照応解析手法を利用した属性–評価値対および意見性情報の抽出

小林のぞみ　　飯田龍　　乾健太郎　　松本裕治
奈良先端科学技術大学院大学 情報科学研究科
{nozomi-k,ryu-i,inui,matsu}@is.naist.jp

1 はじめに

近年，Web 上の掲示板やレビュー記事といったテキストデータから意見や評判などの情報を抽出する研究が注目をあびている．これらの研究は，(a) 文や文章単位で肯定的か否定的かを分類する文書分類的アプローチ [1, 6, 8, 11, 13] と (b) 文章内の意見に該当する箇所を抽出する情報抽出的アプローチ [4, 7, 10] に分けられる．

我々の目的は，特定の商品やサービスに対する評価をその根拠とともに抽出することである．そのため，問題を後者の情報抽出の立場で考える．以下，このタスクを意見抽出と呼ぶ．

評価と根拠はさまざまな形でテキスト中に現れる．その主要なものの一つに属性と属性値の対がある．属性とはある対象 (商品) のある側面を表す表現を指し，属性値はその属性の値を表す表現を指す．我々はこの対に対象，評価を合わせた 〈 対象，属性，属性値，評価 〉 の 4 つ組で意見を抽出するタスクを考える．意見抽出の例を図 1 に示す．例えば「フィットのキビキビした走行性能にもすごく満足していた」から 〈 フィット，走行性能，キビキビ，満足 〉 という組を抽出する．ただし実際には属性値と評価の区別は困難であるため[1]，本研究では評価と属性値を合わせて評価値とし，〈 対象，属性，評価値 〉 の 3 つ組の抽出問題を解く．

このように問題を設定すると，意見抽出は属性と評価値を抽出する問題と抽出した情報が意見性を持つか否かを判定する問題に分けることができる．意見性とはどのような評価値を意見と見なすかの基準であり，要望や評価を意見とみなすなどさまざまな基準が考えられる．属性-評価値対を抽出する問題は，3 節で詳しく述べるように，属性と評価値が依存関係にない場合や属性が明示的に記述されない場合も扱う必要があり，問題の構造が照応解析と似ている．そこで本稿では，両者の類似点・相違点を論じ，照応解析手法を適用することで実際にどの問題が解決し，どの問題が残るかを経験的に明らかにする．

Extracting Attribute-Value Pairs and its Opinionicity Using an Approach to Anaphoric Resolution
Nozomi Kobayashi, Ryu Iida, Kentaro Inui and Yuji Matsumoto
Nara Institute of Science and Technology

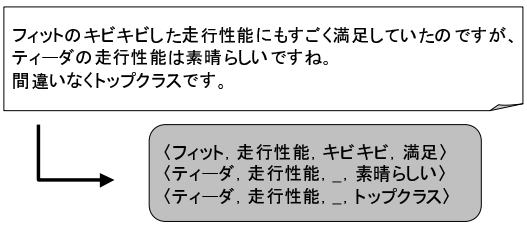

図 1: 意見抽出の例

2 先行研究

意見抽出の先行研究として，辞書と抽出パタンを用いる方法 [7, 10]，項構造解析による方法 [4] などがある．

村野ら [7]，立石ら [10] は抽出パタンと抽出対象となる評価を表す表現の辞書を用いて意見を抽出している．彼らは抽出パタンと辞書を人手で作成しているが，従来の情報抽出の問題と同様，抽出パタンの作成コストと網羅性，抽出精度の問題が残る．

一方，Kanayama[4] 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，比較的安価なコストで対の抽出を実現している．

しかし照応，省略などの問題により，述語項構造を頑健に解析することは困難である．後述する意見タグ付きコーパスで調査したところ，属性と評価値が文節係り受け関係にある対は 43%しかなかった．係り受け関係にない対 50 件を調査したところ，最も多かったのは「〈 振動 〉$_a$ や 〈 騒音 〉$_a$ が 〈 少ない 〉$_v$ 」[2]のように属性が並列になっている場合であり (約 3 割)，次いで「〈 パワー 〉$_a$〈 不足 〉$_v$ 」など，一文節に属性と評価が現れる場合 (約 2.5 割) であった．その他にも「デザイン $_i$ は好みが分かれるでしょうが，私は (ϕ_i ガ)〈 好き 〉$_v$ です」のように属性が省略されている場合や「内外装の 〈 デザイン 〉$_a$ がモデル末期となった現在、同クラスの最新の車と比べると斬新性という点においてやや 〈 見劣りする 〉$_v$ 」のように属性と評価値が長距離依存である場合「子供っぽい 〈 スタイリング 〉$_a$.〈 最悪 〉$_v$ 」のように属性と評価が文を越える場合が見られた．述語項構造を頑健に解析するにはこれらの問題を解く必要がある．

[1]「車内は広くて落ち着く」の場合「車内」が属性で「広い」が属性値であり「落ち着く」は評価であると考えられるが「落ち着いた内装」の場合「落ち着く」は内装の属性値と考えられる．

[2]$\langle\rangle_a$ は属性を表し，$\langle\rangle_v$ は評価値を表すとする

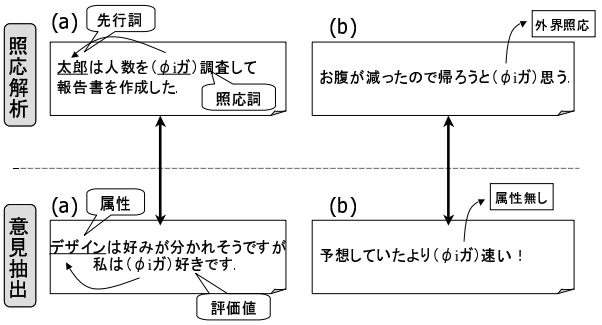

図 2: 照応解析と属性-評価値対抽出

我々は照応解析と意見抽出の問題の類似性に着目し，照応解析手法を適用することの有効性を検証し，実際にどの問題が解けて，どの問題が残るかを明らかにする．

3 意見の抽出手法

レビューサイトや掲示板など意見抽出の対象となる文章ではどの対象についての記述かが自明である場合が多いため，ここでは ⟨ 対象，属性，評価値 ⟩ のうち，とくに ⟨ 属性，評価値 ⟩ の抽出に焦点を当てる．

意見抽出には図 2 に示すように (a) 評価値に対応する属性を同定する，(b) 評価値が属性を含意して談話中に属性が陽に現れない場合に評価値が属性を持たないことを判定するという 2 つの問題がある．一方照応解析では，(a) 照応詞候補に対して先行詞を同定する，(b) 先行詞が談話中に存在しない場合に，照応詞が真の照応詞でないことを判定するという 2 つの問題がある．この 2 つの問題は構造が似ているため，類推的に同じ手法で解析できるのではないかと考える．

照応解析では，照応詞に対して先行詞候補を先に同定して照応詞が先行詞を持つか否かを判定するという方法で比較的良い結果が得られている [3]．そこで，意見抽出でも評価値に対して属性候補を先に同定し，評価値が属性を持つか否かを判定する方法で良い結果が得られるのではないかという仮説を立てる．

また意見抽出では評価値が意見であるかどうかを判定する必要がある．この問題を意見性判定と呼ぶ．意見性判定は，例えば「故障は多い」は意見で「車が多い」は意見ではないということを判定する問題である．この問題は上の「多い」に見られるように，とり得る属性によって意見性の有無がわかれるため，対判定と同様に属性を先に同定し，それを手がかりとして提示することで良い結果が得られるのではないかと予測できる．

この考えに基づいて，意見抽出の問題を以下の手順で解析する．

1. 辞書を用いて評価値と属性の候補 (以下，属性候補，評価値候補) を探す
2. 属性同定: 評価値候補に対して対となる属性候補を同定する
3. 対判定: 評価値候補が属性を持つか否かを判定する
4. 意見性判定: 評価値候補が意見性を持つか否かを判定する

表 1: 属性-評価値の出現傾向

属性と評価値が同一文内に出現	2337	(82.5%)
属性が評価値の一文前に出現	60	(2.1%)
属性無し評価値	424	(15.0%)
評価値出現総数	(2821/2833)	(99.6%)

2，3 には機械学習ベースの照応解析の手法を適用する．4 は意見性を持つか否かの 2 値分類問題と考え，機械学習を用いて解く．

ここで，照応解析と意見抽出のもう一つの相違点について触れておく．照応解析では照応詞に対して一つの先行詞を同定すれば良いのに対し，意見抽出では「⟨ 振動 ⟩$_a$ や ⟨ 騒音 ⟩$_a$ が ⟨ 少ない ⟩$_v$ 」のように評価値が複数の属性をとる場合がある．しかし，まずは評価値に対して一つの属性を同定し抽出することに焦点をあて，この問題はここでは扱わないことにする．

4 評価実験

前節で立てた 2 つの仮説が正しいか，照応解析の手法を意見抽出に適用することで依存関係にない属性-評価値対をどの程度抽出できるかを検証するために評価実験を行なった．

4.1 設定

4.1.1 意見性

どのような記述を意見とみなすかはタスクに依存するため一般的には決められない．今回の実験では「商品またはその側面に対する記述者本人の評価」が意見性を持つと定義し，要望や反事実文の場合は意見性は持たないとした．例えば「インテリアは最高です」は意見性を持つが「安かったらいうことなし」のような仮定表現は意見性を持たないとする．

4.1.2 意見タグ付きコーパス

車に関するレビュー記事 287 記事，4442 文を対象に属性と評価値のタグを付与した．何が評価値であるかは難しい問題であるが，タグ付与者の主観で 4.1.1 の意見性を満たす属性，評価値と思われる表現にタグを付与した．付与されたタグの数は属性有り評価値 2409，属性無し評価値が 424 であった．

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 1 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．そこで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補を探索する．

4.2 実験方法

前処理として茶筌[1]と CaboCha [2]を用いて形態素解析と係り受け解析を行う．属性同定，対判定，意見性判定のモデルの学習にはそれぞれ Support Vector Machine[12] 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．

[1] http://chasen.naist.jp/
[2] http://chasen.org/~taku/software/cabocha/

表 2: 実験結果

手順		属性有り評価値		属性無し評価値		属性-評価値対全体	
手順 1	精度	62.7%	(1112/1774)	37.5%	(201/536)	56.8%	(1313/2310)
	再現率	46.2%	(1112/2409)	47.4%	(201/424)	46.3%	(1313/2833)
手順 2	精度	48.8%	(1049/2151)	35.2%	(56/159)	47.8%	(1105/2310)
	再現率	43.5%	(1049/2409)	13.2%	(56/424)	39.0%	(1105/2833)
手順 3	精度	70.2%	(1307/1863)	34.2%	(166/485)	62.7%	(1473/2348)
	再現率	54.2%	(1307/2409)	39.2%	(166/424)	52.0%	(1473/2833)
手順 4	精度	80.1%	(1209/1509)	33.4%	(148/443)	69.5%	(1357/1952)
	再現率	50.2%	(1209/2409)	34.9%	(148/424)	47.9%	(1357/2833)

4.2.1 辞書

属性候補と評価値候補の同定には辞書を用いるため，ある程度大規模な辞書が必要となる．そこで，タグ付きコーパスとは重複しない，車に関する記事約 23 万文から文献 [5] で提案した手法を用いて属性表現辞書と評価値表現辞書を作成した．手法については文献を参照されたい．さらに，タグ付きコーパス内で頻度が高く，上記辞書に存在しなかった表現を追加して，属性表現辞書 3777 表現，評価表現辞書 3962 表現を辞書とした．

コーパス中のタグが付与された表現のうち，辞書に存在したのは属性有り評価値 1795，属性無し評価値 351 であった．この値が今回の実験で抽出できる最大数であるため，対抽出の再現率の上限値は 75%，属性無し評価値の再現率の上限は 83%になる．

4.2.2 タスク設定

入力は 4.1.2 で述べた意見タグ付きコーパスである．タグの付与された属性-評価値対のうち，属性表現辞書と評価表現辞書に存在する対が抽出対象となる．

まず，属性表現辞書と評価値表現辞書をコーパスに適用して，属性候補および評価値候補を探す．次に，各評価候補に対して，属性候補があればそれを同定し (属性同定)，評価値が対を持たない場合は属性が空であることを判定 (対判定) する．これらの処理により属性と評価値の対が抽出される．以下，この対を属性が空である場合も含めて属性-評価値対と呼び，属性同定と対判定をあわせて対抽出と呼ぶ．また，抽出した対が意見性を持つか否かを判定し (意見性判定)，意見性を持つ属性-評価値対を正解として出力する．

評価尺度には精度と再現率を使用する．それぞれは以下の式で求める．

$$\text{再現率} = \frac{\text{正しく抽出できた意見性を持つ属性-評価値対の数}}{\text{意見性を持つ属性-評価値対の総数}},$$

$$\text{精度} = \frac{\text{正しく抽出できた意見性を持つ属性-評価値対の数}}{\text{システムが意見性を持つと判断した属性-評価値対の数}}.$$

4.2.3 モデルの適用順

前節で述べた，(1) 属性同定後に対判定を行なう方が対判定の結果が良い，(2) 属性同定後に意見性判定をする方が意見性判定の結果が良いという 2 つの仮説を検証するために，以下の 4 種類の手順の比較実験を行なった．

手順 1: 意見性判定+属性同定+対判定
手順 2: 意見性判定+対判定+属性同定
手順 3: 属性同定+意見性判定+対判定
手順 4: 属性同定+対判定+意見性判定

手順 1 と手順 2 を比較することで (1) の仮説を検証でき，手順 1，2 と手順 3，4 を比較することで (2) の仮説を検証できる．

属性同定には飯田らのトーナメントモデル [2] を使用した．この手法では，複数存在する属性候補に対して候補間で勝ち抜き戦を行ない最尤の属性候補を決定する．対判定では，属性候補と評価値の対の情報を用いて，評価値が属性を持つか否かを判定する．手順 2 の対判定では属性同定が後にくるため Soon らのモデル [9] を使用した．このモデルでは評価値に対してすべての属性候補が候補となるか否かを判定し，どの属性候補とも対にならなければ評価値は対を持たないと判定する．また意見性判定では，評価値の情報から意見性を判定する．先に属性を同定する場合は属性の情報も利用する．手順 4 の意見性判定では，対判定の情報が使用できるため，属性無し評価値の場合と属性有り評価値の場合で別々にモデルを作成した．

4.2.4 素性

属性候補と評価値候補から以下の 2 種類の素性を抽出した．

a. 対象とする文節とその係り先，係り元の文節の表層文字列や品詞
b. 候補間の距離情報，属性候補と評価値候補の対が共起用例に含まれるか，属性候補と評価値候補が係り受け関係にあるか

b. の共起用例は「トルク–ある」「燃費–悪い」のような属性と評価値の対情報である．4.2.1 で述べた 23 万文の記事集合から，辞書中の属性と評価値が係り受けの関係にあった場合に対を抽出した．その数は約 4,8000 用例であった．

属性同定と対判定ではともに a と b を抽出した．また手順 1，2 の意見性判定では評価値に対して a のみを抽出するのに対して，手順 3 では a と b の両方を抽出した．手順 4 の意見性判定は，評価値が属性をもつ場合は手順 3 と同じ情報を抽出し，評価値が属性無しの場合は手順 1，2 と同じ情報を抽出した．

4.3 結果・考察

意見抽出の実験結果を表 2 に示す．精度および再現率は 4.2.2 で述べた式で求めた．

3 節で述べた仮説を検証するために，対抽出と意見性判定の結果を独立に求めた．その結果を表 3 に示す．

表 3: 対抽出と意見性判定の結果

	手順	精度		再現率	
対抽出	係り受け関係	41.2%	(1154/2803)	47.9%	(1154/2409)
	対判定 → 属性同定	62.4%	(1933/3098)	68.2%	(1933/2833)
	属性同定 → 対判定	74.0%	(2029/2741)	84.2%	(2029/2409)
意見性判定	意見性判定	75.3%	(1739/2310)	61.4%	(1739/2833)
	属性同定 → 意見性判定	79.5%	(1933/2430)	68.2%	(1933/2833)

表中の「係り受け関係」は，辞書にある属性表現と評価値表現が係り受け関係にある場合に属性-評価値対として抽出したときの精度と再現率を表している．

4.3.1 対判定に関する考察

表 3 より，評価値に対して対となる属性を同定して，評価値が属性を持つか否かの判定を行なう方が逆の場合より精度が 12%，再現率が 16% 向上したことが分かる．この結果は「属性同定の結果を提示することで対判定の結果が良くなる」という仮説を支持している．

また，属性同定を行なった後に対判定を行なう方法で係り受け関係にある属性-評価値対，係り受け関係にない属性-評価値対をどの程度抽出できたかを調査した．その結果，係り受け関係にある対のうちの 96%と，係り受け関係にない対の 53%をそれぞれ抽出できていることがわかった．係り受け関係にない対で抽出できなかった事例 100 件を調査したところ，今回対象としなかった，評価値が複数の属性をとり得るという問題が 25%を占めていた．この問題は今後の課題であるが，複数の属性を抽出対象とすることで精度の低下が予想されるため，対とならない属性をうまく棄却するために何らかの制約を導入する必要があると考えられる．

4.3.2 意見性判定に関する考察

表 3 より，先に属性を同定して意見性を判定するほうが精度と再現率が少し良いことがわかる．対判定と同様に，前節で立てた仮説は正しかったといえるだろう．

しかし属性同定を先に行なう手順でも，精度，再現率は高い値ではない．誤った事例を分析したところ，誤りの典型的な例はなく，問題はさまざまであった．その中で比較的多い誤りとして「〈 座面 〉$_a$ が 〈 下がってくれれば 〉$_v$ 」のような仮定表現や条件表現を含む場合に誤って意見性ありと判断する例があった．この問題については今後詳しく分析し，問題を解くために必要な情報を明らかにしたい．

5 おわりに

本稿では，照応解析と意見抽出の問題の類似性に着目し，学習ベースの照応解析手法を意見抽出に適用することを試みた．実際に評価実験を行なった結果，照応解析の問題と同様に，対となる表現の情報を提示することで対判定や意見性判定の解析精度が良くなることがわかった．しかし，対判定や意見性判定の解析精度は高くはないことや評価値が複数の属性と対になる場合が大きな問題であることも分かった．今後は事例をさらに詳しく分析し，これらの問題を解決するために必要な情報を明らかにしたい．

参考文献

[1] Kushal Dave, Steve Lawrence, and David M. Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *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. 文脈的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] 飯田龍, 乾健太郎, 松本裕治, 関根聡. 最尤先行詞候補を用いた日本語名詞句同一指示解析. 情報処理学会論文誌, Vol. 46, No. 3, to appear.

[4] Hiroshi Kanayama and Tetsuya Nasukawa. Deeper sentiment analysis using machine translation technology. In *Proceedings of the 20th International Conference on Computational Linguistics*(*COLING*), pp. 494–500, 2004.

[5] Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi, and Toshikazu Fukushima. Collecting evaluative expressions for opinion extraction. In *Proceedings of the 1st International Joint Conference on Natural Language Processing* (*IJCNLP*) , pp. 584–589, 2004.

[6] Tony Mullen and Nigel Collier. Sentiment analysis using support vector machines with diverse information. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), pp. 412–418, 2004.

[7] 村野誠治, 佐藤理史. 文型パターンを用いた主観的評価文の自動抽出. 言語処理学会 第 9 回年次大会 発表論文集, pp. 67–70, 2003.

[8] Bo Pang and Lillian Lee. A sentiment education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics* (*ACL*) , pp. 271–278, 2004.

[9] W. M. Soon, H. T. Ng, and D. C. Y. Lim. A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, Vol. 27, No. 4, pp. 521–544, 2001.

[10] 立石健二, 石黒義英, 福島俊一. インターネットからの評判情報検索. 情報処理学会研究報告 NL144-11, pp. 75–82, 2001.

[11] Peter D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (*ACL*), pp. 417–424, 2002.

[12] V. N. Vapnik. *Statistical Learning Theory*. Adaptive and Learning Systems for Signal Processing Communicat ions, and control. John Wiley & Sons, 1998.

[13] Hong Yu and Vasileios Hatzivassiloglou. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), pp. 129–136, 2003.