

商品レビューの複数の観点からの有用性の評価

曾田 颯人

北陸先端科学技術大学院大学
s1910131@jaist.ac.jp

白井 清昭

北陸先端科学技術大学院大学
kshirai@jaist.ac.jp

1 はじめに

ユーザが商品の購入を検討する際、他のユーザによって書かれた商品のレビューを参考に行われることは日常的に行われている。しかし、レビューの中には、ユーザにとって役に立つものもあればそうでないものもある。膨大な量のレビューの中から有用なレビューを見つけるのは多大な労力を要するため、これを効率よく見つける技術が求められる。このような背景の下、レビューの有用性を自動判定する研究が行われてきた。しかし、先行研究の多くは単にレビューが有用であるか否かを判定するが、どのようなレビューが有用かはユーザによって異なると考えられる。例えば、商品を使用した体験が役に立つと考えるユーザもいれば、他の商品との比較を重視するユーザもいるだろう。

本研究では、商品レビューの有用性を複数の観点から評価し、その評価結果を包括的にユーザに提示するシステムを提案する [1]。様々な観点による有用性の評価結果を閲覧できるようにすることで、ユーザが自身の嗜好に合わせて有用な商品レビューを見つける作業をサポートすることを狙う。

2 関連研究

山澤らは、Amazon¹⁾のレビューを対象に、ユーザが内容を信用して利用できる文(有用文)を判別する Support Vector Machine (SVM) を学習した [2]。佐々木らは、人手による有用なレビューの分析結果を基に、機械学習を用いて有用なレビューを判別する手法を提案した [3]。Fan らは、製品のメタデータ(タイトル、ブランドなど)とレビューテキストの両方を入力とし、レビューが有用か否かを分類するディープニューラルネットワークを提案した [4]。Vo らは、Rodak ら [5] が提案したレビューの特徴を素性として使用し、Elastic Net 回帰を使用してレビューの有用性を予測し、レビューをランク付けし、その

上位のレビューをユーザに提示するシステムを提案した [6]。

これらの先行研究は、レビューが有用か否かのみを判定しているのに対し、本研究では、有用性の観点をいくつか定義し、個々の観点毎にレビューの有用性を判定することで、ユーザが多角的な視点から有用なレビューを見つけることを支援する。

3 提案手法

3.1 有用性の観点

本研究では、レビューの有用性を判定する先行研究 [3, 6, 7] の知見や著者らによる経験などを踏まえ、以下の7つの観点からレビューの有用性を評価することを提案する。

観点1 評価表現に対する根拠がある

単に商品の善し悪しを評価するだけでなく、その善し悪しの根拠となる事実もあわせて書いているとき、そのレビューの有用性は高いと評価する。図1の例では、「便利です」という評価の根拠として「大きさもちょうど良い」を示している。

観点2 商品に関係のある言及が多い

レビューの中には、商品に言及している文もあれば、ショップの対応や配送業者の対応など、商品以外のことに言及している文もある。商品に言及している文が多ければ多いほど有用性は高いと評価する。図1の例では、「発送までが遅かったのが残念でした」という文は商品に言及していない。

観点3 他の商品と比較している

レビューが評価対象の商品と他の商品と比較しているとき、ユーザにとって参考になる情報を提供している可能性が高いため、有用性が高いと評価する。図1の例では、「商品Bに比べて軽そう」といった比較がある。

観点4 実際に商品を使用した(あるいはしていない)と推測できる

1) <https://www.amazon.co.jp/>

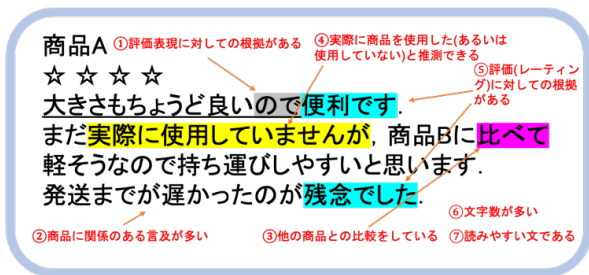


図1 レビューの例

実際に商品を使用していない人のレビューは有用性が低いと評価する。図1の例では、まだ実際に使用していませんが、という句は、ユーザが商品を実際には使っていないことを表し、このレビューの信頼性が低くなる。

観点5 評価(レーティング)に対する根拠がある

多くのECサイトや口コミサイトでは、ユーザは商品に対して星の数などでレーティングをつけることができるが、そのレーティングの根拠となる文が示されているとき、有用性は高いと評価する。図1の例では、星の数は4と高く、「便利です」といった評価表現はその理由の説明とみなせる。

観点6 分量が多い

ごく簡単に感想を書いた短い一文のレビューより、商品に対する意見・感想を詳細に書いた長いレビューの方が、有用性が高いと評価する。

観点7 文章が読みやすい

読みやすい文章で書かれたレビューほど有用性が高いと評価する。

以下、観点1、観点2、観点3による有用性の評価手法については3.2、3.3、3.4項で、これら以外の観点による有用性の評価については3.5項で述べる。

3.2 根拠を含む文の検出

本項では、レビュー文が根拠を含むか否かを判定する手法について述べる。基本的な考えとしては、根拠を示す接続詞で終わる文節(例は図1の「良いので」)が評価表現を含む文節(例は図1の「便利です」)に係るとき、そのレビュー文は根拠を含むと判定する。具体的には、以下のいずれかの条件を見たす文を根拠を含む文として検出する。

条件1 連用形 → 評価表現

用言の連用形で終わる文節が評価表現を含む文節に係る。

(例) 片手でも持ちやすくて 実用的です²⁾

2) 点線の文節が実線の文節に係ることを表す。以降の例でも

のが、ので、のは、為、ため、点が、くて、
のも、ところが、ところも

図2 根拠を示す接続詞

条件2 根拠を表す接続詞 → 用言

根拠を示す接続詞で終わる文節が用言を含む文節に係る。

(例) 耐熱性能も良いのでライフも期待できる

上記を踏まえ、以下の手続きでレビュー文に根拠が含まれるかを判定する。

1. CaboCha[8] を用いて文節の係り受け解析を行う。
2. 評価表現もしくは用言(動詞、形容詞、形容動詞)を含む文節Eを検出する。評価表現は以下の2通りの方法で検出する。
 - ・日本語評価極性辞書(用言編)[9]に含まれる評価表現
 - ・感情分析APIであるCOTOHA API[10]によって検出された評価表現
3. 文節Eに直接係る文節Rを検出する。文節Rが根拠を示す接続詞(図2)を含むかどうかを判定する。
4. 文節RとEが上記の条件1、条件2のいずれかを満たすとき、根拠を含むと判定する。それ以外は含まないと判定する。

最終的なシステムでは、上記の手法で根拠を含む文の検出を試み、レビューに根拠を含む文が出現するか否かをユーザに提示する。また、含む場合には検出された文を強調して提示する。

3.3 商品への言及度の判定

本項では、商品レビューが評価対象の商品について言及している度合を「商品言及度」(以下、単に言及度と記す)と定義し、これを推定する手法について述べる。言及度はユーザレビューの有用性を表す指標の1つとしてユーザに提示することを想定する。

本研究では、評価対象の商品のカテゴリに固有の単語がどれだけ多く含まれるかで言及度を定量化する。例えば、商品カテゴリがペットのとき、「大型犬」「フード」「匂い」「食いつき」などはペットに関連する商品に言及するときによく使われる単語であり、またこのような単語が多いレビューほど商品に言及していると言える。

同様に表記する。

まず、商品カテゴリ毎に単語の重要度を算出する。商品カテゴリ c におけるキーワード k の重要度 $Sig(k, c)$ は、TF-IDF の考え方に基づき、式 (1) のように定義する。

$$Sig(k, c) = tf_{k,c} \cdot \log \frac{N_c + 1}{cf_k} \quad (1)$$

$tf_{k,c}$ は商品カテゴリ c のレビュー集合に含まれるキーワード k の相対出現頻度、 cf_k (category frequency) は k が出現するカテゴリの数、 N_c はカテゴリの総数である。重要度の算出は楽天データ [11] における楽天市場の商品説明文の集合を用いる。商品カテゴリは「ビューティ」「文房具」などの 18 種とする。重要度を算出するキーワードは名詞もしくは複合名詞に限定する。

与えられたレビュー r の言及度を式 (2) のように定義する。

$$Ex(r, c) = w \cdot \sum_{k \in K_r} Sig(k, c) + (1 - w) \cdot \log len(r) \quad (2)$$

言及度は、レビュー内の全てのキーワード (K_r) の重要度の和と、レビューの長さ (文字数) $len(r)$ の重みつき和とする。 c は、レビュー r が評価対象とする商品のカテゴリである。 $len(r)$ を考慮したのは、長いレビューほど商品への言及度が高いという直観に基づく。重み w は実験的に決定する。

3.4 比較の有無の判定

本項では、簡単なルールベースの手法で複数の商品を比較する文を検出する手法について述べる。Amazon の家電とホームカテゴリの商品レビューを 1,201 件取得し、複数の製品を比較しているレビューを 176 件見つけた。これらのレビューから、他の商品との比較があるレビューを検出するルールを策定した。ルールは以下の 3 つに大別される。

- 比較を表すキーワード + 評価表現

複数の商品の比較を示唆するキーワードを含み、そのキーワードの後に評価表現が出現するとき、そのレビューは比較ありと判定する。「くらべる」「違う」「反面」など、34 個のキーワードからなるルールセットを作成した。

- レビューの最初の文で比較を示唆する表現

「〇〇から買い換えました」などの表現がレビューの最初の文に出現するとき、レビュー全体で他の商品との比較があると判定する。「買い換え」「以前」など 6 つのキーワードを設定

し、それが最初の文に出現するときに比較ありと判定するルールセットを作成した。

- 「より」を含む文

以下の例のように、「名詞 + より」という文節が評価表現を含む文節に直接係るとき、比較を含むレビューと判定する。

(例) 他の製品より 良い と思います。

最終的なシステムでは、「他の商品と比較している」という観点からの有用性を示すために、レビューの中に比較があるか否かを表示することを想定している。ここでのルールは、比較を含むレビューの検出について、再現率よりも精度を重視して設計している。

3.5 その他の観点からの有用性の評価

これまで述べてきた 3 つの観点以外については、現時点では評価方法を検討している段階である。

観点 4 「実際に商品を使用したと推測できる」については、予備調査として、Amazon に投稿された本、電子機器、ホーム・キッチンカテゴリのレビューを 1,385 件取得し、商品を使用していないと思われるレビューを手で判定した。その結果、該当するレビューはわずか 11 件であった。実際に商品を使用していないレビューの数が非常に少ないことから、これを自動検出する手法の検討は行わなかった。とはいえ、別の商品カテゴリのレビューや Amazon 以外のサイトに投稿されたレビューでは商品を使用していないレビューが多く存在する可能性もあり、今後も調査を続けたい。

観点 5 「評価 (レーティング) に対する根拠がある」については、レビューの極性判定を行い、その極性がレーティングに近いかどうか、また極性判定の信頼性が十分高いかどうか、などの手法によって評価する予定である。

観点 6 「分量が多い」については、単語数や文字数などで比較的容易に定量化できると考えられる。

観点 7 「文章がよみやすい」については、文書の平易さや読みやすさを判定する先行研究があり、これらの技術を適用することが考えられる。

4 評価実験

本節では観点 1、観点 2、観点 3 で有用性を判定する提案手法の評価実験について述べる。

4.1 根拠を含む文の検出手法の評価

Amazon に LED 電球, ルンバ e5, インパクトドライバ, 洗顔料について投稿されたレビューをランダムに 50 件取得した。これらのレビューに対し、根拠を含む評価文が現われるかを著者 2 名が独立に判定した。提案手法は根拠を含む文を抽出する手法であるので、商品に対する評価とその根拠が複数の文に書いてある場合は除外し、ひとつの文に書かれているかを判定した。二者の判定の一致率は 0.76, κ 係数は 0.56 であった。

この評価データに対し、3.2 項で述べた手法で根拠を含む文を検出し、その精度、再現率、F 値を調べた。また、根拠の有無の判定が一致する割合 (正解率) も算出した。結果を表 1 に示す。再現率は比較的高いが、特に評価者 2 に対する精度が低く、F 値も 0.7 程度に留まっている。

表 1 根拠を含む文の検出の評価結果

	精度	再現率	F 値	正解率
評価者 1	0.61	0.80	0.68	0.64
評価者 2	0.45	0.75	0.57	0.54

4.2 商品への言及度の判定の評価

言及度の計算に用いた 18 のカテゴリについて、それぞれのカテゴリに投稿されたレビューをランダムに選択し、レビューの組を 10 件ずつ、計 180 件作成した。これらに対し、著者 2 名が独立に、どちらのレビューがより商品に言及しているかを判定した。このとき、どちらの言及度が大きいかが判定が難しい場合には「不明」とした。最終的に、評価者のどちらかが不明と判定した 24 件のデータを除く 156 件を評価データとした。二者の判定の一致率は 0.904, κ 係数は 0.807 であった。この評価データに対し、それぞれのレビューの組に対して 3.3 項で述べた手法でレビューの言及度を算出し、言及度の大きいレビューが人手による判定の結果とどれだけ一致するか (正解率) を調べた。ベースラインとして、常に文字数の多いレビューを言及度の大きいレビューとして選択する手法 (式 (2) の w を 0 とした場合) と比較した。図 3 は、重み w を 0.990 から 1 まで 10^{-6} 刻みで変動させたときの正解率の変化を示している。

ベースライン (図 3 の直線) の正解率は、評価者 1

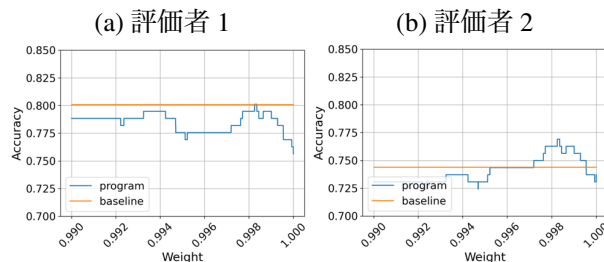


図 3 商品への言及度の評価結果

は 0.801, 評価者 2 は 0.740 であった。提案手法については、評価者 1, 2 とともに $w = 0.9983$ のときに正解率が最大となり、その値はそれぞれ 0.801, 0.769 であった。提案手法がベースラインと同等もしくは上回ることを確認した。ただし、パラメタ w の最適化は本来はテストデータとは別の開発データで行うべきである。開発データを用いた w の最適化は今後の課題である。

4.3 比較の有無の判定の評価

3.4 項で説明した個々のルールについて、それを用いて比較を含むレビューを検出したときの精度を測った。ルールの設計に用いた家電・ホームカテゴリのレビューに対する精度 (クローズド) と、新たに取得した「ベビー」カテゴリのレビュー 557 件に対する精度 (オープン) を調べた。いくつかのルールについて、検出精度が十分に高いことを確認した。例を表 2 に示す。

表 2 比較を含むレビューを検出するルールの精度

ルール	クローズド	オープン
優れて	0.86	1.00
比べる	0.77	0.82
他のメーカー	0.71	0.67
(先頭) 買い替え	0.56	1.00
「より」	0.70	0.50

5 おわりに

今後の課題として、これまで着手していない 4 つ観点からの有用性評価を実現することが挙げられる。一方、3 つの観点からの有用性の判定について、本論文で提案した手法は比較的基本的なものであり、これらをより洗練することも必要である。また、7 つの観点からの有用性の評価結果をどのようにユーザに提示し、ユーザのレビュー検索を補助するシステムを実装することも重要な課題である。

参考文献

- [1]曾田颯人. 商品レビューの複数の観点からの有用性の評価. 修士論文, 北陸先端科学技術大学院大学, 3 2021.
- [2]山澤美由起, 吉村宏樹, 増市博. Amazon レビュー文の有用性判別実験. 情報処理学会研究報告自然言語処理 (NL), Vol. 2006, No. 53 (2006-NL-173), pp. 15–20, 2006.
- [3]佐々木優衣, 関洋平. 商品レビューを対象とした有用性の定義と判別. 第 6 回データ工学と情報マネジメントに関するフォーラム (DEIM2014) B5-1, 2014.
- [4]Miao Fan, Chao Feng, Lin Guo, Mingming Sun, and Ping Li. Product-aware helpfulness prediction of on-line reviews. In *Proceedings of the World Wide Web Conference*, pp. 2715–2721, 2019.
- [5]J. Rodak, M. Xiao, and L. Longoria. Predicting helpfulness ratings of amazon product reviews. Technical report, Stanford University,, 2012.
- [6]Chau Vo, Dung Duang, Duy Nguyen, and Tru Cao. From helpfulness prediction to helpful review retrieval for online product reviews. In *Proceedings of the Ninth International Symposium on Information and Communication Technology (SolCT 2018)*, pp. 38–45, 2018.
- [7]木浪貴博, 小林亜樹. 商品レビュー文における主観的表現と有用性に関する検討. 情報処理学会第 77 回全国大会講演論文集, pp. 159–160, 2015.
- [8]工藤拓, 松本裕治. チャンキングの段階適用による日本語係り受け解析. 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834–1842, 2002.
- [9]小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol. 12, No. 3, pp. 203–222, 2005.
- [10]NTT コミュニケーションズ. COTOHA API. <https://api.ce-cotoha.com/contents/index.html>. (2020 年 12 月閲覧).
- [11]Rakuten Institute of Technology. 楽天データ公開. https://rit.rakuten.co.jp/data_release_ja/. (2020 年 12 月閲覧).