

読み手の視点からの宿泊施設レビューの分析 —インパクトを与える要因の分析と抽出法—

安藤 まや

1. はじめに

インターネット上には、商品について顧客が書いた感想や評価が大量のレビューとして存在し、それがレビューの読み手による購買行動の意思決定に影響を与えることが注目されている。自然言語処理やデータマイニングにおける従来の評判分析の研究では、書き手が書いた内容を対象とした問題設定がされてきた。しかし、レビューの利用形態を考えた時、顧客候補である読み手が重視しているレビューの情報を把握することが重要となる。本研究では宿泊施設に関するレビューを使用して、レビュー中の文を単位として、宿泊施設を利用するという意思決定に重要な影響を与える場合を「インパクト文」と定義する。そのようなインパクト文を詳細に分析し、その特徴を統計的な有意性をもって示した。また、将来の自動的な処理を念頭に、インパクト文に多く含まれる名詞およびその特性の分析を行った。まず、大規模なコーパスを用いて名詞の上位・下位関係の自動抽出を行い、高い正解率を得た。そして、名詞の上位・下位関係をインパクト文の自動抽出に応用する見通しを示した。また、連想実験によって名詞の特性を抽出し、上位・下位関係にある名詞の典型性を定式化し、定量的に有効性を示した。

2. インパクト文の特徴分析

宿泊施設に関するレビューから 3,000 文をランダムに抽出し、「その宿泊施設に行きたくなる」もしくは「行きたくなくなる」ほどの強いインパクトを受けた文を人手で判定してもらい、次の 4 つの観点、即ち①書き手のレビューの評価、②レビューの記述内容、③食事、風呂などのトピック、④インパクトに影響を与える要因について、具体的に以下のように分析した。

まず、書き手がレビューに付与した宿泊施設の評価とインパクト文との関係を分析した。ポジティブなインパクト文は書き手が高い評価を下したレビューに多く含まれ、低い評価を下したレビューになるほど少なくなることから、書き手と読み手の評価が一致する傾向があることを示した。一方、ネガティブなインパクト文は、書き手が最低評価を下したレビューのみならず最高評価を下したレビューにも含まれていることから、読み手の評価と書き手の評価が必ずしも一致しないことを示した。このことから、自然言語処理分野における従来の評判分析の研究のような書き手中心の内容分析では不十分で、読み手の視点からの分析が重要であることを示した。

第2に、インパクト文と対象商品に関するレビューの記述内容との関係を分析した。レビューには、買い手による商品の評価、売り手の対応への評価、商品の購入理由など多様な内容が含まれている。しかし、これらを網羅する分類項目が存在しないため、商品情報、売り手、買い手の3つの側面に基づいて、さらに詳細で網羅的な23の分類項目を作成した。これらの分類項目は宿泊施設以外の商品であるバッグ、自転車、香水などの幅広い種類の15商品にも適用可能で、汎用性が高いことを示した。このような分類項目とインパクト文との関係を定量的に分析した結果、ポジティブなインパクト文では商品を販売する時の売り手の具体的な行動や、売り手の行動などに対する評価・感想が統計的に有意に多く出現することを示した。ネガティブなインパクト文では、「壁にひびが入っている」などの誰が見てもそうと認められる商品に関する情報および商品を販売するときの売り手の具体的な行動が統計的に有意に多く出現することを示した。

第3に、宿泊施設のレビューに含まれる立地、部屋、食事、風呂、サービス、設備・アメニティなどのトピックとインパクト文との関係を分析した。統計的に有意に多く出現するトピックは、ポジティブなインパクト文は食事とサービス、ネガティブなインパクト文は部屋とサービスであることを示した。逆に、統計的に有意に少なく出現するトピックは、ポジティブなインパクト文は立地と設備・アメニティ、ネガティブなインパクト文は立地と食事であることを示した。

最後に、同じトピックについて記述された文について、インパクト文と非インパクト文の違いを明確にするために、具体例を詳細に分析した結果、インパクトに影響を与える要因として「商品の特殊性」、「描写の具体性」、「書き手の満足度」が重要であることがわかった。この3つの要因とインパクトとの関連の強さを実験によって検証し、これらの要因がインパクトの判定に強い影響を与えることを定量的に示した。

3. 名詞およびその特性に関する研究

大規模コーパスから名詞の上位・下位関係の自動抽出および連想概念辞書を利用した名詞とその名詞と関係する形容詞や動詞などの特性についての分析を行った。

例えば「蛤のような貝」などの定型表現を用いて、延べ 38 年分の新聞記事を使用し、名詞の上位・下位関係を自動的に抽出した。上位語および下位語が共起する 7 種の定型表現を作成し、大規模な新聞記事をコーパスとして基本名詞の下位語を抽出した。具体名詞と抽象名詞についてそれぞれ正解率を算出し高い水準となることを示した。また、コーパスの種類によって出現する名詞が異なることから、コーパス特有の名詞、例えば新聞記事コーパスに特有の事件や条約などの上位・下位関係を自動抽出するために新たな実験を行った。定型表現「A のような B」の A および B に名詞相当語が出現した場合に A を下位語、B を上位語として名詞の上位・下位関係として抽出した。その結果、我が国の既存の大規模な EDR 電子化辞書と比較して 10 倍近い規模の名詞の上位・下位関係が抽出可能であるという見通しを得た。EDR 電子化辞書には存在しないデータとして、インターネット上のテキスト分析にも応用可能な固有名詞や時代性のある名詞が抽出されることを示した。また、その結果を利用し、インパクト文および非インパクト文に出現する名詞の上位語を調査した結果、それぞれに異なる種類の上位語が抽出され、名詞の上位・下位関係の情報がインパクト文および非インパクト文に出現する名詞の違いを自動抽出するために応用可能であるという見通しを示した。

次に、主に例示を表す定型表現である「A などの B」, 「A という B」, 「A のような B」, 「A といった B」を取り上げ、それらには大きく分けて例示, 説明, 範疇, 属性, 比喩の 5 種類の用法が認められ、定型表現の種類によって抽出される名詞に違いがあることを示した。上位語が固有名詞を下位語に持ちやすいか否か、また上位語が下位語に固有名詞を多く含む場合は上位語が「有意志」であるか否かによって定型表現の用法ごとに出現する名詞の上位・下位関係の頻度に明らかに違いがあることを示した。

名詞を刺激語とした連想概念辞書を用いて、ある名詞の下位に属する名詞の典型性を分析した。典型性とは、下位語が上位語の特徴をよく表していることを指す。例えば「鳥」の下位語として、「スズメ」や「ハト」などは典型性が強いが、飛べない「ペンギン」などは典型性が弱いとされる。刺激語である名詞とその名詞と関係する形容詞類や動詞などの特性を表す連想語との定量的な距離を使用して、名詞間の類似度を計算した。そ

の類似度を用いて、上位・下位関係にある名詞における特性の継承の傾向を分析し、上位・下位関係にある名詞に共通する重要な特性として、名詞の部分、名詞と関係する形容詞類、名詞と関係する動詞の3つを抽出した。例えば、「ウサギ」の部分は「目」や「耳」、形容詞類は「かわいい」や「元気な」、動詞は「跳ねる」や「食べる」となる。これらを用いて典型性を定式化し、定量的に有効性を示した。

4. まとめ

本研究は、「読み手の視点による宿泊施設のレビューの分析」と「名詞の上位・下位関係の自動抽出及びその特性の分析」の2つで構成される。前者では、レビューの読み手にインパクトを与える特徴として、レビューの記述内容とトピックを抽出し、統計的に有意な意味を持つことを示した。また、インパクトに影響を与える重要な要因として、「商品の特殊性」、「描写の具体性」、「書き手の満足度」を示し、これらがインパクト判定に与える影響を定量的に示した。後者では、延べ38年分の新聞記事から名詞の上位・下位関係の自動抽出を行って、高い抽出精度を得た。さらに、名詞の上位・下位関係がインパクト文の特徴の自動抽出に応用可能な見通しが得られた。上位・下位関係にある名詞の典型性を定式化し、定量的に有効性を示した。