意見情報抽出のための評価対象・評価視点間の関係同定

小林のぞみ　　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,inui,matsu}@is.naist.jp

1 はじめに

ブログジャーナリズムのような新語の出現に象徴されるように，個人がネット上に発信する意見情報の重要性が広く認知されるようになった．これに伴い，個人の意見を Web 上のテキストから収集し，分析する技術に関する研究が活発になっている．こうした研究は文書分類に基づくアプローチと情報抽出に基づくアプローチに大別することができる．前者では，テキストを記事や段落（または文）の単位で肯定的意見か否定的意見かといった意見クラスに分類する課題を設定し，機械学習に基づく様々な手法の実験結果が報告されてきた [1, 11, 15]．一方後者は，意見の構成要素とそれらの間の関係をテキストから抽出する課題に取り組むもので，情報抽出あるいは関係抽出の技術を適用する試みの報告が増えつつあり (例えば，[12, 4, 5, 13])，我々もテキスト中の意見情報を〈評価対象，属性，評価値〉の 3 つ組として抽出する技術の研究を進めてきた [6]．

テキスト中の意見情報には「〈アコード〉の〈エンジン〉の〈音〉が〈良い〉」のように，対象やその部分，属性と，それに対する評価値という形で構造化できるものが比較的多く，これらの関係を同定できれば意見の収集や要約に役立つ．本稿では，この部分問題として，対象から属性に至る階層関係の同定に焦点をあて，人手による辞書の作成を仮定せずにこの問題を解く方法を論じる．また，作成した意見タグ付きコーパスを用いて実験を行ったので，その結果を報告する．

2 意見情報抽出の問題設定

意見情報抽出を，何を抽出する問題とするかについてはさまざまな議論がなされているが，我々はまず以下のような評価に着目する．

(1) 評価の主体が評価対象のクラスに対して期待している標準値から実際の評価対象がどれくらい偏位しているか，そしてそれが「望ましい」偏位であるか「望ましくない」偏位であるかに関する評価主体の認識であり，その言明

この定義に基づき，評価を構成する以下の 4 種類の情報を抽出する問題を考える．

評価主体　評価を行った主体を指す表現

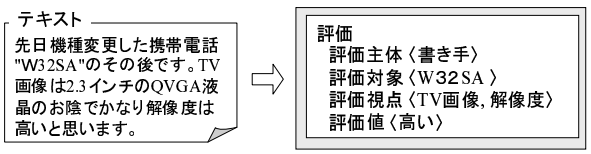

図 1: 抽出したい評価

評価対象　製品やサービス，人物，組織，政策など，特定のクラスの実体を指す表現

評価視点　評価対象のどの構成要素あるいは属性が評価の焦点になっているかを示す表現

評価値　評価対象全体あるいは評価視点に対する評価の値を示す表現

図 1 に具体例を示す．評価の主体はテキストの書き手，対象は「Ｗ３２ＳＡ」という携帯電話の機種である．

評価対象と評価視点の関係には主として全体–部分関係（製品と部品など），全体–役割関係（店と店員など），対象–属性関係（製品と性能など）がある．評価対象全体に対する評価の場合は評価視点は空であると見なす．また，評価視点がさらにその内部に階層構造を持つ場合があることに注意する必要がある．例えば，図 1 の例で「高い（評価値）」のは「（（Ｗ３２ＳＡの）ＴＶ画像の）解像度」であり，評価視点を構成する「ＴＶ画像」と「解像度」の間に対象–属性関係を認めることができる．我々はこれを〈ＴＶ画像，解像度〉のようなリストで表し，最上位の評価対象から評価視点の階層の末端に至る〈Ｗ３２ＳＡ，ＴＶ画像，解像度〉のようなパスを評価対象のチェーンと呼ぶ．

これらの構造を人手で抽出可能か否かを調べるために，評価値と評価対象のチェーンにタグを付与した意見タグつきコーパスを作成した [7]．タグ付与の対象は Web 上の blog 記事で，対象ドメインはレストランや携帯電話など，4 ドメイン，規模は約 2,900 記事である．うち，階層関係はレストランドメインで約 5,300 対出現しており，評価対象のチェーンに関する作業者間の一致率を見たところ，タグの付与範囲のゆれを除けばある程度 (7 割) は一致するという結果が得られた．

2.1 情報抽出問題へのアプローチ

我々は，上述の意見抽出の問題を (1) 評価値と評価視点（評価対象）間の関係同定，(2) 評価視点間の階層関係の同定の二つの問題に分ける．また前処理として，評価対象 (固有名詞) と同一指示対象の同定が必要である．意見情報抽出の流れをまとめると以下のようになる．

Identifying aspect-aspect/subject-aspect relations for opinion extraction
Nozomi Kobayashi, Kentaro Inui and Yuji Matsumoto
Nara Institute of Science and Technology

1. 評価対象の同定: テキストのどの部分が評価対象となっているかを当てる
2. 同一指示対象の同定: 例えば「久しぶりに ⟨ 珍萬 ⟩$_i$ にいってきました．⟨ ここ ⟩$_i$ の炒飯はかなりおいしいです．」の場合に，「ここ」が評価対象「珍萬」と同じであることを同定する
3. 評価視点 (評価対象) –評価値対の抽出: 評価値と評価されている評価視点 (評価対象) を抽出する
4. 評価対象のチェーンの抽出: 評価対象に至るまで評価視点間の階層関係を同定する

1 と 2 は一般的なタスクで，多くの手法が提案されている．3 については，問題の構造が照応解析の問題と似ていると考え，照応解析で用いられている機械学習手法を適用することを提案した [6]．そこで本稿では 4 に焦点をあて，人手による辞書の作成を仮定せずに解く方法を論じる．

3 評価対象・評価視点間の関係同定

本稿では，文内の評価対象・評価視点間の関係同定タスクをある評価視点「B」が与えられたときに，「B」と全体–部分関係，全体–役割関係，対象–属性関係にある評価視点もしくは評価対象「A」を同定する問題と考える (図 1 の例では「解像度」が「B」，「TV 画像」が「A」になる)．

全体–部分関係とは，評価対象もしくは評価視点とその部位，部品など (車とエンジン，店と内装，携帯電話と液晶など) を指し，全体–役割関係は，ある概念に依存して何らかの役割を表わす関係で，たとえば，学校と教師，病院と医師，店と店長，店と店員などがそれにあたる [10]．対象–属性関係とは，評価対象もしくは評価視点とその性質，特徴など (エンジンと音，内装とデザイン，車と色，店と雰囲気など) を指す．以下，これらの関係をまとめて **aspect-of** 関係と呼ぶ．

関係が文内に出現する場合とそうでない場合では，出現の仕方がかなり異なるため，ここでは文内に出現する場合に絞って問題を考える．

文内に出現する aspect-of 関係は「携帯電話の液晶」のように「A の B」の形で出現することが多い．実際，後述の評価に使用したデータでは，文内に出現する対のうち 3 割が「A の B」の形で出現していた．しかしながら，表層的に同じ接続助詞「の」で結ばれていても二つの名詞句の間にはさまざまな関係があり [9]，「A の B」の形で出現していても aspect-of 関係であるとは限らない．たとえば，「私の車」「人気のランチ」「最近の携帯電話」は「A の B」の形をしているが，aspect-of 関係ではない．そのため，「A の B」で出現したという情報だけでなく，二つの間にどのような関係があるかを判定することが重要になる．

3.1 提案手法

提案手法では，人手で作成した辞書を仮定せずに (1) 与えられた評価視点と候補の間のうちに aspect-of 関係があるか否かの情報，(2) 与えられた評価視点と候補との共起情報，を使用して aspect-of の関係にある評価対象・評価視点の対を抽出する．以下，aspect-of 関係をどのように推定するか，最尤候補をどのように同定するかについて述べる．

3.1.1 関係推定

この問題は，与えられた評価視点「B」とある候補「A」が R である確率 $P(R|A,B)$ が最大になる R を求める問題である (R は ⟨ 全体–部分 ⟩，⟨ 全体–役割 ⟩，⟨ 対象–属性 ⟩ もしくは ⟨ その他 ⟩ とする)．

そこで，あらかじめ人手で関係のラベルをデータに付与しておき，そのデータを用いて確率モデルを推定することを考える．モデルの推定には最大エントロピー法を使用する．

素性には，A と B それぞれの分類語彙表でのクラス番号 (上位二桁) とどのような評価値と共起したかを使用する．

3.1.2 最尤候補の同定

関係推定モデルが高確率で「その他」と推定した名詞句は候補から除外し，それ以外の候補の中で最も「B」と対になりそうな候補を同定する．同定の際には以下の情報を使用する．

(1) 与えられた評価視点と候補が aspect-of 関係か否か
(2) 与えられた評価視点と候補との共起情報

(1) は関係推定モデルの結果，(2) には「A の B」の共起情報をそれぞれ使う．しかし「A の B」は上述したように aspect-of 以外の関係も多く含んでいるため，そのまま使用したのでは効果が得られないと予想できる．そこで以下の式から求めた「A の B」の生起確率を使用することを考える (R_{asp} は aspect-of 関係を指す)．

$$P(A|B,R_{asp}) = \frac{P(R_{asp}|A,B)}{P(R_{asp}|B)}P(A|B)$$

$P(A|B)$ を求める際に，PLSI(Probabilistic Latent Semantic Indexing)[3] を使用して確率のスムージングを行う．PLSI の結果を使用した $P(A|B)$ の式は次式で与えられる．

$$P(A|B) = \frac{\sum_{z\in Z} P(A|z)P(B|z)P(z)}{P(B)}$$

上記の情報に加え，文節の主辞の品詞，機能語などの統語的な情報も使用し，文の構造を学習できる Bact[8] を用いてモデルを学習する．Bact は入力された木構造のデータから，部分木を素性とする decesion stumps を抽出し，それを弱学習器として用いる boosting アルゴリズムである．Bact を使用することで，「A の B」の形で出現していれば A は B と対になりやすいというようなパタンを学習できると期待できる．

4 評価実験

(1)aspect-of とそれ以外の関係がどの程度推定できたか，(2) 与えられた評価視点に対して，どの程度正しく対 (以下，評価視点対と呼ぶ) を抽出できたかを評価するための実験を行なった．

4.1 事例収集

「A の B」の事例の収集には大規模なデータが必要である．今回は Web 上の blog，7ヶ月分 (約 175 万ページ) から「A の B」の出現頻度，「A が評価値」の出現頻度をそれぞれ求めた．また，各内容語の tf·idf もこのデータ集合から求めた．

4.2 関係推定

収集した「A の B」の事例のうち，「B」を実験に使用したデータ中で評価視点になった表現に絞り，2,767 対について全体–部分関係，対象–属性関係，全体–役割関係とそれ以外のタグを人手で付与した．

このうちの 1,657 対を訓練事例として使用し，残り 1,110 対を評価に使用した．ここで，「その他」のラベルが付与された事例が他に比べて非常に多かったため「その他」の事例はランダムにサンプリングして使用した．最大エントロピー法は Maxent パッケージを使用した[1]．

4.3 最尤候補の同定

モデルの学習および評価には Bact を用い，評価は 5 分割交差検定で行なった．

4.4 学習・評価事例

2 で述べた意見タグつきコーパスのうち，レストランドメインのデータ 460 記事を対象に実験した．460 記事中に存在する文内の評価視点 (評価対象) –評価値の数は 289 である．

このデータに対し，前処理として茶筌[2]と CaboCha[3] を用いて形態素解析と係り受け解析を行った．

4.4.1 素性

候補表現の分類語彙表でのクラス番号上位二桁，文節の主辞の品詞，機能語などの情報を使用した．また，候補表現の tf·idf の値が候補内で何位であったか，候補が aspect-of 以外か，$P(A|B, R_{asp})$ の値が候補中で何位であったかを素性として使用した．

4.5 結果・考察

4.5.1 関係推定

関係推定の結果を図 2 に示す．人手でラベル付けされた事例 1,110 について，出力された確率が「その他」が高かった場合は「その他」，それ以外は aspect-of 関係として精度と再現率を以下の式で求めた．1,110 事例のうち，736 がその他で 374 が aspect-of 関係であった．

$$\text{再現率} = \frac{\text{正しく同定できた「A の B」の数}}{\text{「A の B」の数}},$$

$$\text{精度} = \frac{\text{正しく関係が同定できた「A の B」の数}}{\text{システムが出力した「A の B」の数}}.$$

図 2 から，ある程度の精度で推定ができているといえるが，まだ改良の余地がある．精度向上のために，ラベル有り事例から推定したモデルを使用して，大量のラベルの付与されていない事例のラベルを推定し，ブートストラップを行う手法が考えられるが，今回の問題設定では効果が得られなかった．今後は使用する素性を取捨選択するなどして精度の向上に努めたい．

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

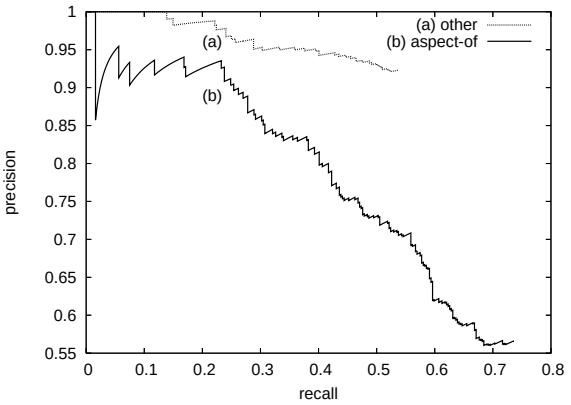

図 2: 関係推定の結果

4.5.2 最尤候補同定

「A」の候補は名詞を含む文節とし，関係推定モデルが確率 0.98 以上で「その他」と判断された場合は候補から除いた．削除された例はいずれも対とはならない表現であったが，その数は 8 つと少ないものであった．しかし閾値を下げると対となる候補まで削除してしまうため，現状では推定結果をそのまま候補選択に使用するのは難しい．

次に候補同定の結果について議論する．ベースラインとして，「A の B」の形で「A」が出現していた時に正解だとした際の精度，再現率を導入し，そこからどの程度精度を低下させることなく再現率を向上させることができたかをみた．ベースラインモデルの精度は 0.75(87/116)，再現率は 0.30(87/289) であった．図 3 は，入力された「B」に対して，Bact のスコアが最も高かった候補を出力した場合の再現率-精度曲線を表す．使用した素性の有効性を調べるために，以下の 4 つの結果を出力した．

(a) 「A の B」の共起情報を使用しない (ML baseline)
(b) $P(A|B, R_{asp})$ の結果を使用
(c) $P(A|B, R_{asp})$ の結果と関係推定モデルの判定結果を使用
(d) $P(A|B)$ の結果と関係推定モデルの判定結果を使用

精度および再現率は次式で求めた．

$$\text{再現率} = \frac{\text{正しく抽出できた評価視点対の数}}{\text{評価視点対の総数}},$$

$$\text{精度} = \frac{\text{正しく抽出できた評価視点対の数}}{\text{システムが出力した評価視点対の総数}}.$$

上記の結果から，共起情報を使用することで，精度，再現率が向上していることがわかる．しかしながら，推定された関係を共起の計算や素性に使用することによる効果はほとんどみられなかった．誤った例を分析したところ，「ここの鰻」「稲庭うどんの歯ごたえ」のように，「の」で結ばれた二つは対になりやすいというパタンや，「〈ボルシチ〉$_A$ は優しい〈味〉$_B$ で、特に気に

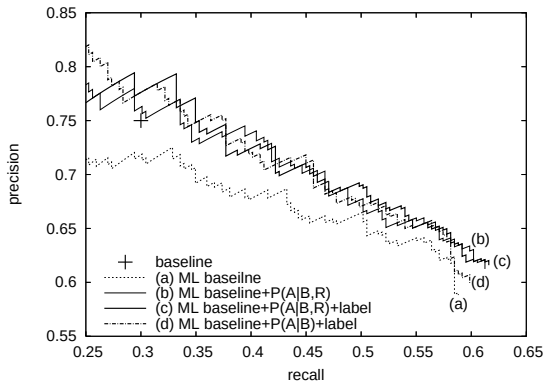

図 3: recall-precision 曲線 (最尤候補同定)

入った。」のように，文頭に「A」が出現しやすいというパタンをうまく学習できていたのだが，高い重みがついたために，以下のような場合に誤った候補を抽出していた．

(2) ⟨ 和食屋さん ⟩$_A$ ですが、⟨ お手ごろ価格 ⟩$_{error}$ の ⟨ ランチ ⟩$_B$ がおいしいです。

また，今回導入した共起の情報にも高い重みがついていたため，推定が間違った際の結果に影響したと考えられる．今後，関係推定の精度向上にむけてさらに分析したい．

その他に，誤った事例の中で Bact のスコアが高いものを分析したところ，文境界がうまく判定できていないために複数の文が一文として扱われていることがノイズになっているようであった．今回は簡単なヒューリスティクスで文境界を判定していたのだが，福岡らの手法 [2] の適用を考えたい．

5 関連研究

意見抽出の先行研究では，主に評価値と評価視点 (もしくは評価対象) の抽出問題に焦点が当てられてきた．例えば，立石ら [14]，Popescu ら [12] はパタンを用いて評価値と評価視点を抽出している．また，Hu ら [4] は，評価視点とその視点に対する肯定/否定の値を同定する手法を提案している．しかしながら，今回我々が取り組んだような評価視点間の階層関係の抽出を扱った研究は今のところ報告されていない．

今回対象とした問題は，関係抽出問題[4] の一つと考えることができる．関係抽出の問題は，一般に表現のラベル (例えば人名，会社名，役職名) などが与えられた上でその間の関係を同定する問題であるが，今回は，表現にラベルが付与されていない状態で共起の情報などを使用して，求めたい関係になっている対を抽出する問題を考えた．

6 おわりに

本稿では，⟨ 評価主体，評価対象，評価視点，評価値 ⟩ を抽出する問題の設定について述べ，そのうちの評価対象・評価視点間の関係を抽出する手法について提案した．また評価実験を通して，現段階の関係推定の精度では，共起情報の見積もりに十分ではないことがわかった．今後はそれらの問題を解決するために必要な情報などを分析すると同時に，文間に出現する場合にどの程度今回の手法が有効かを調査したい．また，今回使用した意見タグ付きコーパスの問題点も明らかになってきたので判定モデルの洗練と同時にコーパスも洗練していく予定である．

なお，意見タグ付きコーパスの一部は，文部科学省科学技術研究補助金・基盤研究 A「円滑な情報伝達を支援する言語企画と言語変換技術」(研究代表者: 佐藤理史氏 (名古屋大学) , 課題番号: 16200009) の支援を受けて作成した．

[4] Automatic Content Extraction
http://www.itl.nist.gov/iaui/894.01/tests/ace/

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* (*WWW*), pp. 519–528, 2003.

[2] 福岡健太, 松本裕治. Support vector machines を用いた日本語書き言葉の文境界推定. 言語処理学会 第 11 回年次大会, pp. 1221–1224, 2005.

[3] Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual ACM Conference on Research and Development in Information Retrieval*, pp. 50–57, 1999.

[4] M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proc. of the Tenth International Conference on Knowledge Discovery and Data Mining*, pp. 168–177, 2004.

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

[6] Nozomi Kobayashi, Ryu Iida, Kentaro Inui, and Yuji Matsumoto. Opinion extraction using a learning-based anaphora resolution technique. In *The Second IJCNLP, Companion Volume to the Proceeding of Conference including Posters/Demos and Tutorial Abstracts*, pp. 175–180, 2005.

[7] 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. 情報処理学会研究報告 NL-171, pp. 111–118, 2006.

[8] 工藤拓, 松本裕治. 半構造化テキストの分類のためのブースティングアルゴリズム. 情報処理学会論文誌, Vol. 45, No. 9, pp. 2146–2156, 2004.

[9] 黒橋禎夫, 酒井康行. 国語辞典を用いた名詞句「a の b」の意味解析. 情報処理学会研究報告 NL-129-16, pp. 109–116, 1999.

[10] 溝口理一郎. オントロジー工学. オーム社, 2005.

[11] 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.

[12] Ana-Maria Popescu and Oren Etzioni. Extracting product features and opinions from reviews. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), pp. 339–346, 2005.

[13] 鈴木泰裕, 高村大也, 奥村学. Semi-supervised な学習手法による評価表現分類. 言語処理学会 第 11 回年次大会 発表論文集, 2005.

[14] 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.

[15] 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.