

インパクトの視点に基づくWEB上のユーザレビューの分析

安藤 まや 石崎 俊
慶應義塾大学 政策・メディア研究科

1. はじめに

これまでの評判分析に関する研究では、書き手の視点からの問題設定が行われてきた。対象に関する書き手の評価をレビュー文章から推定する問題や[B. Pang 2002][Kudo 2004][岡野原 2007]、書き手が評価対象のどのような特徴や属性に注目しているかを同定、分類する問題[飯田 2005]である。しかしながら、評判分析の応用として重要なのは、読み手の視点である。レビューを読んだ人が、その影響を受けて商品を購入したいと思うかどうか、店舗そして読み手の双方にとって要となるレビューの存在意義である。読み手の心に響くインパクトがあるレビューこそが重要なレビューであり、それは書き手の絶対的な評価との必然的な関係はない。書き手の視点と読み手の視点の関係に関しては、読み手によるレビューの有効性という指標が存在し、それを利用した研究も行われている[Ghose 2007]。しかし、この情報は入手困難であり、詳細な分析に耐えられるほどの情報量が含まれていない。

本研究では楽天から公開された宿泊施設のユーザレビューを被験者に読んでもらい、その宿泊施設に行きたくなくなる程のインパクトを受けた文を抽出してもらうことによって読み手の視点からの評判分析を目指す。インパクト文について3つの観点から分析を行う。

- ・書き手の判定と読み手の判定の違いの分析 (第4章)
- ・「判断対象」と「判断」の観点からの分析 (第5章)
- ・個人的嗜好の分析 (第6章)

2. 先行研究

主観的な評判情報を positive か negative に分類する研究がこの10年近く盛んに行われてきている[Kudo 2004][B. Pang 2002]。最近では、2分類だけではなく複数の段階の評判分類を行っている研究もある。[岡野原 2007]は、Amazonの本のレビュー(英語)に対し、5段階の評価分類を機械学習の手法を使って実現している。本研究では、評判分析を2分類や5段階といった書き手の視点からの分類問題としてではなく、読み手の心に響くインパクトのある表現を分析することを目的としている。書き手が高評価だと判定したレビューが読み手にとっても高評価なレビューとなるのではなく、読み手にとって重要な情報にのみ重きを置いているところが本研究の特徴である。

また、読み手が下したレビューに対する評価を用いた研究が行われてきている。[Ghose 2007]はAmazonのレビューに対し、読み手がレビューに対し「helpfulness」であると認めた数をもとに、レビューの有効性のランキングを予想するメ

カニズムを提案している。ただし、この研究の目的は、「helpfulness」の値を統計処理の目的値として使用しているだけであり、書き手と読み手の評価の違い、どのような評価対象が読み手にとって「helpful」であったのか、「helpfulness」の個人の嗜好といった詳細な言語的分析は行われていない。

3. データの作成

インパクトの分析をするためには、インパクトのある文を収集する必要がある。しかし、インパクトのある文の数は少なく、すべてのレビューを読んで探すのは難しい。そのため、楽天トラベルのレビューから「！」が文末に2つ以上連続して使われている文をランダムに500文抽出する。「！」はその直前にある文の意味を変えない表現であり、インパクト文と多く共起する表現であると考えられる。「！」を含む文を分析対象とすることで、インパクトのある文を比較的容易に収集できると考えた。

この500文から被験者20名にインパクトがある文を選んでもらう。被験者は、20代から60代までの、各世代の男性2名、女性2名で、合計20名である。作業目的は「良くも悪くもユーザの心をぐっと掴むような、極端なレビューを集めること」とする。「インパクトのある表現」を「この宿泊施設に是非とも行ってみたいと思うもの」と「この宿泊施設に絶対に行きたくないと思うもの」と説明し、個人的な嗜好をできるだけ排除し、一般的にインパクトがあると考えられる文を選んでもらった。被験者全員が、インパクト有と判定した文はなかったが、5人以上がインパクト有と判定84文を、本稿では「インパクト有」とし、それ以外を「インパクト無」とする。

4. 書き手の判定と読み手の判定の違いの分析

書き手と読み手の感覚にはどのような違いがあるのだろうか。その分析のために、書き手(宿泊施設利用者)が付与した「総合評価」の星の数を利用して、書き手の判定と読み手の判定の違いを調査した。図1はすべてのレビューの星ごとの分布を示し、図2はインパクト有と判定された84文が含まれるレビューの星ごとの分布を示している。

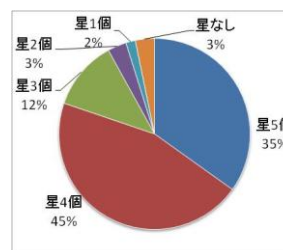


図1 すべてのレビューの星ごとの分布

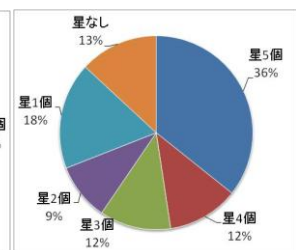


図2 インパクト有の星ごとの分布

直感的には極端な評価を示す星 5 個（最高評価）と星 1 個（最低評価）は、読み手がインパクトを受けるようなレビューが多く含まれると考えられることから、インパクト文での割合が高くなり、星 2~4 個は、極端な評価ではないため、割合はそれほど高くはないと考えられる。実際には、星 1 個の割合はすべてのレビューの中では 2% 足らずだったのが、インパクト有では 18% を占めている一方、星 5 個の割合には全く変化がなかった。星 2~4 個の割合はいずれも 10% 程度あり、直感よりも読み手に影響を与えることがわかった。ポジティブとネガティブなレビューを比較すると、前述したように星 1 個の割合が増えているのに加えて、星 4 個は割合を 45% から 12% と大幅に減らし、星 2 個は 3% から 9% に増えた。これらのことから、ネガティブであることがよりインパクトを与えている場合が多いことがわかった。インパクト判定を自動で行う際には、高評価のレビューと低評価のレビューを平等に扱わず、適切な差をつける必要がある。

「総合評価」の星 2~4 個のレビュー内容を確認すると、良い評価と悪い評価が混在しているのが確認された（例文 (1)(2)）。インパクト判定に使われた文を【】で囲んで示す。

例文 (1) (宿泊施設名)さんのベッドの寝心地が大好きで、…(中略) …良く利用させてもらっています♪…(中略) …【テーブルの上はホコリまみれ、カーテンを開けると蜘蛛の巣の固まりが…2泊予定だったので、次の日の清掃に期待しましたが残念ながら期待はずれ、ベッド横のサイドテーブルは手に白つくほどのホコリ!!】…(中略) …また機会があれば利用させてくださいね♪ (立地 4: 部屋 2: 食事 0: 風呂 3: サービス 3: 設備・アメニティ 3: 総合 3)

例文(1)はレビューの最初と最後にポジティブな記述が、その間に「部屋」に関するネガティブな記述がある。書き手は「総合評価」に星 3 個、「部屋」に星 2 個を付与しているが、被験者の 20 人中 18 人がインパクト有と判定するほどの内容となっている。

例文(2) 赤ちゃんプランで宿泊させてもらいました。いろいろなもの（ベビーカーやパジャマなど）をレンタルできるのでとても助かりました。スタッフの方の対応も細やかでとても素晴らしかったです。【朝食のバイキングも沖縄ならではの料理が多数あって、とても美味しかったし大満足でした!!】しかし、私たちが泊まった部屋は道路沿いで、…(中略) …しっかりと寝ることができませんでした。…(中略) …本当に騒音がひどかったです。赤ちゃんプランの人にはこの部屋は向かないと思います。…(中略) …このことがとても大きかったので 2 点にしましたが、スタッフの方に対しては大変満足 5 点満点です!! (立地 4: 部屋 1: 食事 4: 風呂 3: サービス 5: 設備・アメニティ 5: 総合 2)

例文(2)は、総合評価では星 2 個をつけているが、レビュー内には「食事」に関するポジティブ記述と「部屋」に関するネガティブ記述が含まれている。「食事」は、星 4 個の評価だが、20 人中 7 人がインパクト有と判定している。

全体的な統計結果と、上記に挙げたような例を見ると、読

み手と書き手との判定の差が確認できる。レビューのポジティブ、ネガティブの判定のみならず、読み手にとってインパクトがあるレビューが何かを見つけることは、非常に重要な課題である。

今回のインパクトの調査は文単位で行っている。これは書き手が付与した星がレビュー単位で判定していることと食い違っているが、読み手は全体評価ではなく具体的な事柄が書かれている文の単位でインパクトを判断している場合が多いと考えられるため、このような調査方法に有効性があると考えている。今後、さらなる分析を行う予定である。

5. 「判断対象」と「判断」の観点からの分析

読み手が、判断材料に使うようなインパクト文とはどのようなものなのかを調査するために、被験者がインパクト有と判定した文を、宿泊施設利用者が取り上げたトピックである「判断対象」と、それについてどう感じたのかという「判断」の 2 つの側面から分析する。分析は筆者の手作業で行った。

「判断対象」について分析を行った結果、13 に分類された（表 2）。また、「判断」については、4 つの大きなグループと 9 つのより細かいグループに分類できることがわかった（表 3）。2 つ表には、分類項目ごとにインパクト有と無が占める割合に有意な差があるかを示すためのカイ 2 乗検定の結果を提示し、インパクトの有無の分析に使用する。

双方の分類では重複を認めている。例えば、1 文の中に「判断」の項目が複数含まれている場合は、すべての「判断」を抽出する。「浴槽は十分すぎる広さで大満足でした!!」という文では、【浴槽も十分すぎる広さ】から「事象（描写）」、【大満足】から「気分」の両方の頻度を数えている。

5. 1 宿泊施設利用者の「判断対象」の分類

宿泊施設利用者がどのようなことについてレビューを書いたのかを示す「判断対象」について分類した。楽天トラベルで星の項目となっている「立地」「部屋」「食事」「風呂」「サービス」「設備・アメニティ」の他に、筆者が独自に設定した「スタッフ」「プラン」「ホテル」「価格」「景色」を追加した。インパクト有と無でカイ 2 乗検定を行った結果、1% 以下で有意差が認められたものが 3 つ、5% 以下のものが 1 つ見られた（表 2）。

表 2 宿泊施設利用者の「判断対象」の有意差に関するカイ 2 乗検定結果

1%以下	部屋、食事、サービス
5%以下	景色
なし	立地、風呂、設備、スタッフ、プラン、ホテル、価格

「!」を 2 つ以上含む 500 文中で、多く取りあげられた「判断対象」は多い順に「食事(81 文)」「サービス(75)」「風呂(55)」「価格(50)」「部屋(46)」「設備・アメニティ(46)」であった。一方、特に読み手が着目しているのは、カイ 2 乗検定で 1%

以下の有意差が認められる「部屋」「食事」「サービス」と、5%以下で差が認められる「景色(21)」であることがわかった。読み手が影響を受ける「判断対象」には偏りがあり、書き手が書いた「判断対象」を抽出することがそのまま読み手にとって有用な情報になるわけではないことがわかった。

5. 2 宿泊施設利用者の「判断」の分類

宿泊施設利用者の感覚を表す「判断」について、インパクト有と無でカイ2乗検定を行った結果、表3にあるように「五感」と「事象（描写）」に対して1%以下で有意差が認められた。それ以外の項目もこれまでの評判分析で重要視されてきたと考えられるが、今回の調査でそれだけでは読み手にとって重要な情報とはなりえないことがわかった。

5. 3 2つの側面からの分析

「判断対象」と「判断」の組み合わせに対して分析を行う。今回、インパクト有の中で「判断対象」と「判断」の頻度が5以上の場合を中心に分析を進めることとする。表4に「判断」と「判断対象」が共起した場合のカイ2乗検定の結果を示す。「-」は、頻度が5未満を示し、「なし」は有意性がないことを示す。以降、「食事+五感」「サービス+気持ち」「部屋／食事／サービス／景色+事象（描写）」について述べる。

【食事+五感】

「五感」と「食事」の組み合わせについて分析する。「五感」に含まれるインパクト文は「嗅覚」を表現している1文を除いて「味覚」に関するものであった。統計的にも、「食事」と「味覚」が組合わさっている文数はインパクト有で84文中

12文、インパクト無で416文中19文となっており、有意水準1%以下で差が見られる。さらに詳細に分析する。

例文(3) 朝食、おいしいですね！！

例文(4) 珈琲&トースト、ボイルドエッグ、サラダ、フルーツヨーグルトのシンプルな朝食メニューながら、これがかかりうま〜い！！

例文(4)のように「トースト」「サラダ」など1つの器に入れて出せるものを「料理名」とした場合、「料理名」が含まれている文数はインパクト有では6文、インパクト無では5文で、「料理名」が含まれている場合には有意水準1%以下で差が見られる。また、「料理の種類が多い」という意の記述が、インパクト有で3文、インパクト無で0文あり、「料理の種類が多い」ことも読み手に影響を与えられられる。

【サービス+気持ち】

「満足」「嬉しい」といった宿泊施設の利用者の気持ちを表す「判断」は「判断対象」が「サービス」の時にのみ有意水準2.5%以下で差が見られた。「サービス」には「ウェルカムフルーツ」といった「宿泊施設のサービス内容」と例文(5)のように「宿泊施設の人間の行為」が含まれる。「気持ち」は、単体で有意性が認められた「部屋」とも頻度6で共起していたが、ここでは優位性は認められなかった。「気持ち」は、「サービス」と共起するときのみ、読み手に影響を与えることがわかった。

例文(5) 予約時に彼との記念日だとお伝えした所、当日夕食時にサプライズで嬉しいサービスがあり、寿のお箸まで用意してくださいました！！

表3 宿泊施設利用者の「判断」の分類

要因		説明	例	インパクト有 (84文)	インパクト無 (416文)	カイ2乗
宿泊者の感覚	気持ち	利用者の感情を表すもの	楽しい、腹が立つ	19.1% (16/84)	19.0% (79/416)	なし
	五感	人間の五感に関するもの	おいしい、臭い	15.5% (13/84)	5.1% (21/416)	1%以下
対象物の評価	事象 (描写)	利用者が対象について判定しているもの	最高、最低、素晴らしい、部屋が広い、食事が多い	81.0% (68/84)	51.0% (212/416)	1%以下
読み手・宿泊施設へのメッセージ	再来	その宿泊施設にまた来る、もう来ないといったこの先の書き手の行動・意志を表すもの	また来ます、もう使いません	6.0% (5/84)	11.5% (48/416)	なし
	要望	宿泊施設に対するリクエスト	してほしい	1.2% (2/84)	3.9% (16/416)	なし
	推薦	人に勧める意の表現	お勧め	1.2% (1/84)	1.7% (7/416)	なし
その他		上記に含まれないもの	ありがとうございます。	5.5% (4/84)	25.5% (90/416)	-

表4 「判断」と「判断対象」の共起

	部屋	食事	風呂	サービス	設備	価格	景色
事象 (描写)	1%以下	1%以下	なし	1%以下	なし	なし	5%以下
気持ち	なし	-	-	2.5%以下	-	-	-
五感	-	1%以下	-	-	-	-	-

【部屋／食事／サービス／景色＋事象（描写）】

「事象（描写）」はインパクト有となった「判断」の中でも最も頻度の高い項目である。その中でさらに「判断対象」の「部屋」「食事」「サービス」と共起すると有意水準 1%以下で差が見られる。読み手はレビューの中で具体的な事象とこの3つの対象が結びついた時によりインパクトを受けることが確認できた。また、楽天が設定した星の項目以外にあらたに「景色」が重要な要素であることがわかった（例文(6)）。

例文(6) 部屋いっぱいの大きな窓の外に広がる光輝く素晴らしい夜景を呑むほど美しく、圧巻でした！！

6. 個人の嗜好の分析

今回の実験では、被験者の判定にばらつきが見られた。少なくとも 1 人以上の被験者がインパクト有と判定した文は 500 文中 289 文で、そのうち 1 人のみがインパクト有と判定したものは 94 文であった。一方、10 人以上がインパクト有と判定したものは 24 文で、非常に少ないため、これらの状況を明らかにするために、個人の嗜好について調査を行った。

各被験者が判定したインパクト文の中に「判断対象」の項目がどのくらいの割合かを調査した。「立地」「部屋」「食事」「風呂」「サービス」「設備・アメニティ」「価格」「景色」の 8 つの項目で調査したところ、個人の嗜好が 4 つのグループに分類できることがわかった（図 3～6）。

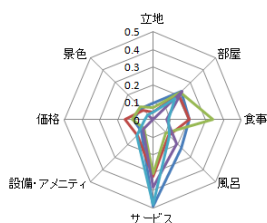


図 3 サービス重視型

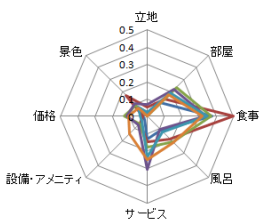


図 4 食事重視型

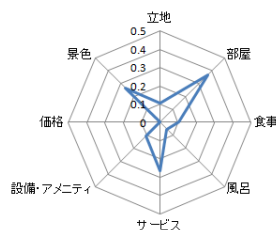


図 5 部屋・景色重視型

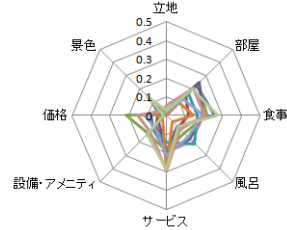


図 6 バランス型

「サービス（図 3）」「食事（図 4）」「部屋・景色（図 5）」にそれぞれ格別の興味を示すタイプと、特定の項目に突出しない判定をした被験者がいることがわかる（図 6）。被験者 5 人以上がインパクト有と判定したもののうち、「サービス」「部屋」「食事」「景色」について、有意な差が認められたが、すべてのユーザがこれらを選んでいるわけではなく、一部のユーザが積極的に選んでいることがわかった。これらは嗜好によって読み手がいくつかのタイプに分けられることを示しており、読み手の感覚により近い情報を提示する際に重要な役割を果たすと考えている。

将来的にインパクトのあるレビューを抽出し、ユーザにとって有効なレビューを提示するシステムの構築を考えると、例えば、10 人以上の被験者がインパクト有と判定した 24 文だけでは充分ではない。各ユーザの嗜好を考慮した内容を多く出す等の工夫をすることによって、それぞれのユーザにとって有効なレビューを提示することが可能となると考えられる。

7. まとめ

ユーザレビューの分析に関する研究では、従来、ポジティブかネガティブかなどの書き手の視点から行う評判分析が主であった。本研究では、「読み手の心に響く表現」を「インパクトのある表現」と定義し、読み手の視点から評判分析を行った。レビューに付与された星の分布を調査することで、書き手と読み手の判定の違いが明らかになった。また、レビューの分類項目のすべてが読み手に影響を与えるのではなく、「料理名＋味覚」「サービス＋気持ち」「部屋／食事／サービス／景色＋事象（描写）」が読み手に影響を与えることがわかった。さらに個人の嗜好の存在を確認し、その分析を行った。

謝辞

本研究を進めるにあたり、データを提供してくださいました楽天技術研究所に心より御礼申し上げます。

参考文献

- [1] 飯田龍, 小林 のぞみ, 乾 健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出を目的とした機械学習による属性・評価値対同定. 情報処理学会研究報告. 自然言語処理研究会報告 2005(1), pp.21-28, 2005-01-11.
- [2] 岡野原 大輔, 辻井 潤一. レビューに対する評価指標の自動付与. 自然言語処理. 14(3). pp. 273-295, 2007.
- [3] 高村 大也, 乾 孝司, 奥村 学. 極性反転に対応した評価表現モデル(言語の統計モデル). 情報処理学会研究報告. 自然言語処理研究会報告, 2005(73), pp.141-148, 2005.
- [4] Kudo, T. and Matsumoto, Y. (2004). "A boosting algorithm for classification of semi-structured text." In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 301-308.
- [5] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques". Proc. of the Conference on Empirical Methods in Natural Language Processing(EMNLP), pp.79-86, 2002.
- [7] Anindya Ghose and Panagiotis G. Ipeirotis. Designing novel review ranking systems: Predicting usefulness and impact of reviews. In Proceedings of the International Conference on Electronic Commerce (ICEC), pp303-309, 2007. Invited paper.