

# 属性-評価ペアを単位とした評判情報の要約

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

菊池 悠太\*<sup>1</sup>    高村 大也\*<sup>2</sup>    奥村 学\*<sup>2</sup>  
Yuta Kikuchi    Hiroya Takamura    Manabu Okumura

\*<sup>1</sup>東京工業大学 総合理工学研究科

Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

\*<sup>2</sup>東京工業大学 精密工学研究所

Precision and Intelligence Laboratory, Tokyo Institute of Technology

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.

## 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。人々の意見が見られる場所として、ブログを始めとした様々な媒体が考えられるが、本研究では、それらの中で、通信販売サイトにおけるレビュー文書に焦点を当てる。レビュー文書は、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。そのため、レビュー文書中に書かれた評判情報を分析し、要約するという技術は重要であり、近年そのニーズは高まっている。

本研究では、ある製品について書かれたレビュー文書クラスタの要約を生成する。具体的には、レビュー文書クラスタに含まれる全ての文から意見を表す評価表現を抽出し、それら全体の内容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。なお、評価表現として、属性-評価ペアを利用する。また、抽出したペア集合全体から、全体の内容を被覆する、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。ここで、文書クラスタとは、書き手の異なる一つ以上の文書を含んだ文書の集合である。

## 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。従来の文書要約と同様のアプローチで行われた研究には [Nishikawa 10], [Shimada 11] 等がある。いずれも、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。

連絡先: 菊池悠太, 東京工業大学精密工学研究所, 横浜市緑区長津田町 4259 R2-728, 045-924-5295, kikuchi@lr.pi.titech.ac.jp

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 [Kim 11] 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。いわば、データの構造化や可視化に焦点を当てた研究である。

## 3. 属性-評価ペアとその抽出規則

属性語とは、対象となる製品が評価される側面を表す語であり、評価語とは、その側面に対する具体的な評価を表す語である。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 [飯田 05]。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。また、従来研究では3つ組よりも更に複雑な構造化に焦点を当てた研究も存在するが [小林 06]、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、複雑な構造化や抽出手法については、今回は考えないことにする。

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、以下に定義した規則を用いて属性-評価ペアを抽出する。評価対象の属性は、一般的に名詞(句)であり、評価は主に形容詞によって表される。そこで、構文解析器の出力結果から、文節末が“が”あるいは係助詞である名詞節が、形容詞節に係っている構造に対し、係り元の語を属性語、係り先の語を評価語とし、それらを対にして属性-評価ペアとして抽出する。なお、構文解析器としては CaboCha を用いた。また、コーパス全体で頻度が高い属性-評価ペアのうち、ノイズであるものを人手で除去している。

## 4. 提案手法

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タから、3節における規則により抽出されたペア集合を入力として与え、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。

#### 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 [高村 10]。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、扱う単位が文ではなく属性-評価ペアである。それに伴い要約長制限が文字列長ではなく選択するペアの数となっている。

#### 4.2 提案モデル

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。

$$\text{maximize } \lambda \sum_{i,j} e_{ij} z_{ij} + (1 - \lambda) \sum_i f_i x_i$$

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

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

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

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

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

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

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を、 $e_{ij}$  で表し、これをペア間係数と呼ぶことにする。また、レビューコーパスにおけるペア  $p_i$  の出現頻度を  $f_i$  とする。よって、目的関数内の二つの項、 $\sum_{i,j} e_{ij} z_{ij}$  と  $\sum_i f_i x_i$  の第一項は、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となっている。第二項は、5.3 節で述べるベースラインと同等の意味合いを持つ。これらをトレードオフパラメータ  $\lambda$  でバランスを取ったもの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りモデル化する。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。これについては次節で、いくつかの類似度を定義する。

#### 4.3 ペア間係数

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、 $p_j$  がそれに被覆されるか否かの度合を、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。

ここで、留意したいいくつかの類似度の定義を示すが、その前に、類似度を作成するために利用したコーパスについて説明する。本研究では、二種類のレビューコーパスを用いる。一つ目は Amazon.co.jp と価格.com から収集したレビューコーパスである。これは、掃除機、デジタルカメラ、Mp3 プレイヤーに対し、合計 565 製品に対して 1900 のレビュー文書が存在している。評価に用いるデータセットは本コーパスから選んで作成している。二つ目は、より大規模なレビューコーパスとして、楽天技術研究所により公開された楽天市場のみんなのレビュー・口コミ情報 (以下、楽天コーパス)\*<sup>1</sup> を利用した。楽天市場内の約 5000 万商品に対する約 1660 万のレビュー文書が含まれており、このうち掃除機、デジタルカメラ、Mp3 プレイヤーに対するレビューのみを用いた。合計で 15788 製品に対して 64324 のレビュー文書が存在している。本コーパスは類似度の計算のために利用した。サイズの小さい一つ目のコーパスのみではデータがスパースになってしまうため、大規模な本コーパスによって解消することが狙いである。

- *cooc*

各ペアについて、出現した各レビュー文書に対応した製品番号を素性とした製品番号ベクトルを作成する。ある二つのペアの類似度は、それぞれの製品番号ベクトルの Jaccard 係数とする。同じ製品について書かれた複数のレビュー文書間では、同じ内容の評価を異なる表現で表されている可能性が高い。そこで、二つのペア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と考え、本類似度を作成した。

- *coocr*

本類似度は、基本的には *cooc* と同様である。*cooc* において製品番号ベクトルを構築したレビューコーパスは、Amazon.co.jp と価格.com のものであった。*coocr* は、これに大規模な楽天コーパス内のレビューデータによる情報を加えたものであり、より密な製品番号ベクトルが得られることが期待できる。

- *bow<sub>sent</sub>*

レビューコーパス内での分布類似度を用いる。具体的には、各ペアについて、そのペアが出現した全ての文内の出現単語を用いて文脈ベクトルを構築する。そして、二つのペアの類似度を、それぞれの文脈ベクトル間の余弦類似度とする。分布類似度は、意味的に似ている語句はその出現文脈も似ている傾向にある、という分布仮説 [Harris 54] と呼ばれる性質を利用している。

- *bow<sub>aep</sub>*

*bow<sub>sent</sub>* と同様に、分布類似度を用いた類似度である。*bow<sub>sent</sub>* とは異なる点として、出現単語を抽出する対象を、属性-評価ペアだけではなく、それを構成する属性語、評価語と共起した語も含めるといった点である。ただし、その重みはペア単位のものの半分に設定している。

- *bow<sub>google</sub>*

本類似度では、検索エンジンである Google.co.jp\*<sup>2</sup> を用いて、ペアの類似度を定義する。各属性-評価ペアについて、クエリとして検索をかけた検索結果のスニペット中に含まれる文を用いた分布類似度を用いる。なお、自然なクエリで検索を行うために、ペアを“が”という助詞でつなげる。

\*1 <http://rit.rakuten.co.jp/rdr/index.html>

\*2 <https://www.google.co.jp/>

以上五つの類似度を設定した。これらの類似度を用いて提案モデルを構築し、後の実験でその性能を評価する。

なお、本研究で提案するモデルは、大きく二種類に分けることができる。bow<sub>acp</sub>, bow<sub>sent</sub>, bow<sub>google</sub> など、文脈類似度を用いたモデルと、cooc, cooc<sub>c</sub> など、製品の共起情報に基づく類似度である。以降、前者を総称して bow モデル、対して後者を製品共起モデルと呼ぶ。

## 5. 実験

提案モデルを評価するための実験を行う。実験は、アノテートされた 25 製品のデータセットを用いて行った。

提案モデルは NP 困難な問題であるが、今回の実験設定における規模では、整数計画問題のソルバーである CPLEX を用いることで最適解の計算が可能である。

### 5.1 データセット

Amazon.co.jp、価格.com における製品レビューを収集し、評価のためのデータセットを構築した。掃除機、デジタルカメラ、Mp3 プレイヤーの三つの製品ドメインより構成され、各ドメイン毎の製品数は、掃除機 5 製品、デジタルカメラ、Mp3 プレイヤー各 10 製品である。レビュークラスタ内の文数は平均 154 文であった。収集した合計 25 製品の各レビュー文書クラスタに対して、三名のアノテータに以下の手順で参照要約、参照ペアを作成するように依頼した。アノテータはまず、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。その後、各参照ペアに対して、同じ意味を表すレビュー文書クラスタ中の属性-評価ペアを紐付ける。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だろう。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、このようなペアも同時に探し出し、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。その結果、上記の例で言えば、「吸引力-強い」、「吸引力-強力」、「吸う力-大きい」という三つのペアによる参照ペア群が出来上がる。このように作成された参照ペア群集合を用いて、提案手法が出力したペア集合について 5.2 節で述べる被覆率により評価を行う。また、アノテートの際に、ゼロ照応・文外照応を行なった際は、照応を行ったことを表すタグを付与した。本研究では、ペアの抽出においてゼロ照応・文外照応は考慮しておらず、今後の課題とする。

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点あげられる。具体的には「その製品を購入するか迷っている人にとって有用で、複数製品について比較する上で読みやすい長さ」という非常に曖昧な定義で要約の長さを指定した。これは、製品の特徴は様々なものが存在するため、一意に要約長を制限することが、必ずしも最適ではないと判断したためである。モデル中に、内容の充実度と要約の長さのトレードオフを取り込み、要約長をシステムに選択させることができれば、より有用な要約を作成することが可能になると期待できる。

ただし今回は、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、明示的に設定することとし、モデルにそのような機能を組み込むことは今後の課題とする。

### 5.2 評価方法

本研究の評価指標は、複数文書要約の評価指標として平尾らが提案した被覆率を用いる [平尾 07]。被覆率は、異なる表

現で同じ意味を表す文が複数存在するという、コーパスの冗長性を考慮した評価指標、及びアノテーションの枠組みである。

オリジナルの被覆率は文を単位として考えたが、ここではペアを単位とした被覆率を考える。また詳しくは割愛するが、オリジナルの被覆率では部分点が存在したが、ペアを単位とした被覆率では部分点を定義することが困難であるため、簡素化した評価式を用いる。具体的には、参照ペア群のうち、システムがカバーした参照ペア群の割合が、システムの被覆率である。アノテータ毎の参照ペア群との被覆率の平均値を、そのレビューデータにおけるシステムのスコアとし、評価を行う。なお、本評価は提案手法が内部で用いる類似度の違いによる代表ペアの選択についての評価であり、実応用における有用性の評価はおこなっていない。関連研究の多くは評価軸毎にレーティングを行うことで、粗いレベルでの製品の比較を行うものであり、本研究が目指す要約の形とは役割が異なっている。そのため、そのような関連研究と、本研究の比較評価の方法は今後検討していく必要がある。

### 5.3 ベースライン

実験にあたり用意したベースラインについて説明する。提案手法と同様に、要約対象となるペア集合を受け取り、コーパス全体における出現頻度を用いて代表ペアを選択する。つまり、ベースラインでは出現頻度が高いものから順に代表ペアを選択する。このようなベースラインを設定した理由として、コーパス全体で出現頻度が高いと言うことは、それだけ言及される可能性が高いという事が挙げられる。また、別の理由として、出現頻度順に並べたとき、全く同じ意味を表す属性-評価ペアが続く可能性は、そうでない場合よりも低いことが予想できる。ここで、掃除機において、吸引力に関する様々な表現が、他の動作音などのトピックとくらべて軒並み高頻度であった場合は、本ベースラインは常に吸引力に関するペアしか選択しないことになり、良い結果は得られないであろう。ただし、実際は動作音や排気、本体のサイズなど、種々のトピックに対して評価が分散することが考えられ、このようなベースラインでも一定の評価が得られることが期待できる。以降、本ベースラインを *freq* と呼ぶ。

### 5.4 実験結果

製品ドメイン毎の  $\lambda$  の変化による平均被覆率の変化を図 1, 2, 3 に示す。なお、 $\lambda = 0.0$  のときのモデルの出力は、ベースライン *freq* と同じである。各グラフを見ると、すべての製品ドメインにおいてベースラインが高スコアである事がわかる。

また、これら三つの図を見て分かる通り、どの製品ドメインにおいてもグラフが歪な曲線を描いている。図 3 において、製品共起モデルが bow モデルに比べて高いスコアを安定的に出しているなど、いくつかの特徴は見えて取れるが、全体的には値の上下が激しい、不自然な形をしている。このようなグラフになった理由として、製品によってデータの特徴が異なることが考えられる。図 4 は、掃除機における 5 つの製品について、 $\lambda$  による平均被覆率の変化をプロットしたグラフである。図から分かるように、製品によって描く曲線の傾向が大きく異なっている。これは、製品のレビュー文書から得られるデータによって、目的関数の第一項、第二項どちらの尺度がより影響するかが異なるために生じた現象である。図 1, 2, 3 は、このように個々の製品で異なる弧を描くグラフの平均をとってしまったために起きると考えられる。実際にある製品についてのレビュー文書クラスタの要約を生成する場合、 $\lambda$  の値はある値に定めなければならない。そのため、 $\lambda$  を自動的に決定する必要がある。しかし、前述の通り、製品ドメイン内でもその特徴の違いが大

きいため、この方法は上手くいかないことが予想される。

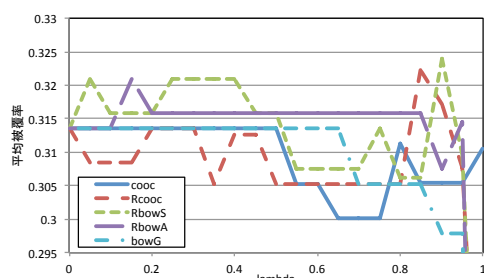


図 1: 掃除機

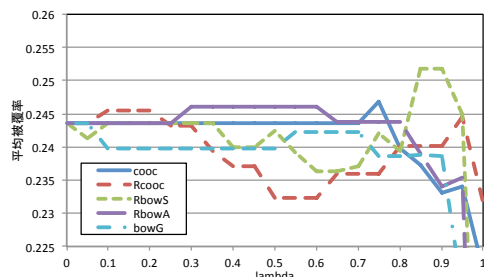


図 2: デジタルカメラ

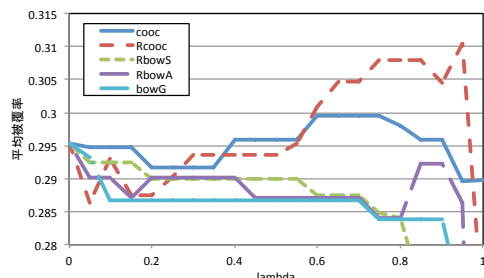


図 3: Mp3 プレイヤー

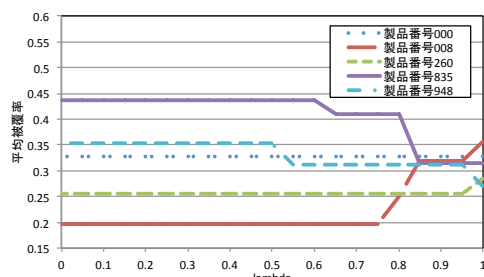


図 4: 掃除機五製品における平均被覆率の変化。

表 1: 実験結果

類似度	mp3	掃除機	カメラ	average
<i>cooc</i>	0.302	0.343	0.244	0.296
<i>cooc<sub>r</sub></i>	0.292	0.343	0.236	0.290
<i>bow<sub>sent</sub></i>	0.265	0.302	0.245	0.271
<i>bow<sub>aep</sub></i>	0.287	0.308	0.239	0.278
<i>bow<sub>google</sub></i>	0.278	0.259	0.234	0.257
<i>freq</i>	0.295	0.314	0.244	0.284

そこで、同じ製品ドメインの中でも、製品単位で似た製品を選別し、似た製品の  $\lambda$  の情報のみを用いることで、この問題に対処する。具体的には、要約対象の製品ドメインに属するすべての製品について、そのレビュー文書クラスから属性-評価ペア集合を抽出する。このペア集合は、その製品の特徴が現れていることが期待できる。ある製品の要約を生成するとき、

このペア集合同士の Jaccard 係数が最も高かった製品の  $\lambda$  のみを参照し、その中で最も高い数値であった  $\lambda$  で実験を行う。推定した  $\lambda$  を用いて実験を行なった際の、モデル毎の平均被覆率を表 1 に示す。表を見ると、二つのモデル *cooc*、*cooc<sub>r</sub>* において、ベースラインを上回る結果が得られた事がわかる。次節では、主にこの表について考察を行う。

## 6. 考察

表 1 で、類似製品のみを用いて  $\lambda$  を推定した結果、製品共起モデルがベースラインを上回る結果となった。また表を見ると、全体的に製品共起モデルが、bow モデルのスコアを上回っていることが分かる。このことから、本タスクにおいては形態素情報に基づいた分布類似度よりも、製品共起情報が有用であることが分かる。

実験結果全体を通して、単に頻度を基にしたベースライン *freq* が、全体的に高い値であることが分かる。大きな理由としては、5.3 節で述べた、本ベースラインを設定した二つ目の理由が、大きく関係している。すなわち、コーパス全体で出現頻度の高い順に属性-評価ペアを並べた場合、そこに同じ意味のペアばかりが連なることは少ないという直感である。

## 7. まとめ

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。モデルについては、トレードオフパラメータにより重み付けられた二つの項を持つ施設配置問題として定式化を行なった。また、提案モデルを評価するための新しいデータセットを作成し、抜粋型要約の一評価指標として提案されている被覆率を適用することで、モデルの評価を行った。

実験の結果、施設配置問題の枠組みにおいて、ペアの類似度として製品の共起情報を用いたものが最も良いスコアとなる事がわかった。コーパスの分析により、製品により重要となる目的関数の要素が異なることを示し、類似した製品のみを用いて目的関数のトレードオフパラメータ  $\lambda$  の選択を行なった。結論として、評価表現に多様性のある製品ドメインにおいて、本研究で提案する枠組みが有用に働くことを示した。今後は、ペアの再現率の向上や感情極性の利用などに取り組む。

## 参考文献

- [Harris 54] Harris, Z.: *Distributional Structure*, in *Word*, Vol. 10(23), pp. 146–162 (1954)
- [Kim 11] Kim, H. D., Ganesan, K., Sondhi, P., and Zhai, C.: *Comprehensive Review of Opinion Summarization* (2011)
- [Nishikawa 10] Nishikawa, H., Hasegawa, T., Matsuo, Y., and Kikui, ichiro G.: *Opinion Summarization with Integer Linear Programming Formulation for Sentence Extraction and Ordering*, in *COLING (Posters)*, pp. 910–918 (2010)
- [Shimada 11] Shimada, K., Tadano, R., and Endo, T.: *Multi-aspects review summarization with objective information*, *Procedia - Social and Behavioral Sciences*, Vol. 27, pp. 140–149 (2011)
- [高村 10] 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*, Vol. 25, No. 1, pp. 174–182 (2010)
- [小林 06] 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*, *Technical Report 2006-NL-171* (2006)
- [飯田 05] 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*, *Technical Report 2005-NL-165* (2005)
- [平尾 07] 平尾努, 奥村学, 福島孝博, 難波英嗣, 野畑周, 磯崎秀樹: *抜粋による複数文書要約を評価するためのコーパスと評価指標*, *情報処理学会論文誌データベース (TOD)*, Vol. 48, No. SIG14(TOD35), pp. 60–68 (2007)