

レビューテキストを対象とした評価条件の抽出手法

中山 祐輝 藤井 敦

東京工業大学大学院情報理工学研究科計算工学専攻

1 はじめに

情報化技術の発展により、個人が様々な情報を手軽に発信できるようになった。その中には、利用者の意思決定支援に有益であるような主観的な評判情報が膨大に含まれている。評判情報とは「商品Aの〇〇が良い」という「評価」や「商品Bの××がうれしい」という「感情」が記述されている情報である。

しかし、情報過多により有益な評判情報を利用者自身で発見することは、多大な時間コストを要する。よって、評判情報の抽出、整理を自動的にを行い、評判検索や評判の提示に応用できる技術が求められている。本論文では、これらの技術を総称して評判分析と呼ぶ。近年、レビューテキスト、ブログ記事、Twitterなどのメディアから、評判分析が盛んに行われている。そのタスクの一つに、テキストから評価対象、評価属性、評価表現を抽出し、〈評価対象、評価属性、評価表現〉の三つ組の関係を同定するタスクがある[1, 7]。例えば、「ホテルAのアメニティが良い」において、評価対象は対象商品を示す「ホテルA」である。評価属性は「アメニティ」のような評価対象の特徴や構成要素を表す。評価表現は「良い」のような主観的な評判である。

従来のタスクで抽出された評判は、全ての利用者に当てはまることを想定している。しかし、評判はある条件によって異なる場合があり、必ずしも全ての利用者に当てはまるとは限らない。例えば、「ホテルBは、ミナミエリアで遊ぶには良いロケーション。」という評価文における従来のタスクは、〈ホテルB, ロケーション, 良い〉という三つ組の関係を同定することである。しかし、「良い」という評価表現は「ミナミエリアで遊ぶには」という条件をもって成立する評価表現であり、ビジネス目的の宿泊客に当てはまるとは限らない。つまり、このような条件を考慮することで、個々の利用者に当てはまる評判をより獲得できる。本論文では、評判中の評価表現が成り立つための条件を**評価条件**と呼ぶ。

我々は評価条件を6種類に分類する。以下で、評価条件を含む例文を種類ごとに示す。なお、本論文で示される全ての例文は、実際に楽天トラベルのレビューサイトに投稿された例文である。本論文中の全ての下線は、評価条件を表す。

- **利用者**：特定の利用者で評価表現が成立する
(例) 高齢者には ベッドが小さかったです。
- **利用状況**：特定の場面や時点で評価表現が成立する
(例) 朝食は、パンですが、連泊するときは物足りないと思います。

- **利用目的**：特定の利用目的で評価表現が成立する
(例) 部屋は綺麗で ビジネス利用には十分です。
- **反実仮想**：現在は条件を満たしていない。しかし、いつか満たせば評価表現が成立する
(例) 料理は 金額が安ければ満足です。
- **その他の条件**：ある条件が満たしているなら評価表現が成立する
(例) 朝食は、クロワッサンで、一泊だけならいい。
- **意見保持者**：意見を述べている特定の人物・団体に評価表現が成立する
(例) お風呂は 子どもが大変気に入っていました。

意見保持者の抽出は先行研究がある[4]。本研究では、意見保持者はある利用者の視点から見た評価であると考え、意見保持者も評価条件に含める。評価条件の分類は、特定の利用者に向けた評判のみを知りたいなど、評判検索時の条件の絞り込みに役立つ。

本研究は評判検索と提示のために、評判が記述されているテキストから、利用者や条件を考慮した評判分析を行うことを目的としている。その第一歩として、本論文では旅行サイトのレビューテキストを対象とした評価条件の抽出手法を提案する。本研究のタスクは評価条件の抽出と分類、「出張には」と「ビジネスには」のような評価条件の同義表現をまとめるためのクラスタリング、そして四つ組の同定の三つに大別される。

2 関連研究

評価条件は「高齢者には」などの特定の利用者における立場や視点を含んでいる。Linら[3]は、パレスチナからの視点で見たニュース記事とイスラエルの視点で見た同じ内容のニュース記事を同定している。また、異なるニュース内容について同じ視点から見られているかも同定している。Mullenら[5]は政治対話集から、共和党、保守党、民主党、自由党といった、どの政治的立場かを分類している。LinらとMullenらにおいて、原文中にはイスラエルにとっては「賛成」や民主党からみると「良い」などの評価条件が含まれていると予想される。LinらとMullenらは立場や視点に基づいて記事と発言をそれぞれ分類しているのに対して、本研究はこのような立場や視点に関する記述を抽出する点で異なる。

西川ら[6]は旅行のレビューテキストを対象として、レビュー投稿者の利用目的や状況を「ビジネス」、「カップル・恋人」、「家族旅行」、「友達」、「一人旅」のいずれかに自動分類する手法を提案している。西川らは特定の利用者、目的、状況における評価を考慮している点では、

本研究と類似している。しかし、西川らはレビューテキスト全体を特定の利用目的や状況に基づいて分類している。それに対し、本研究は多様な利用者、目的、状況の評価をレビュー本文から抽出する点で異なる。

3 評価条件の抽出手法

本研究はレビュー文を文節毎に区切り、各文節が評価条件を構成する文節か否かを同定する2値分類問題として扱う。本論文では、評価条件を構成する文節を評価条件文節と呼ぶ。分類器の学習にはSVMを用いる。本研究で用いる素性は、レビュー文に存在する評価属性および評価表現との関係に基づく。評価属性と評価表現の抽出は本研究の対象外であるので、既存の手法によって予め抽出しておく。本論文では、小林ら[2]の評価属性と評価表現のタグ付け仕様にしがたい、手動で抽出する。

次に、図1に基づき、モデル構築のために利用する素性の詳細について説明する。図1の例文は、文節ごとに区切られており、文節番号および各文節の係り受け結果が付加されている。文節番号は、後の文節になるにつれて大きくなるように各文節に付加されているとする。さらに、 $\langle \rangle$ と \square 内の語はそれぞれ手動で抽出した評価属性と評価表現である。我々は評価条件抽出のために七つの素性を用い、各文節 s は7次元のベクトル s を持つ。

$$s = (A_s, B_s, C_s, D_s, E_s, F_s, G_s) \quad (1)$$

なお、評価属性と評価表現が含まれる文節は評価条件文節でないという前提で、文節 s から除く。以下、式(1)中の各素性と素性値について説明する。

● 素性 A：評価表現との文節距離

評価表現の後文節には「快適と言える」や「リーズナブルだと思う」といったモダリティ表現が付加されやすく、評価表現よりも後の位置に出現する文節は評価条件になりにくい。つまり、評価条件文節は評価表現よりも前の文節に存在しやすい。また、評価条件文節は評価表現との文節距離が小さい位置に出現しやすい。文節距離とは文節番号の差を表す。

これら評価条件の特徴に基づき、評価表現が含まれる文節を h とすると、文節 s における素性Aの素性値 A_s を以下のように定義する。

$$A_s = \begin{cases} \min_{h \in H} t(h) - t(s) & \text{if } \min_{h \in H} t(h) - t(s) > 0 \\ 10 & \text{otherwise} \end{cases} \quad (2)$$

$t(s), t(h)$ は各々の文節番号を表す。 H は1文中において、評価表現が含まれる文節の集合である。式(2)は文節 s の文節番号が文節 h の文節番号よりも小さければ、その差を素性値とする。なお、評価表現が1文中に複数個含まれている文も存在する。その場合は、 H 内に含まれる文節との文節距離のうち、最小の文節距離を素性値とする。また、文節 h よりも後の位置に出現する文節の素性値は、文節 h との文節距離が大きい文節は評価条件文節になりにくいことを考慮し、文節距離が大きい値を示す10と設

定した。例えば、図1での s ="静かな"における素性値は4となる。

● 素性 B：評価表現との係り受け距離

評価条件文節は評価表現が含まれている文節との係り受け関係が成立しやすい。ここで、本論文における文節 s と文節 s' に係り受け関係があるとは、 $t(s') - t(s) \geq 0$ であり、 s から係り受けの経路をたどって、 s' へ到達可能であるとする。

文節 s における素性値 B_s は、文節 h と係り受け関係があれば、経路内でたどった文節数とする。そうでなければ-1とする。例えば、 s ="静かな"の素性値は4である。

● 素性 C：評価属性に係っていないか

評価属性を含む文節と係り受け関係にある文節は、評価属性を修飾しているだけで、評価条件文節ではない傾向がある。例えば、図1中の「電車の」という文節は、「音」という評価属性を修飾しているだけで、「電車の」は評価条件文節ではない。文節 s における素性値 C_s は、 s と評価属性を含む文節に係り受け関係がなければ1を、そうでなければ0とする。

● 素性 D：特定の助詞もしくは助動詞が文節内に含まれるか

評価条件の末尾には、特定の助詞もしくは助動詞が付きやすい。助詞や助動詞は、動作の行われる対象、相手、時、場所、そして目的を限定したり、条件節をなす用法を持っているからである。そのような助詞や助動詞が文節内に含まれていれば、評価条件文節である可能性が高い。

そこで、素性Dにおける文節 s の素性値 D_s は、 s 内に表1に示す特定の助詞もしくは助動詞が形態素として含まれ、かつ品詞細分類が一致すれば1とし、そうでなければ0とする。

● 素性 E：「ので」の接続助詞が含まれていないか

「ので」という接続助詞は理由や根拠を表す。それを含む文節は評価条件文節ではなく、根拠を述べている部分の文節である傾向にある。文節 s における素性値 E_s は、「ので」が文節 s 内の形態素に含まれていなければ1、そうでなければ0とする。

● 素性 F：特定の助詞が含まれる文節との係り受け関係があるか

評価条件は複数の文節で構成されることがある。また、図1中の「静かな環境を求める方には」という評価条件のように、評価条件文節の各文節は、特定の助詞が含まれる文節との係り受け関係が成立する傾向にある。

そこで、素性値 F_s は文節 s が表1に示す形態素と品詞細分類が含まれる文節と係り受け関係があれば1とし、そうでなければ0とする。なお、 s と s' が同じ文節を指す場合も、到達可能であると考え、特定の助詞もしくは助動詞が含まれる文節の素性値も1とする。

● 素性 G：「ので」の接続助詞が含まれる文節との係り受け関係がないか

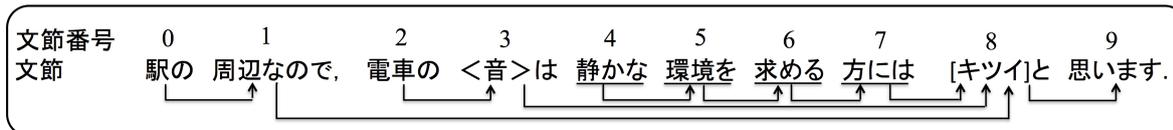


図 1: 各素性値計算の説明で用いる例文

根拠もまた、複数の文節で構成されることがある。つまり、「ので」が含まれる文節と係り受け関係にある文節は、根拠を構成する文節と考えることができる。図 1 中の「駅の」という文節は「ので」が含まれる「周辺なので」の文節と係り受け関係にある。素性値 G_s は文節 s が「ので」が含まれる文節と係り受け関係になれば 1 と、そうでなければ 0 とする。「ので」が含まれる文節の素性値は、0 とする。

表 1: 素性値 D_s と素性値 F_s の計算で用いる特定の助詞と助動詞のリスト

形態素	品詞	品詞細分類
が, に, で, にとつて, として	助詞	格助詞
は, も	助詞	係助詞
の	助詞	連体化
と, だと, ば, にして	助詞	接続助詞
なら, たら	助動詞	*

*は、品詞細分類の一致を考慮しないことを表す

4 評価実験

4.1 実験方法

レビューテキストには楽天と国立情報学研究所が提供している楽天トラベルのレビューデータ (2010 年 7 月 13 日時点で 34 万 8564 件) の一部を用いた。まず、小林らの手法にしたがい、レビューデータ 675 件に対して評価属性と評価表現を抽出する。その後、評価条件のタグ付けを行う。タグ付けの結果、評価条件が 1 個以上含まれているレビューデータは 182 件となった。次に、レビューデータ 182 件から、評価条件が含まれている 259 文を抜き出し、形態素解析と係り受け解析を行い、文節ごとに正例または負例のラベルを付与する。形態素解析と係り受け解析は MeCab¹ と CaboCha² をそれぞれ用いた。解析の結果、正例データ 641 文節と負例データ 1036 文節が得られた。また、評価属性、評価表現のいずれかを含む文節が 564 文節となった。

正例データと負例データ計 1677 文節に対して、文を単位とした 10 分割の交差検定により訓練を行い、分類器を学習する。構築された分類器の入力をテスト事例として、テスト事例が正しく予測されるかを検証する。また、素性ごとの有効性を評価するために、計 128 通りの素性の組み合わせによる実験を行った。ベースラインは、テスト事例に対してランダムにラベルを出力する手法と

¹<http://code.google.com/p/mecab/>

²<http://code.google.com/p/cabocho/>

した。

テスト事例が正しく分類されたかを示す評価尺度として、文節単位での精度、再現率、F 値を用いる。また、評価条件を完全に抽出できたかを示す精度、再現率、F 値を考える。これらは同じラベルが連続して続く文節を単位とした評価値である。

4.2 実験結果と考察

表 2、表 3 に文節単位と評価条件単位で評価したときの実験結果をそれぞれ示す。表中の太字の行は、本手法での 128 通りある素性の組み合わせのうち、F 値の最も高かった素性の組み合わせにおける各評価値を示している。手法が素性名となっている行は、素性一つのみで訓練した場合の各評価値を示している。なお、素性 B, C, E, G のいずれかで訓練した場合は、全てのテスト事例が負例と予測され、各評価値が 0% であった。紙面の都合上、表中にはこの結果を示していない。ベースラインの精度は、テスト事例において負例が多数派であることを考慮し、予測結果を反転させた場合の値を示している。表 2 から、素性 A と素性 D が他の素性よりも有効な素性であることがわかる。

我々は提案手法の誤り分析を行い、誤りの原因を特定した。以下、事例の多かった誤りの原因から順番に説明する。なお、各原因は False Positive か False Negative のどちらか一方を引き起こす。

1. 正例の訓練データに多く含まれる素性ベクトルと類似する負例データがある (False Positive)

訓練データ中の正例には A_s が小さい、かつ A_s と B_s の差が比較的小さい、そして D_s または F_s が 1 である事例が多く含まれていた。「静かな環境を求め方にはおすすめできません。」中の「求める」や「方には」という評価条件文節はその典型的なパターンである。このような正例パターンをもつ事例がテストデータ中の負例データにも含まれていたため、誤分類された。

実験では評価条件を含む文のみを対象に抽出手法を適用した。しかし、レビュー 675 件の中に評価条件が含まれている文は 259 文しかなく、レビュー 1 件あたりでは、負例ラベルが正例ラベルよりも多く含まれている。よって、レビュー 1 件に対して評価条件を抽出することを考えると、誤りが頻発し、精度の低下が懸念されるので、この誤りは優先的に対策を講じるべきである。

2. 素性 A の素性値が大きい (False Negative)

訓練データ中の負例には、 A_s の値が比較的大きい事例が多く含まれていた。このような事例がテス

トデータ中の正例にも含まれていたため、正例が負例であるという誤分類を引き起こした。特に、 A_s が6以上になると、必然的に正例の全てが負例に誤分類された。

3. 評価条件文節が評価属性と係り受け関係にある (False Negative)

図1の「電車の」のように、評価属性の連体修飾語になっている文節は負例が多かった。しかし、「値段の割にしっかりした部屋」のように、「しっかり」という評価表現が「部屋」という評価属性の連体修飾語になっているとき、「値段の」という評価条件文節と「部屋」に係り受け関係があることにより、「値段の」が負例に誤って分類された。

4. 係り受け解析が誤っている (False Negative)

「煙草の臭いが苦手な私には本当にありがたかったです。」では、「煙草の臭いがある」という誤った係り受けが生じた、これにより、 A_s と B_s の差が大きくなり、「煙草の」という評価条件文節を誤って負例と予測した。もし「臭いが」という評価条件文節の次の係り受け先が「苦手な」になっていれば、正しく分類された。

5. 評価条件と評価表現が異なる文に存在する (False Negative)

「JCBホールでのライブに参加するため 宿泊させて頂きました。まさにホールの目の前という理想的な立地でした。」において「理想的」という評価表現は、「マラソン大会参加のため」という目的を表す評価条件で成立する。しかし、各評価条件文節は1文中に評価表現が含まれないため、 A_s が10となり、負例と誤分類された。本手法は評価条件と評価表現が異なる文に存在するパターンに対応できず、改善する必要がある。

6. 助詞が省略されていることによる誤分類 (False Negative)

「三角公園までは わかりやすいですが、そこからホテルまで すこし見つけにくいかもしれません。」は「ホテルまで」と「すこし」の両文節間に助詞が省略されていることにより、 D_s と F_s の値が0となり、誤分類を引き起こした。

7. その他の誤り

形態素解析の誤りによって、特定の助詞を認識できなかったことや文節区切りが異常になり、誤分類を引き起こした事例が見られた。また、評価表現が評価条件の前の文節に存在することで誤った事例も観測された。

5 おわりに

評判は全ての利用者に適合せず、ある条件をもって成立する場合があります。従来のタスクでは考慮されていなかった。本論文ではこの条件を評価条件を定義し、その抽出手法を提案した。実験において、提案手法はいくつかの誤りのパターンが見受けられた。

今後は考察で挙げた誤り1,2に対する方策として、文節距離に依存しない、語に基づく素性を追加する。また、

表 2: 文節単位の実験結果

手法\尺度 [%]	精度	再現率	F 値
ベースライン	61.32	49.85	55.00
{A}	58.45	56.81	56.58
{D}	53.52	64.05	58.05
{F}	23.00	46.95	30.83
{A,B,C,D,F,G}	74.01	77.46	75.57

表 3: 評価条件単位の実験結果

手法\尺度 [%]	精度	再現率	F 値
ベースライン	3.95	7.10	5.07
{A}	13.70	27.51	18.20
{D}	8.53	9.50	8.96
{F}	13.28	17.19	14.88
{A,B,C,D,F,G}	44.63	47.61	46.04

評価条件文節か否かは、前後の文節に依存すると考え、系列を意識した素性を追加することも考える。

謝辞

本研究の一部は、科学研究費補助金基盤研究 (B) (課題番号 22300050) によって実施された。

参考文献

- [1] Niklas Jakob and Iryna Gurevych : Using Anaphora Resolution to Improve Opinion Target Identification in Movie Reviews, *In the Proc. of the ACL 2010 Conference Short Papers*, pp.263–268, 2010.
- [2] 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 NL171-18*, pp.111–118, 2006.
- [3] Wei-Hao Lin and Alexander Hauptmann : Are These Documents Written from Different Perspectives? A Test of Different Perspectives On Statistical Distribution Divergence, *In Proc. of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, pp.1057–1064, 2006.
- [4] Bin Lu : Identifying Opinion Holders and Targets with Dependency Parser in Chinese Texts, *In Proc. of the NAACL HLT 2010 Student Research Workshop*, pp.46–51, 2010.
- [5] Tony Mullen and Robert Malouf : A Preliminary Investigation into Sentiment Analysis of Informal Political Discourse, *In AAAI Symposium on Computational Approaches to Analyzing Weblogs*, pp.159–162, 2006.
- [6] 西川崇哉, 岡田真, 橋本喜代太 : レビュー文書の自動分類におけるテキストの前処理手法の検討, *言語処理学会第18回発表論文集*, pp.517–520, 2012.
- [7] Wayne Xin Zhao, Jing Jiang, Hongfei Yan and Xiaoming Li : Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid, *In the Proc. of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp.56–65, 2010.