

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

飯田龍[†] 小林のぞみ[†] 乾健太郎[†] 松本裕治[†]
 立石健二[‡] 福島俊一[‡]

本稿では、文章に記述されている意見を抽出するタスクについて述べる。我々は、意見を〈対象、属性、評価値〉の3つ組として定義し、文章からその3つ組を抽出する手法を提案する。具体的には、意見抽出の問題を、(a) 辞書に存在する属性候補集合と評価値候補集合から評価値候補と対となる属性を同定する問題と、(b) 同定した対が意見性を持つか否かを判定する問題の2つの問題に分解し、それぞれ機械学習に基づく手法を用いて解析することにより、属性-評価値対を同定する。提案手法を用いて評価実験を行った結果を報告するとともに、今後の展望についても議論する。

A Machine Learning-Based Method to Extract Attribute-Value Pairs for Opinion Mining

Ryu Iida[†] Nozomi Kobayashi[†] Kentaro Inui[†] Yuji Matsumoto[†]
 Kenji Tateishi[‡] Toshikazu Fukushima[‡]

This paper addresses the task of extracting opinions described in a given document collection. Assuming that an opinion can be represented as a tuple $\langle Subject, Attribute, Value \rangle$, we propose a computational method to extract such tuples from texts. In this method, the major task is decomposed into (a) the process of extracting *Attribute-Value* pairs from given texts and (b) the process of judging whether each extracted pair is expressed as an opinion of the author, to both of which we apply machine-learning techniques. We also report on the present results of our experiments discussing future directions.

1 はじめに

ウェブ上のレビュー記事や blog などの文章から評判情報を獲得する研究 [1, 7, 9, 11] が注目をあびている。これらの研究では、テキスト分類や情報抽出の技術を応用して実現されており、意見の記述されている範囲（例えば、文章や文、節など）の特定やその意見が肯定的に記述されているか、それとも否定的な記述かの判定（以後、p/n 判定と呼ぶ）などさまざまな問題を扱っている。

我々は立石らの研究 [12] に基づき、レーダーチャートを作成することで文章中の意見を要約するという立場で研究を進める。この枠組では、要約処理を以下に示す4つのサブタスクに分解して考えることができる。

- (1) 属性、属性値、主観評価となりうる表現の収集
- (2) テキスト中の属性と属性値（もしくは主観評価）

の対の抽出

- (3) 抽出した対の p/n 判定
- (4) 判定した p/n の値を用いた意見の要約

それぞれのサブタスクについて図1と対応付けて簡単に説明しよう。(1) まず、抽出すべき表現を限定するための辞書を構築する。抽出したい表現の多くはドメインに依存しており、その表現の辞書を作成することがまず必要となる。この処理では、我々が意見の構成素として定義した3種（属性、属性値、主観評価）の表現を獲得する。(2) 次に(1)で構築した辞書を用いて意見の抽出を行う。具体的には、図1の出力結果〈車1, 値段, 高い〉のように、対象、属性、評価値（属性値もしくは主観評価）の3つ組として意見を抽出する。(3) さらに、抽出した意見の p/n 値を判定する。例えば、意見〈車1, 値段, 高い〉が否定的であり、また意見〈車1, 内装, きれいだ〉が肯定的であると判定する。(4) 最後に、チャートの各軸に割り当てられた意見のうち肯定的な意見がどの程度存在するかを求めることでレーダーチャートを作成する。(1)の辞書獲得の内容は

[†] 奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology

[‡] NEC インターネットシステム研究所
Internet Systems Research Laboratories, NEC Corp.

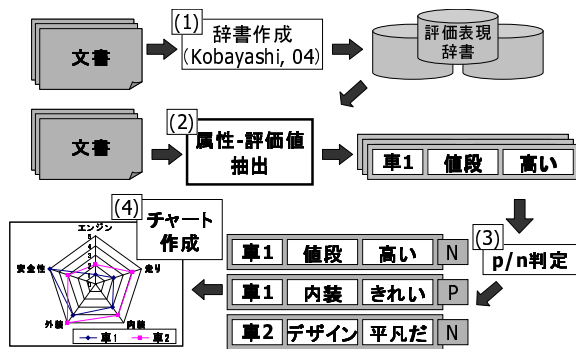


図 1: レーダーチャート作成の概観

文献 [4] でおおむね実現されており、(3) の p/n 判定は、辞書の表現を半自動で収集する際に p/n の値を手で付与するというわずかなコストで実現可能である。また、このほか、属性をチャートのどれかの軸に割り当てるとい作業が必要になるが、これはユーザがどのような属性をどの軸に加えたいか、またその粒度をどのように調節したいかに依存するため、自動的に行える類の作業ではない。これに対し、(2) の属性と評価値 (属性値もしくは主観評価) の対の抽出は、ドメインに依存した辞書さえあれば、ある程度機械的に処理できる。そこで、本稿では特に (2) のタスクに着目し、精度良く属性と評価値の対を抽出するための手法について述べる。

2 節では抽出対象となる意見をどのように形式化するかについてまとめ、3 節では、2 節で定義した意見を抽出する手法を提案する。次に、その提案手法を用いて対象ドメインが車とデジタルカメラの文章から、属性と評価値の対を抽出する評価実験を行った結果を 4 節にまとめ、その結果を考察する。5 節で関連研究との比較を行い、最後に 6 節でまとめる。

2 意見の定義

先行研究における意見の単位はさまざまなものがあるが、本研究では、文章内に記述された意見の要約を目的とするため、意見はチャートのある軸に対応付けられるような表現となっていることが望ましい。そこで、我々は意見を (対象, 属性, 評価値¹) の 3 つ組であると定義する。意見は属性を必ず含む。また、評価値は少なくとも属性値か主観評価のどちらか一方を必ず含む。属性を含むものだけを意見とすることにより、個々の意見をチャートのある軸へ対応付けることができる。

¹ 「車内は広くて落ち着く」の場合の「落ち着く」は「車内が広い」について的主観評価と考えられるが「落ち着いた内装」の場合の「落ち着く」は内装の取り得る属性値と考えられる。このように同一の表現でも、属性値と主観評価のどちらにもなり得るため、今回の実験では属性値と主観評価を区別せずとも評価値として対の抽出対象とする。

抽出対象となる属性, 評価値 (属性値, 主観評価) の特徴と例を示そう。まず、属性とはある対象 (商品) のある側面を表す表現を指す。対象が車の場合は「インテリア」や「エンジン」がそれに当たる。属性値は属性の値を指す表現であり、例えば、属性「室内」に対し「きれい」や「広い」などがそれに当たる。一方、主観評価は属性値に該当しない書き手の個人的な感想などを指す。例えば「気に入らない」「好きだ」「許せない」などの表現がある。

また、抽出対象を選択する際、どのような記述を意見とみなすかという意見性の規定も問題となる。例えば「彼は車 1 の内装がよいと言っていた」のような伝聞を意見とみなすか、また「デザインがもう少し落ち着いた感じならば」のような仮定表現を意見とみなすかといった判断は、抽出する意見情報の用途にも依存するので一般的には決められない。そこで、本研究では、ひとつの例として付録 A.4 に示す意見性の基準を設定し、どの程度意見の抽出が可能かを経験的に調査するというアプローチをとった。

3 属性と評価値の対の抽出手法

レビューサイトや blog のように意見抽出の対象となる文章ではどの対象 (商品) について記述されているかが自明である場合が多い。そこで、2 節で定義した意見を構成する 3 つ組のうち、本稿では特に、意見性を持った (属性, 評価値) の対を対象文章中から抽出することを目的とする。ただし、対の抽出には、以下に示す 2 つの問題が混在している。

- (1) (属性, 評価値) の対を同定する問題
- (2) 同定した対が意見性を持つか否かを判定する問題

この問題設定に対して、提案手法では、対象文章内に存在する評価値となりうる候補 (評価値候補) と対となる属性を同定し、同定した (属性, 評価値候補) の対を用いて評価値候補が意見性を持つか否かを判定する、という 2 段階の処理を行い、最終的に (属性, 評価値) の対を抽出する。

処理の流れを図 2 を用いて説明しよう。図 2 で意見の抽出対象となる文書には、属性候補 AC1, ..., AC6, また評価値候補 EC1, ..., EC3 が出現しており、最終的に抽出したい意見は (AC3, EC1) と (AC4, EC3) であるとする。まず最初に、図 2 (i) の候補抽出を行うことで各評価値候補に対して属性候補集合 $\{AC1, \dots, AC6\}$, $\{AC1, \dots, AC6\}$, $\{AC1, \dots, AC6\}, EC2$ を抽出する。次に、属性候補集合からそれぞれの評価値候補に対して適切な属性を選択し、対 (AC1, EC1), (AC3, EC1), (AC6, EC2), (AC4, EC3) を得る (図 2 (ii))。ここまでの処理で対の形で抽出はできているが、評価値

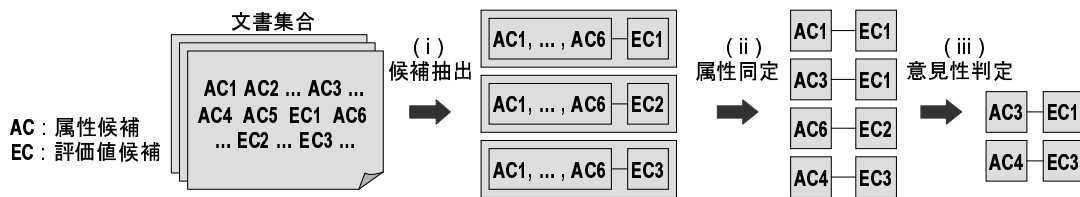


図 2: 属性と評価値の対の抽出

候補それぞれが 2 節で規定した意見性を持っているかどうかはわからない。そこで、それぞれの対に対して意見性を持つか否かの判定を行い、最終的に抽出したい意見である $\langle AC3, EC1 \rangle$ と $\langle AC4, EC3 \rangle$ を抽出する。

以下、個々の処理について、候補の抽出 (図 2 (i)) を 3.1 で、評価値候補と対となる属性の同定 (図 2 (ii)) を 3.2 で、評価値候補の意見性の判定 (図 2 (iii)) を 3.3 で説明する。

3.1 評価値候補と属性候補の抽出

属性辞書と評価値辞書 (つまり、属性値辞書と主観評価辞書) を用いて、まず対象文章内から評価値辞書の項目に該当する評価値候補を探し、その評価値候補に対してある範囲内に存在する属性候補を網羅的に抽出する。この集合を 3.2 の入力とする。

3.2 評価値候補と対となる属性の同定

次に、評価値候補と対となる属性の同定の問題を考える。評価値と属性は係り受けの関係にあるなど局所的に共起している場合もあるが、ゼロ照応などの現象のため、単純にパターンを用いて抽出するという方法では網羅的に対を抽出することが困難である²。

これに対して、本手法では、評価値候補とそれと対となる可能性のある属性候補集合が与えられた場合に、属性候補集合のそれぞれの要素と評価値候補が対となるか否かの 2 値分類問題を考える。この問題設定により、パターンベースの手法と比べて属性と評価値の統語的な制約が緩和されるため、より柔軟な組合せの可能性を検討することができる。以下で今回導入した 2 通りの解析手法を説明する。

3.2.1 Soon らのモデルを用いた属性の同定

属性と評価値の対という文脈を先行詞と照応詞の対と置き換えると、これまで提案されてきた照応解析のモデルをそのまま適用することが可能となる。さまざまな機械学習を用いた照応解析のモデルが存在するが、今回は最も単純な Soon ら [10] のモデルを適用する。

²今回作成したタグ付きコーパス中で属性と評価値の対が係り受けの関係として出現した割合は、車の場合では全体の 68.4%、デジカメの場合でも 61.1%であった。

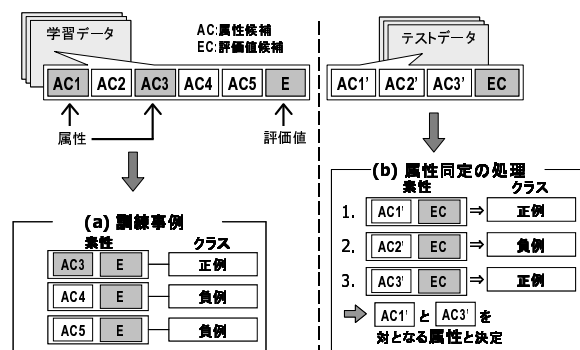


図 3: Soon らの解析モデルによる属性同定

このモデルでは、属性候補集合内のそれぞれの候補が評価値候補と対となるか否かの 2 値分類問題を解き、評価値候補に対して複数の属性を同定する。正解データの中には、一つの評価値候補に対して複数の対を持つ場合もあるが、この手法ではそのような事例に対しても網羅的に対を抽出できる可能性がある。

学習・分類のやり方は、Soon らの解析モデルにおいて、照応詞と先行詞をそれぞれ評価値、属性に置き換えて解釈したものにおおむね相当する。図 3 を用いて説明しよう。図 3 のタグ付きコーパスから抽出した候補集合には、評価値 E に対して、5 つの属性候補 $AC1, \dots, AC5$ が出現している状況を仮定している。AC1 と E, AC3 と E はそれぞれ正しい属性と評価値の対である。この状況で、分類器は各 AC_i ($i \in 1, \dots, 5$) が E と対となる属性か否かという 2 値分類問題を解く。

訓練時には、図 3 (a) のように、評価値と (評価値から最も近い) 属性の対 $\langle AC3, E \rangle$ を正例、評価値と他の (評価値と属性の間の) 各属性候補の対 $\langle AC4, E \rangle$ および $\langle AC5, E \rangle$ を負例として学習する。新しい対の同定の問題を解く際には、訓練時と同様に、属性候補の各々について、それが属性か否かを分類する。そして、分類器が属性と決定したすべての属性候補と評価値候補の対を出力する。分類器が、各属性候補のどれも属性ではないと分類した場合は、対象とする評価値候補は対となる属性を持たないと判断する。図 3 (b) の例では、評価値候補 EC とそれぞれの属性候補との組合せ、 $\langle AC1', EC \rangle$, $\langle AC2', EC \rangle$, $\langle AC3', EC \rangle$ を分類し、その結果 $\langle AC1'$,

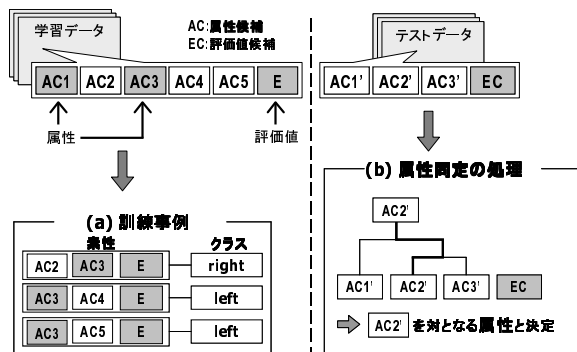


図 4: トーナメントモデルによる属性同定

EC) と $\langle AC3', EC \rangle$ をそれぞれ対として出力している。この 2 つの対が 3.3 への入力となる。

3.2.2 トーナメントモデルを用いた属性の同定

3.2.1 で述べた Soon らの手法では、属性と評価値に関する可能な組合せをすべて出力するため、誤った対も出力する可能性がある。そこで、再現率は犠牲になるが評価値候補に対して適切な属性の一つを選択するという問題設定を考える。

評価値候補に対して最尤の属性を選択する問題には、文献 [2] で提案したトーナメントモデルが適用できる。この手法では、複数存在する属性候補に対して候補間で比較を行い、勝ち抜き方式で対となる属性を決定する。

図 3 の状況を再び図 4 に描く。トーナメントモデルでは、正しい評価値 E と対となる属性は他の属性候補に対して勝ち残る必要がある。そのため、その関係を学習するために、図 4 (a) に示す 3 つの訓練事例を抽出する。クラス right (left) は与えられた属性候補のうち、右 (左) 側の候補が勝ちである (より対となる属性らしい) ことを示している。

トーナメントモデルで解析を行う際には、対象範囲内に存在する属性候補 $AC1', \dots, AC3'$ の間で勝ち抜き戦を行い、最終的に勝ち残った属性候補 $AC2'$ を評価値候補 EC と対となる属性と決定し、 $\langle AC2', EC \rangle$ の対を出力する (図 4 (b))。

3.3 評価値候補の意見性の判定

3.2 で述べた処理だけでは、2 節で定義した意見性の基準を満たさない評価値候補も抽出する可能性がある。そこで、3.2 の出力である \langle 属性, 評価値 \rangle の対が意見性を持つか否かを判定する。

今回定義した意見性を捉えるための手がかりの一つとして表層的な表現があり「~と思う」と記述されている場合は、主観的な記述である場合が多く、また「~と言う人もいる」という表現の場合は他人の意見と判断することができる。そこで、まずは単純に対象となる文節の前後文脈、対となる属性の表現のみを用いて意見性を持つか否かを判定するモデ

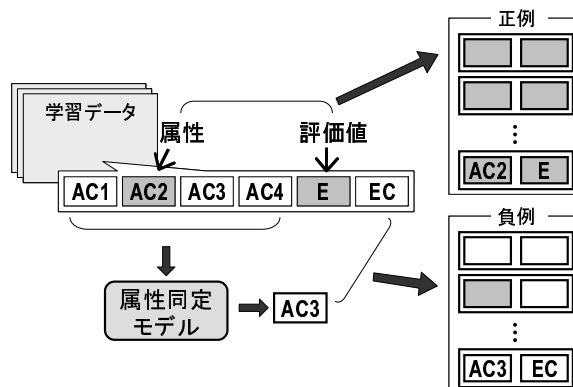


図 5: 意見性判定モデル

ルの作成を考える。

訓練事例の抽出を図 5 を用いて説明しよう。図 5 では、評価値 E に対して 4 つの属性候補 $AC1, \dots, AC4$ が対象範囲に出現している状況を仮定している。E と対となる属性は AC2 である。この状況で、分類器は評価値候補が意見性を持つか否かの 2 値分類問題を解く。訓練時には、評価値と属性の対 $\langle AC2, E \rangle$ を正例とする。また、3.2 で示したモデル (Soon らのモデルもしくはトーナメントモデル) を用いて意見性を持たない評価値候補と対となる属性候補を求め、その対を負例とする。図 5 の例で EC が意見性を持たない評価値候補だったとすると、EC に対して 3.2 の属性同定モデルを用いて対となる属性 AC3 を決定し、その対 $\langle AC3, EC \rangle$ を負例として訓練事例に追加する。このように訓練事例を作成することで、3.2 の段階で評価値に対して適切に属性を決定できた場合はその対の関係を最終的に出力し、また意見性を持たない評価値候補については対として同定した属性候補が不適切な組合せとなり対を棄却できる可能性がある。

4 評価実験

提案手法の有効性を検証するために、属性と評価値の対の抽出の評価実験を行った。

4.1 意見タグ付きコーパスの作成

評価実験のためにレビューサイトから車とデジタルカメラ (以下、デジカメ) に関する記事を抽出し、それらの文章に対し付録 A に示す基準で属性、属性値、主観評価にタグを付与し、意見タグ付きコーパスを作成した。作成したコーパスの規模を表 1 にまとめる。

作成したコーパス内の属性と評価値がどの程度離れているかを調査したところ、表 2 に示すように評価値に対して対となる属性の多くは同一文内もしくは評価値の前文に出現していたため、今回の実験では評価値の同一文内もしくは前文を対象に属性候

表 1: 意見タグ付きコーパスの規模

	車	デジカメ
文章数	287	835
文数	4442	10707
属性-評価値の対	2649	3654

表 2: 属性-評価値の出現傾向

	車	デジカメ
同一文内 (a)	2530	3299
1 文前 (b)	99	244
(a)+(b)	2629	3543
割合	99.2% (2629/2649)	97.0% (3543/3654)

補を抽出をする。

4.2 実験方法

前処理として茶筌 [6] と CaboCha [5] を用いて形態素解析と係り受け解析を行う。学習器は SVM [13] を用いた。多項 2 次カーネルを使用し、10 分割交差検定を行った。

実験では以下に示す 3 つのモデルを比較した。

1. ベースラインモデル (4.2.3 参照)
2. Soon らのモデル + 意見性判定
3. トーナメントモデル + 意見性判定

4.2.1 辞書の作成とその適用

今回の実験では、属性と評価値の対の抽出を他のタスクと独立に評価するために属性、属性値、主観評価の表現は、4.1 で作成したコーパスにタグ付与された表現を抽出し、それぞれの辞書項目とした。

辞書項目抽出の際は、コーパス内の文を形態素解析、係り受け解析したのち、それぞれのタグが付与された箇所を文節単位で抜き出した。このままでは接辞など余計な部分を含んでおり、辞書項目と入力の文の照合の際のノイズとなるため、抽出したもののうち最右の内容語までを辞書項目とした。実際に抽出した辞書項目の例を以下に示す。

- 属性: ボディ (名詞-一般), デザイン (名詞-サ変接続), シャッター (名詞-一般)
- 属性値: 軽い (形容詞-自立), 時間 (名詞-副詞可能) が (助詞-格助詞-一般) かかる (動詞-自立)
- 主観評価: すばらしい (形容詞-自立), 文句 (名詞-一般) なし (名詞-一般)

コーパスからの辞書項目の抽出は、訓練事例を n 分割交差検定する際、テストデータを除いた $(n-1)/n$ から抽出する場合とすべての訓練事例である n/n から抽出する場合に分けることができる。前者の場合は辞書を別に用意しているオープンな状況を評価をすることができ、後者の場合では対抽出のタスクを他と独立に評価することができる。

また、どの属性がどの評価値と共起していたかという情報 (以後、共起用例と呼ぶ) も辞書項目と同

様にあらかじめ抽出することができる。そこで、以下に示すような共起用例も抽出した³。

- インテリア / センス-が-いい
- オーディオ / 不満
- コストパフォーマンス / 飛び抜ける-て-いい

4.1 に示すタグ付きコーパスから抽出した総項目数は、車では属性辞書が 969、属性値辞書が 625、主観評価辞書が 702 であり、デジカメでは属性辞書が 1112、属性値辞書が 907、主観評価辞書が 930 であった。総用例数はそれぞれ 2131 と 2954 であった。

4.2.2 素性

評価値候補と対となる属性の同定、意見性判定の 2 つの処理では、どのような表現で属性 (もしくは評価値) が記述されているか、また属性と評価値がどのような位置関係で出現しているかの情報を参照する必要がある。そこで、素性には対象とする文節の表層文字列、品詞、属性候補と評価値候補が係り受け関係にあるか否か、属性候補と評価値候補の距離情報を導入した。さらに、4.2.1 で示した共起用例の情報も対の抽出に有効だと考え、属性候補と評価値候補の対が共起用例集合に含まれるか否かも素性として用いた。

4.2.3 ベースラインモデル

提案手法と比較を行うためのベースラインモデルを作成した。この単純なモデルでは、4.2.1 で示した共起用例を使用し、抽出した評価値候補と属性候補集合のうち共起用例と同じ組合せである場合にその対を正解として出力する。

4.3 実験結果

3 つのモデルの実験結果を表 3 に示す。評価を行う際は 4.2.1 で抽出した辞書を用いて候補の抽出を行うが、辞書の適用誤りや同一文節内に複数の評価値が存在するなどの原因により、解析の対象となる候補は実際にタグ付与された候補数より少なくなる。そのため、システムの解析結果から抽出した候補中に含まれる正解を分母として求めた再現率を再現率 a 、タグ付きコーパスにタグ付与された総数を分母として計算した再現率を再現率 b として 2 種類の結果を示した。

結果より、ベースラインモデルは、共起用例が十分にある場合は網羅的に正解を抽出できているが、9/10 の用例を用いた場合は再現率、精度ともに低いことがわかる。これに対して、学習を用いた 2 種類のモデルは 9/10 の場合には再現率は低いが、精度が約 70% 前後である。チャート作成には再現率

³ $(n-1)/n$ と n/n の 2 通りの抽出を行った。

表 4: 抽出誤りの原因

	車	デジカメ
属性と評価値の対の同定誤り	18	9
商品全体の評価	19	7
条件や仮定の表現	10	4
上位の属性を同定	8	1
書き手特有の現象	3	7
他の商品の評価	0	13
事実を述べている	0	10
タグ付与の揺れ	18	22
その他	30	35
合計	106	108

より抽出された結果の正しさを重視するためこの結果は好ましいものといえる。

次に、2つの提案手法の間で再現率、精度を比較する。Soonらの手法では、評価値候補に対して属性候補を複数出力するため、トーナメントモデルを用いた場合より再現率が高くなっている。逆に、より正確に属性候補を抽出できると考えられるトーナメントモデルを用いたモデルの方が抽出精度が良い。

また、全体的に車の記事を用いた場合よりデジカメの記事を用いた場合が再現率、精度ともに悪くなっている。この2つの記事の内容を比較することは困難だが、少なくともいくつかの記述の特徴の違いがある。車に関する記事の場合は、対象となる商品についての言及がほとんどであるのに対して、デジカメの記事では「商品1に対して、商品2の方がグリップがよい」のように他商品との比較が多かった。今回導入した素性の多くは bag-of-words 素性であったため、比較の構造的な特徴を捉えることができていない。今後は、そのような表現の構造を捉えるための素性について検討する必要がある。

4.4 誤り分析

表3のトーナメントモデル+意見性判定の実験結果のうち、9/10の辞書を用いて得られた精度が未知データに対する精度となり、チャート作成への応用を考えるとこの数値の向上を目指す必要がある。そこで、この誤り事例のうち、車、デジカメの2つのドメインで100件ずつ誤り分析を行い、誤りの主な原因を調査した。調査結果を表4にまとめる。

いくつかの主要な原因を示そう。まず〈属性と評価値の対の同定誤り〉については、「荷物_(属性)が広い_(評価値)」のように属性「荷物」が評価値「広い」の選択制限を満さない場合でも、誤って対として抽出してしまっている。また〈上位の属性を同定〉に関しては、「音_(属性)の悪過ぎる_(評価値) サウンドシステム_(属性候補)」で正解となる対は「音-悪過ぎる」であるが、「音」の上位属性となる「サウンドシステム」と「悪過ぎる」の対を誤って出力してしまった。こ

の2種類の誤りを改善するには、質の良い選択制限知識の導入やドメインに特化したオントロジが必要となり、特にドメイン依存の知識獲得は今後のおおきな課題と考えられる。

「内装がきれいならば」のように条件や仮定表現など手がかりとなる表現を含む事例でも、意見性を持つと誤判定したことがあった。これに関しては誤った事例をさらに調査し、追加すべき素性を検討したい。

また、表4によるとタグ付与の揺れが多い。ある名詞句が属性となるか、また、意見性の判断など最終的な決定は作業者の主観に依存する部分が多いため、付録Aに示す仕様の検討も今後の課題である。

5 関連研究

評判分析や評判検索の分野では、おおきく分類と抽出の2つの方向で研究が進んできた。

分類の立場から進められている研究 [1, 7, 9, 11] では、文や文章を単位に p/n 判定を行う。一例をあげると、Turney[11]らの研究では、肯定、否定を表す語と対象文章に含まれる形容詞句、副詞句との共起から句の p/n 値を求め、文章中に出現する p/n 値を持つ句の値の平均をとることで文章を p/n に分類する。

これに対して抽出の観点からは、文章全体を分類対象とせず、文章内から意見を表している箇所を抽出するという立場で研究がなされている。Nasukawaら [8] はあるトピック(商品名、会社名、属性名など)が与えられた場合に、そのトピックの意見を抽出し、それらの p/n 値を判定する。Nasukawaらの枠組の中で進められている研究の一つに Kanayamaらの研究 [3] があり、これが我々の目的と最も近い。彼らはトランスファ方式の機械翻訳の考え方を属性と評価値の対の抽出に導入しており、述語項構造を抽出の単位とし、翻訳パタンと語彙の辞書を評価抽出用のパタン集合と評価語彙に置き換え、安価なコストで対の抽出を実現している。

これに対して、提案手法では、評価値候補に対して属性を同定し、評価値候補の意見性を判定している。これは、本質的には Kanayama らの枠組と同様に、意見に関係する述語項構造の抽出を問題とする。ただし、述語項構造を頑健に解析することは困難であるため、特に意見性を持った〈属性, 評価値〉という限られた構造にのみ焦点を当て、機械学習の技術を導入することでどの程度の意見が抽出可能かを調査する試みである。

表 3: 実験結果

手法			車		デジカメ	
ベースライン モデル	10/10 の辞書	精度	74.2%	(2302/3102)	59.7%	(3110/5209)
		再現率 <i>a</i>	93.2%	(2302/2470)	93.0%	(3110/3344)
		再現率 <i>b</i>	86.9%	(2302/2649)	85.1%	(3110/3654)
	9/10 の辞書	精度	43.2%	(493/1142)	26.2%	(608/2324)
Soon らの モデル+ 意見性判定	10/10 の辞書	精度	88.1%	(2183/2479)	84.7%	(2775/3276)
		再現率 <i>a</i>	88.4%	(2183/2470)	83.0%	(2775/3344)
		再現率 <i>b</i>	82.4%	(2183/2649)	75.9%	(2775/3654)
	9/10 の辞書	精度	74.8%	(1079/1442)	67.5%	(1329/1968)
トーナメント モデル+ 意見性判定	10/10 の辞書	精度	92.1%	(2226/2416)	87.7%	(2814/3207)
		再現率 <i>a</i>	90.1%	(2226/2470)	84.2%	(2814/3344)
		再現率 <i>b</i>	84.0%	(2226/2649)	77.0%	(2814/3654)
	9/10 の辞書	精度	78.3%	(1042/1331)	70.2%	(1328/1893)
		再現率 <i>a</i>	57.1%	(1042/1825)	49.4%	(1328/2690)
		再現率 <i>b</i>	39.3%	(1042/2649)	36.3%	(1328/3654)

6 おわりに

本論文では、機械学習を用いて対象の属性表現と対となる評価値を抽出する手法を提案した。この手法では、属性表現と評価値の対の同定と評価値の意見性判定を独立なものと考え、評価値候補と対となる属性を同定したのち、その対を用いて評価値候補の意見性を判定する。提案手法を用いて車とデジカメの記事を対象に対の抽出実験を行い、それぞれ 78.3%と 70.2%の抽出精度を得た。

今回は、一例として書き手個人の主観的な記述のみが意見性を持つという基準を設定し、どの程度意見の抽出が可能かを調査した。ただし、抽出する意見情報は用途に依存して変わるため、別の意見性の基準を設定した場合の提案手法の有効性を調査する必要がある。

また、今後さらなる精度向上のためには、ドメインに依存した良質の知識を獲得し、本稿で対象とする属性と評価値の対の抽出に用いる必要がある。そうした知識獲得の方法を検討するとともに、既存の辞書獲得手法 [4] で構築された辞書の利用や立石ら [12] の意見要約の枠組に提案手法を導入し評価を行いたい。

参考文献

- [1] Dave, K., Lawrence, S. and Pennock, D. M.: Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews, *Proceedings of the 12th International World Wide Web Conference (WWW2003)*, pp. 519–528 (2003).
- [2] 飯田龍, 乾健太郎, 松本裕治: 文脈の手がかりを考慮した機械学習による日本語ゼロ代名詞の先行詞同定, 情報処理学会論文誌, Vol. 45, No. 3, pp. 906–918 (2004).
- [3] Kanayama, H. and Nasukawa, T.: Deeper Sentiment Analysis Using Machine Translation Technology, *Proceedings of the International Conference on Computational Linguistics (COLING)*, pp. 494–500 (2004).
- [4] Kobayashi, N., Inui, K., Matsumoto, Y., Tateishi, K. and Fukushima, T.: Collecting Evaluative Expressions for Opinion Extraction, *Proceedings of the 1st International Joint Conference on Natural Language Processing (IJCNLP)*, pp. 584–589 (2004).
- [5] 工藤拓, 松本裕治: Support Vector Machine を用いた Chunk 同定, 自然言語処理, Vol. 9, No. 5, pp. 3–21 (2002).
- [6] 松本裕治, 北内啓, 平野善隆, 松田寛, 高岡一馬, 浅原正幸: 形態素解析システム『茶釜』 version 2.2.9 使用説明書, 奈良先端科学技術大学院大学 (2002).
- [7] Mullen, T. and Collier, N.: Sentiment Analysis using Support Vector Machines with Diverse Information Sources, *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 412–418 (2004).
- [8] Nasukawa, T. and Yi, J.: Sentiment Analysis: Capturing Favorability Using Natural Language Processing, *Proceedings of the second International Conferences on Knowledge Capture (K-CAP)*, pp. 70–77 (2003).
- [9] Pang, B. and Lee, L.: A Sentiment Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, *Proceedings of the Conference on Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 271–278 (2004).
- [10] Soon, W. M., Ng, H. T. and Lim, D. C. Y.: A Machine Learning Approach to Coreference Res-

olution of Noun Phrases, *Computational Linguistics*, Vol. 27, No. 4, pp. 521–544 (2001).

- [11] Tarney, P. D.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 417–424 (2002).
- [12] 立石健二, 福島俊一, 小林のぞみ, 高橋哲朗, 藤田篤, 乾健太郎, 松本裕治: Web 文書集合からの意見情報抽出と着眼点に基づく要約生成, *情報処理学会研究報告 NL-163*, pp. 1–8 (2004).
- [13] Vapnik, V. N.: *Statistical Learning Theory*, Adaptive and Learning Systems for Signal Processing Communications, and control, John Wiley & Sons (1998).

A 意見タグ付与の基準

意見タグ付与の基準について, 車に関する例文を用いて説明する. 例文中の太字が実際にタグが付与される部分を示し, 下線部はタグを付与しない部分を示す.

A.1 タグの種類

付与するタグは, 対象, 属性, 属性値, 主観評価の 4 種類である. 今回対象とするコーパスでは, 対象はある商品の名称を指し, 属性は商品のある側面を表す. また, 属性値は属性の値であり, 主観評価は対象, 属性, 属性値の 3 つ組 (もしくは属性値の欠落した 2 つ組) についての書き手の意見を表す. 一つの意見は, 対象, 属性, 評価値の 3 つ組で構成されると定義する. 評価値は属性値と主観評価の 2 つで構成されるが, どちらか一つが欠落していてもよいものとする. したがって「商品 A の値段が高い (属性値)」や「商品 B のエンジンが気に入らない (主観評価)」は意見とみなす.

以下に, 車の記事におけるそれぞれのタグの例を示す.

対象: フォルクスワーゲンポロ, メルセデス・ベンツ A クラス

属性: コストパフォーマンス, ボディーサイズ, エンジン, 足回り

属性値: 音が静か, 値段が高い, 足元が少し狭い, ロードノイズが少し大きい

主観評価: 音が静かなのが気に入らない, 直進安定性が素晴らしい, 満足のいくスタイル, 圧倒的なボディ剛性

A.2 タグの範囲

タグは基本的に文節に付与する. 実際に作業をする際は, CaboCha[5] を用いて文節区切りにした結果が見える状態で作業を行う. 文節は以下のように ”/” で区切られている.

..... とにかく / スムーズな / 走行性能。
エンジンは / 軽自動車とは / 思えない / ほど / トルクフルで、 / ...

但し, A.2.1 から A.2.4 のように文節内の表現だけでは属性値・主観評価とみなせない場合は文節を越えてタグを付与する.

A.2.1 主観評価, 属性値

使い物に/ならない, 奥が/深い, 一味/違う

A.2.2 ある, ない, 思う などの表現

ブレーキに/不安が/ある, ストレスを/感じる

A.2.3 述語のみでは評価とみなせない場合

静かに/回る, そこそこ/走ってくれる

A.2.4 否定表現が文節を越える場合

よろしく/ありません, 最高とは/言えません, 広く/ない

A.3 タグ付与の手順

以下の手順でタグを付与する.

1. テキスト中の属性にタグを付与する
2. その属性と対応する対象, 属性値, 主観評価にタグ付与する
3. タグ付与の対象となる属性が無くなるまで, 1, 2 の作業を繰り返す

ただし, 属性の出現しない評価値にはタグを付与しない.

A.4 意見性の基準

今回導入する基準では, 記事の書き手の主観的な意見のみが意見性を持つと設定する. そのため, 以下に示すような条件, 仮定表現や, 書き手以外の意見については, 意見性が無いと判断しタグ付与の対象外とする.

書き手以外の意見はタグ付与しない.

エンジンが / 良いという / 人も / いる。

条件, 場所, 状況は属性に含めない.

直線での / 走り, FF としては / 回頭性,
後席の / 膝元空間, スタート時の / 加速,
4000 回転を / 超えてからの / フィーリング

条件や仮定を表す部分に記述された属性はタグ付与しない.

エンジンが / 良ければ、 / 買いた

A.5 タグ付与の注意点

A.5.1 属性

属性が階層で書かれている場合, 最下層の属性にのみタグを付与する.

タイヤの / グリップ感, デザインが / 良い / 車体, ドア-ロックの / 音, ホールド感も / ある / シートの / 出来, シートなどの / 絶妙な / バランス

車の種類は属性としない.

魅力的な / スポーツセダン, 1 ランク上の / 国産車, 世界最高水準の / ステーションワゴン

A.5.2 主観評価, 属性値

程度を表す表現にはタグ付与しない.

異常に / 軽かった, 超 / 安定している, 広く / 重宝している

A.5.3 その他

属性と属性値が一文節に出現する場合は両方のタグを付与する.

高品質, 低燃費

「~さ」や「~性」のように属性とも属性値とも考えることができる表現は属性とみなす.

ハンドリングの / 良さ, エンジンの / 静粛性

反実仮想の場合はタグを付与しない.

これで / アイドリング時の / エンジン音が / 静かだったら / 問題なしなんだけどな!