設配置問題とは、顧客の利得が最大となるような最適な施設の配置を求める最適化問題の総称である。高村らは、施設配置問題の一種であり、施設配置数に制限を設けた予算付きメディア問題として文書要約のモデル化を行った [1]。彼らは、文書クラスタ中の文を顧客とみなし、施設として選択された文集が、文書クラスタ内の全ての文を内容的に被覆するようモデル化を行った。彼らは、ある文 x_j が、施設として選択された文 x_i に、顧客として割り当てられた時に得られる利得を、文中の単語の被覆関係を用いた包含関係のスコアとする非対称なスコアを定義した。

彼らのモデルでは、顧客そのものを施設の配置候補としている点で、一般的な施設配置問題とは異なっている。

本研究における提案モデルは、基本的には高村らのモデルと同一であるが、扱う単位が文ではなく属性-評価ペアという点と、要約長制限が文字列長ではなく選択するペアの数としている点で異なる。

4.2 提案手法のモデル

文書クラスタ内のすべてのペアが、要約として選択されたいずれかのペアによってできる限り表現されるような要約を作るモデルは、以下の整数計画問題で定式化できる。

$$\text{maximize } \sum_{i,j} e_{ij} z_{ij}$$

$$\text{s.t. } z_{ij} \leq x_i; \quad \forall i, j \quad (1)$$

$$\sum_i x_i \leq K; \quad \forall j \quad (2)$$

$$\sum_i z_{ij} = 1; \quad \forall j \quad (3)$$

$$z_{ii} = x_i; \quad \forall i \quad (4)$$

$$x_i \in \{0, 1\}; \quad \forall i \quad (5)$$

$$z_{ij} \in \{0, 1\}; \quad \forall i, j \quad (6)$$

決定変数 x_i は、ペア p_i が要約として選択されたとき 1 となり、そうでないとき 0 となる。決定変数 z_{ij} は、要約として選択されたペア p_i にペア p_j が割り当てられた場合に 1 となり、そうでないとき 0 となる。 $z_{ij} = 1$ のときは、ペア p_i は選択されている必要があり、これは制約式 (1) により保証されている。ペア p_i がペア p_j を被覆している度合いを、 e_{ij} で表し、これをペア間係数と呼ぶことにする。よって、目的関数 $\sum_{i,j} e_{ij} z_{ij}$ は、文書クラスタ内の全てのペアが要約によって表現されている度合いを表していることになり、これを最大化することが我々の目的となる。また、(2) は、要約として選択されるペアの数が K 以下になることを保証する制約式である。また、(3) は、すべてのペアがいずれかのペアに割り当てられることを保証する。さらに、(4) は、要約として選択されたペアは、そのペア自身によって割り当てられることを意味している。

以上により、ペア集合全体から、それらを代表する一部

のペア集合を選択する問題を施設配置問題によるモデルとする。ここで重要となるのは、ペア間係数 e_{ij} をどのように設定するかである。これについて次の節で、いくつかの類似度を定義する。

4.3 ペア間係数の計算方法

施設配置問題の枠組みにおいて、二つのペア p_i と p_j に対して、 p_i を要約へと含めた時に、 p_j は要約に含めるべきか否かを、ペア間係数という形で表現する。高村らの文を単位としたモデルでは、二つの文の単語の重複度に着目し、非対称な含意関係を文間係数としてモデルに組み込んだが、ペアを単位とした含意関係の定義は容易ではない。そのため本研究では、ペア間係数をペア間の類似度としてモデルを設計する。

また、二つのペア p_i と p_j 間の類似度を、それぞれの属性語 a_i と a_j 間の類似度と、評価語 v_i と v_j 間の類似度の二つに分割して考えるという方針と、ペア単位で類似度を求める方針の、二つの方針でペアの類似度を考える。

ペア間の類似度を属性語同士、評価語同士の類似度に分割する場合、“音-大きい”と“動作音-うるさい”など、属性語、評価語それぞれが類似しているようなペアに対しては高い類似度が与えられると期待できるが、“排気-きれい”と“臭い-無い”といった、属性語同士も、評価語同士も類似していないが、内容的には類似していると判断できるペアに対して、低い類似度を与えてしまう。

定義のしやすさの観点から見ると、属性語と評価語とに分けて考える場合は、単語間の類似度と同じような枠組みで計算できることが利点として挙げられる。逆に、ペア単位で類似度を定義することは、容易ではない。

以下では、まずはペア単位での類似度について説明し、その後、属性語と評価語とに分割した類似度について述べる。

4.3.1 ペア単位のペア間係数

まず、ペア単位での単語の重複度を利用した類似度 sim_{cov} を定義する。高村らは、文中に含まれる単語の表層的な含意関係を文間係数として用いることで、精度の高い結果を得た。そこで、ペア間係数として、二つのペア p_i と p_j の表層的な重複度を用いる。属性語と評価語で構成されるペアを二語の文とみなし、その一致度を、ペア間係数とした。ここで、一致度とは属性語同士と評価語同士を比べ、一致している数を、2 で割った値である。

次に、各ペアが出現した製品の情報を利用した類似度 sim_{cooc1} を定義する。

ここでは、今回利用したレビューコーパス中で、要約対象の製品と同カテゴリに分類される全ての製品に関するレビュー文書を利用する。

一般に、同じ製品に対するレビュー文書では、同じ内容の評価がされやすいということが考えられる。これは、属

性-評価ペアについても同様に考えられ、もし二つのペアが表す内容に類似性があるならば、それらが同じ製品中で共起する可能性は、そうではないペア同士よりも多いことが予想できる。つまり、二つのペア p_i と p_j が、共通の製品で出現したという情報は、二つのペアの類似度を測る一つの指標となり得る。そこで、そのような考えの下、以下の類似度を定義する。

各ペアについて、そのペアが出現したレビュー文書が言及している製品を素性とするベクトルを作成する。そして、二つのペア間の類似度を、それらのベクトル間の余弦類似度とする。

ここで、利用したレビューコーパスは掃除機カテゴリに関する Amazon.co.jp^{*3}と価格.com^{*4}より収集されたレビュー文書集合である。合計で 565 製品に対して 1900 のレビュー文書が存在しており、総文数は 14101 文であった。これに 3 節で定義した抽出規則を適用すると、3692 ペア (異なりペア数 2188) が抽出された。3 節で示したように、出現回数が 3 回以上のペアのみを残すと 2483 ペア (異なりペア数 1075) となった。

4.3.2 属性語、評価語単位のペア間係数

ここでは、あるペア p_i を、それを構成している属性語 a_i と評価語 v_i に分けて考える。二つのペア p_i と p_j の類似度を考える時、ペア単位で直接考えるのではなく、そのペアが含む属性語 a_i, a_j 間の類似度と、評価語 v_i, v_j 間の類似度を別々に計算し、二つのスコアをマージすることによって、ペアの類似度とする。

この形式では、二つの類似度を用意した。レビューコーパスを用いたものと、それ以外の外部資源を利用したものである。

まず、類似度 sim_{cooc1} と同様に、レビューコーパスにおける共起情報を用いた類似度 sim_{cooc2} について説明する。ここでは、4.3.1 節における仮定、すなわち、同じ製品で共起したペアは類似しているという考えを、属性語、評価語それぞれに分けて実装を行った。

各属性語について、その属性語が抽出されたレビュー文書が言及している製品を素性としたベクトルを作成し、属性語同士の類似度を、それらのベクトル間の余弦類似度とした。評価語についても同様に計算を行った。

そして、二つのペアの類似度を、属性語同士の類似度と、評価語同士の類似度との積とした。

2483 ペアを構成する属性語、評価語の異なり語数は、それぞれ 143, 354 であった。このそれぞれについて、上記のベクトルを作成する。

次に、レビュー文書以外の外部資源を用いた類似度 sim_{google} について説明する。

GoogleNgram とは、Google が公開している、ウェブか

ら抽出した約 200 億文 (約 2550 億単語) の日本語データから作成した n-gram データである。ユニグラムから 7 グラムまでが公開されており、今回は 7gram を利用した。

本稿では、GoogleNgram の 7gram データ中から、属性語を含むデータを全て抽出し、そこからストップワードを除外したものを共起語とし、共起語ベクトルを作成する。そして、属性語同士の類似度を、それぞれの共起語ベクトル間の余弦類似度の値とする。評価語についても同様に類似度を計算し、それらの積を、ペア間の類似度とする。

5. 実験

提案モデルを評価するための実験を行う。実験は、ある製品についてのレビュー文書から、システムが抽出し選択した属性-評価ペアと、同じデータについて人手で作成された参照要約から、参照ペアとして人手で抽出された属性-評価ペアを比較することで、システムが選択したペアの有用性を評価する。

5.1 データセット

掃除機に対するレビューデータに対してアノテートを行った。データは、ある製品に対する、一つのレビュー文書クラスタで、クラスタ中に含まれるレビュー文書数は 27 文書、各レビュー文書中の総文数を数えると 303 文であった。

三人のアノテータに以下の手順で参照要約、参照ペア作成を依頼した。アノテータはまず、ある掃除機についてのレビュー文書クラスタから、その内容を簡潔かつ客観的に表した参照要約を作成する。続いて、作成した参照要約中から、属性-評価ペアを抽出し、参照ペアとして列挙する。これら参照ペアと、提案手法が出力したペア集合を比較し、その一致度を提案モデルの評価とする。

この時、例えば参照要約中に「吸引力が強い」という意見が多い」という文がある場合、「吸引力-強い」という参照ペアが抽出できる。しかし、レビューの本文中では、吸引力の強さに対して「吸引力が強力」、「吸う力が大きい」など表現の仕方に多様性がある。そこで、各参照ペアについて、レビュー本文中で同じ意味で言及されているペアも同時に探し、参照ペアと対応付けを行うことで、評価表現の多様性を吸収する。上記の例で言えば、「吸引力-強い」、「吸引力-強力」、「吸う力-大きい」というペア群に対し、いずれか一つのペアが選択できれば、そのペア群はすべて選択したものとみなす。また、アノテートの際には、ゼロ照応・文外照応を許しており、その際は照応を行ったことを表すタグを付与した。

今回作成したデータの特徴として、参照要約の長さや参照ペアの抽出数に明確な基準を与えていない。理由については 8.3 節において説明するが、今回は「その製品を購入するか迷っている人にとって有用で、複数製品について同

*3 <http://www.amazon.co.jp/>

*4 <http://kakaku.com/>

じものを参照する人にとって読みやすい長さ」という非常に曖昧な定義で要約長を指定した。

実験の際は、アノテータが作成した参照ペアの数を要約長制約として、システムにペアを抽出させる。システムが出力したペアの集合と、参照ペア群（参照ペアとそれに意味的に対応するレビュー文書中のペア集合）の集合とを比較し、システムが出力したペアのうちいくつが、参照ペア群と一致しているかを評価の対象とする。

6. 結果

アノテートされた一つの製品に対するレビュー文書を用いて実験を行った。三人のアノテータにより抽出された参照ペアは、それぞれ 11,13,19 ペアであった。各アノテータの参照ペアと、システムの出力ペアを比較する。この時、システムへ与える出力ペア数の制限 K は、それぞれのアノテータが作成した参照ペア数とした。表 1 に実験結果として、参照ペアとシステムの出力したペアの正解率の平均値を示す。

表 1 実験結果

類似度	sim_{cov}	sim_{cooc2}	sim_{cooc1}	sim_{google}
スコア	0.420	0.407	0.403	0.377

表 1 を見ると、表層的な語の重複度を類似度とした場合に最も良い結果となった。 sim_{cov} は、属性語と評価語同士の二語の一致度を用いた類似度のため、取りうる値は 0.0, 0.5, 1.0 の三つしかない。それにも関わらず設定した類似度の中で最も良くなった原因については考察する。

また、“同じ製品で共起した評価表現は類似度が高い”という仮定により定義した二つの類似度 sim_{cooc2} , sim_{cooc1} が、次いで高いスコアとなった。

7. 考察

類似度 sim_{cov} は、属性語あるいは評価語のどちらかが一致していれば類似度が上がり、逆に表層的に同じ語を共有していなければ類似度が得られないという、シンプルかつ極端な定義である。にも関わらず、今回設定した類似度の中で最も高い評価を得た理由として、掃除機に対するレビュー文書中の表現の多様性の低さが挙げられる。

掃除機の場合、排気、吸引力、音など、購入の際に焦点となる属性が限られており、実際にレビュー文書中に見られる評価表現もそのような表現に集中している。

主要な属性語の少なさから、評価のパリエーションもあまり多くはなく、例えば掃除機において“綺麗”と言った場合、ほぼ全てが“排気”に対する評価である。しかし、これが例えば iPad に対する“綺麗”であれば、“音質”、“画質”など、対応する複数の属性が考えられ、一意に予想することが難しい。

そのため、属性語も評価語も異なるが、同じ意味を表している属性-評価ペアの組がある程度混在するような場合に、提案手法の有効性が発揮されることが予想される。今後はより多くの製品カテゴリで実験を進めることで、提案手法の枠組みの有用性を検討していく。

また、同じ製品で共起した属性-評価ペアは似ている、という仮定を持って設定された類似度 sim_{cooc1} , sim_{cooc2} は、データ量の少なさからくるスパースさの問題で、大きな特徴を出すことが出来なかった。データ量を増やすことで改善されると期待できる例として“静音性-高い”と“騒音-小さい”の二つのペアが挙げられる。これらはどちらも出現回数は少ないながらも、同じ製品で共起したために類似度が高くなった例である。より大規模なデータを用いることで、このような異なる表現で同じ意味を表すペアの類似性を認識できる類似度の作成ができる。

8. まとめと今後の展望

本稿では、評判情報の要約手法として、文そのものではなく属性-評価ペアを単位として要約を行うモデルについて説明した。提案手法を考える上で重要な要素としては、ペアの抽出方法、ペアの選択モデル、ペア間の類似度の定義が挙げられるが、それぞれにおいて多くの課題が残されている。今後はそれぞれの課題を解決していくことで、より良い要約の生成を目指す。

考察で述べたように、設定したペアの類似度によって大きな特徴が観測できなかった理由として、利用したレビューコーパスの性質や規模の小ささが挙げられる。今後はより大規模なレビューデータを用いて、提案する要約手法の有用性を検討していく。また、特徴が現れなかった原因として、ペア抽出の再現率の低さも同時に挙げられる。今後は、品詞を限定した単純な係り受け関係ではなく、より多くの表現から属性-評価が抽出できるよう検討していく。

また、2.3.2 節で述べたような感情極性について考慮することは、ペアの類似度を考える上で重要である。そこで、今後は、評価の感情極性も考慮に入れたペアの類似度について検討を行う。

以降では、本研究が設定している最終的な目標について説明し、次にそのために注目すべき課題について述べる。

8.1 選択されたペアによる生成的要約への取り組み

本研究では、将来的には選択したペアを自然文として再構築するという abstractive な要約の生成を目指している。“吸引力-強い”、“排気-きれい”、“音-うるさい”、という三つのペアが選択され、類似したペアがレビュー文中に多く存在していた場合に、“吸引力は強いし、排気もきれいという意見が多いが、音はうるさく感じる人が多い”というような要約を作成する。これらのペアは、抽出元の文では、1 節で示した例のように冗長な文である可能性が高いが、そ

のような三文を、上記のように簡潔で、重要な情報のみを含んだ文として再構築することが出来れば、有用な要約の作成が可能となることが期待できる。

8.2 属性-評価ペアで表せない重要な文の扱い

ここで、レビューにおいて重要と考えられる評価表現の多くは属性-評価の形で表すことができると思われるが、中にはそうでない例も存在する。例えば、「この製品はコードレスではないのでご注意ください」という文は、どのような属性評価を用いても代替不可能な評価表現であるが、要約として含めることは有効であるように見える。そこで、上記の生成的な要約と同時に、“ペアの形で取れないが、要約として含めるべき重要な表現”を抽出する extractive な要約手法を組み合わせることで、より自然で良質な評判情報の要約を目指す。

ペアの形で抽出できない表現を抽出する要約手法に関しては、ペアが与えられた上で、新たな評価表現を選ぶということを考えると、update summarization^{*5} の枠組みが一つの参考になると考えられる。しかし、ペアにより代替可能な文や表現を抽出しないためにも、ペアの抽出に関して検討を行う必要がある。また、扱う単位の異なる二種類の要約手法を同時に考える場合、それぞれの手法が出力する要約の長さをどのように決定するかを考える必要がある。そのような統一された“要約の内容の充実度”の定義は難しい問題であり、今後十分に検討していく必要がある。

8.3 要約の長さ

またこのトピックに関連し、今後検討していくべきもう一つの課題として、要約の長さがある。具体的には、与えられた文書クラスタの内容の充実度に応じて出力される要約の長さを動的に変化させることができれば、より望ましい要約が生成できる。

評価データの作成時に、アノテータに参照要約の長さに明確な基準を与えなかったのもこのような理由からである。しかし、予め要約長を制約として指定した現状のモデルではそのような挙動は期待できない。今後は、内容の充実度と要約の長さのバランスを取った要約モデルを考えていく必要がある。

8.4 評価を表す述語項構造

また、抽出されたペアのなかには、“掃除-楽しい”、“ゴミ-捨てにくい”などといったペアも見られた。これらは、属性語と評価語という観点では正しいものではないが、評価表現としては、ペアの形で簡潔に分かる有用な情報であるといえる。これらのような、属性-評価ペアではないが、“評価表現を表す述語項構造”をどのように扱うかは今後の

焦点の一つとなりうる。

我々の選択肢としては、まず評価表現の単位を属性-評価ペアに限らず、より一般的な述語項構造まで拡張を行うことが考えられる。

もう一つは、属性-評価ペアを単位としたままで、今後の展望として挙げた、ペア選択と同時に抽出的手法の枠組みとして解くことである。この場合は、評価表現の抽出は、従来の意見抽出タスクと同様のアプローチで取り組むことになる。

参考文献

- [1] 高村大也, 奥村 学: 施設配置問題による文書要約のモデル化. 人工知能学会論文誌 (2010) .
- [2] Hitoshi Nishikawa, Takaaki Hasegawa, Yoshihiro Matsuo and Genichiro Kikui: Opinion summarization with integer linear programming formulation for sentence extraction and ordering. *COLING* (2010) .
- [3] Kazutaka Shimada and Ryosuke Tadano and Tsutomu Endo: Multi-aspects review summarization with objective information. *Procedia - Social and Behavioral Sciences* (2011) .
- [4] Hyun Duk Kim, Kavita Ganesan, Parikshit Sondhi and ChengXiang Zhai: Comprehensive Review Of Opinion Summarization. *Opinion Summarization Survey* (2011) .
- [5] Yue Lue, ChengXiang Zhai and Neel Sundaresan: Rated aspect summarization of short comments. *WWW* (2009) .
- [6] Minqing Hu and Bing Liu: Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD-2004) .
- [7] Minqing Hu and Bing Liu: Opinion Extraction and Summarization on the Web. *AAAI* (2006) .
- [8] 飯田龍, 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出を目的とした機械学習による属性-評価値対同定. *2004-NL-165* (2004) .
- [9] 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. *2006-NL-171* (2006) .
- [10] 工藤拓, 賀沢秀人: 「Web 日本語 N グラム第 1 版」. 言語資源協会発行 (2007) .

*5 <http://www.nist.gov/tac/2009/Summarization/update.summ.09.guidelines.html>