

# 投稿者による評価スコアや感情極性辞書を用いない、文意に基づく宿泊施設不満レビュー自動抽出手法の構築及びその分析による不満予防策の検討

吉田 伊武貴<sup>†</sup> 倉田 陽平<sup>‡</sup>

<sup>†</sup> 東京都立大学大学院都市環境科学研究科観光科学域 〒192-0397 東京都八王子市南大沢 1-1

E-mail: <sup>†</sup> yoshida-ibuki@ed.tmu.ac.jp, <sup>‡</sup> ykurata@tmu.ac.jp

**あらまし** 近年の e コマースの普及によりユーザーの感想や印象が「レビュー」として記録されることが多くなりレビュー分析が盛んに行われている。一方で従来のレビュー分析では極性辞書を用いた単語ベースの感情分析による手法が主流であり精度に欠けることがわかった。そこで本研究では感情分析における精度の改善を目的に不満テキストコーパスと楽天トラベルの宿泊レビューを併用し自然言語処理技術によって不満を抽出する機械学習モデルを構築し宿泊レビューから効率的かつ、より確実な不満抽出手法を検討した。その結果、既存の感情極性辞書を用いた感情分析よりも良い精度を得ることができた。抽出された不満レビューを分析することによって典型的な不満の傾向把握や宿泊業者にとっての不満の予防策の提案などを見出し提案手法の有効性を検証した。

**キーワード** レビュー分析, BERT, 不満テキスト, 感情分析

## 1. はじめに

不慣れた土地で行われる観光には失敗経験やトラブルがつきものである。失敗経験やトラブルは不満を生じさせやすく、観光関連事業者にとって注意すべき課題である。特に COVID-19 パンデミック下の昨今は感染予防のための臨時休業やイベント休止が相次ぎ、予定通りの旅行を失敗した旅行者は潜在的に多数いると思われる。旅行中の失敗経験にはオーバーブッキングやフライト遅延や臨時休業のようにサービス事業者によるサービスの失敗に起因し、旅行者の責に帰さないものと、寝過ごしや道迷いや予約ミスのように旅行者の責に帰するものがある。前者についてはロイヤリティ低下やネガティブクチコミ拡散を予防するなどの方策がとられる。特に顧客の期待通りのサービスが提供できなかった場合に限定してとられる不満緩和方策のことをサービスリカバリと呼ばれ、サービスマーケティング分野において研究[1][2]が蓄積されてきた。武谷(2016)[1]によると具体的方策としては、補償や弁解・正当化・参照的説明・謝罪といった方策が知られている。

また SNS や e コマースの普及によって大量のテキストデータ（ツイートやレビュー）が得られるようになった今日、旅行者が旅行中どのような経験をしたか探るためにそれらのネット上にあふれるテキストデータを活用することが盛んになっている。それらのユーザーのテキストデータに対して、個々のテキストがどのような感情を記述したものなのかをコンピュータに自動判定させる技術（センチメント分析）が開発され、この技術はたとえば齋藤・横川(2016)による位置情報付きツイートに記載された感情に応じたアイコンで表現し、リアルタイムで都市の各所でどのような感情が飛び交っているかを可視化する取り組み[3]などに

応用されてきた。このような事例で用いられてきたセンチメント分析では一般に、単語ごとに設定されたポジティブ度合いやネガティブ度合いの情報が人手により付与された東北大学の乾・岡崎研究室が公開している、日本語評価極性辞書[4]のような感情語辞書を用いて、文章が肯定的か否定的かなどを判定している。よって文脈が全く考慮されておらず強いポジティブ語である「観光」の二文字が文章中に表れただけで肯定的側に判定が引きづられ、このせいで、たとえば宿泊レビューに適用するには評価が誤る危険性があった。また、若者言葉である「死ぬほど楽しい」のような表現は強いネガティブ語「死ぬ」に引きづられ、否定的評価と誤判定される危険性があった。実際に ML-ASK を利用して国立情報学研究所より提供されている不満調査データセット[5](以下不満コーパス)から宿泊・観光・レジャー」のカテゴリに含まれるテキスト 82358 件に感情分析を行ったところ、ネガティブ・ネガティブに近いと判定されたものは約 20000 件程度であり、それ以外は正しく判定できていなかった。表 1 は不満コーパスに対しての ML-ASK の判定結果の一部を抜粋したものである。

表 1 ML-ASK による判定結果

None	Neutral	Negative	Positive	Mostly_negative	Mostly_positive
51561	3404	19207	6599	1789	632

表 1 によると、「宿泊・観光・レジャー」カテゴリの不満コーパスの多くは「None」と判定されることが最も多く、さらに「Positive」と判定されてしまうものもあり精度が高いとは言えない結果となった。

表 2 判定結果例

判定結果	テキスト
None	駐車場が少なすぎる。そして高い!
Positive	サービスはとても充実して満足だったが駐車場が満車だったため契約駐車場を紹介してもらったがすごく分かりにくかった
None	あんなに汚いトイレ入ったの久しぶり。
None	インカのめざめフライドポテトが少し油っぽい。 もう少し油を切ってからお客様に出して欲しい。 胸焼けする。
Positive	お茶のティバックがあったが、数が少なくて喉が渴いた。
Positive	もっと用意していて欲しいです。
Positive	銭湯に入りに行き、洗い場がぬめってて子供が転んだ。 危ないからきちんと掃除してほしい
Positive	ファミレスはもちろん、気軽に食べれる飲食店が少ない。

背景には先述の極性辞書と呼ばれるものを利用した「単語」の出現を考慮して感情分析を行う手法であることから文脈を無視した感情分析に陥りやすいことが影響していることが考えられる。表 2 から誤判定の大まかな傾向として、「してほしい」のような願望を示す言葉が含まれているとポジティブと判定していることがわかった。

このことから本研究では感情分析において文脈の意味を考慮した手法を探り、既存の極性辞書を用いた「単語の出現」をベースとする感情分析に対する有効性を検証する。さらにその手法の実用性を検証するため宿泊施設のレビューに対して適用し不満の予防策につながるか考察を行う。

## 2. 関連研究

Kelly ら(1993)[2]は小売業におけるサービスの失敗とその対処策について調査し、15 のサービス失敗パターンと 12 のサービスリカバリ方策を見出しているが、米国の小売業を対象とした調査であり、その成果が日本の観光関連産業にも援用できるかは疑問の余地がある。本研究は日本国内の宿泊施設に対するレビューを調査対象とするという点と本研究の成果をサービスリカバリではなく不満そのものの予防策につなげる点で独自性がある。

Web 上にあふれる口コミやテキストデータに対してネガティブな情報の抽出を試みる研究として筆者らは旅先でのリスクを避けるために他の旅行者の失敗経験を知ることが重要であるという観点のもと、膨大な位置情報付きツイートの中から、自然言語処理技術を応用し、機械学習の分類モデルによって機械的に失敗談が記されたツイートのみを抽出し、それを用いて、地域固有の旅のリスクを導き出せることを実証した。[6][7]これらの研究における分析手法を本研究においても取り入れ、本研究においてはテキストが不満か否かを判定させる機械学習の言語分類モデルを構築する。

## 3. 研究目的

本研究では宿泊レビューと不満コーパスを併用し自然言語処理によって宿泊レビューから不満が述べられているレビューを抽出する。抽出された不満レビューを分析することによって宿泊業界における典型的な不満の傾向把握や宿泊業者にとっての不満の予防策などを見出し提案手法の有効性を検証する。

## 4. 研究方法

本研究では、株式会社リクルートがじゃらん口コミデータとして提供する"Japanese Realistic Textual Entailment Corpus"(<https://github.com/megagonlabs/jrte-corpus>)に含意関係データセットとして含まれている宿泊レビュー(以下じゃらん口コミコーパス)と不満コーパス、さらに岐阜大学が公開している Twitter 日本語評判分析データセット[8](以下ツイートデータセットコーパス)を使用し、テキストに対して不満を述べているかどうか判定する機械学習の分類モデルを作成する。その際、既存の感情分析ツールと結果を比較し精度を検証し楽天トラベルの宿泊レビュー[9]から不満レビューの抽出を試みる。抽出した不満宿泊レビューに対して自然言語処理における TF-IDF 値や共起ネットワーク図を作成し、宿泊業における不満傾向の把握や不満解消策を検討する。

## 5. 使用データ

機械学習モデル作成のための訓練データ・テストデータを作成するために以下のデータを使用する。本研究の目的である不満抽出のため、不満コーパスを使用する。不満コーパスの中で、宿泊業と関連性があると考えられる「宿泊・観光・レジャー」のカテゴリに含まれるものを本研究で使用する。その件数は 82358 件であった。またポジティブなテキストデータを確保するためにじゃらん口コミコーパスにてポジティブのラベルが付与されているコーパスを使用する。さらにレビューにはポジティブな内容でもネガティブな内容でもない中性的なテキストも含まれることが想定される。そのためツイートデータセットコーパスからラベルがネガティブ・ポジティブ以外のものが付与されているデータを利用する。じゃらん口コミコーパスとイトデータセットコーパスから不満コーパスと同数をランダムサンプリングし、これら 3 つを組み合わせる訓練データとして機械学習モデル作成に使用する。また有効性検証のためテストデータに対する既存の感情分析と比較した有効性の検証を行う。そのために楽天トラベル宿泊レビューを使用する。

## 6. 機械学習手法

使用言語は Python で機械学習フレームワークである PyTorch を用いて実装する.本研究における自然言語処理の機械学習分類モデルには BERT[10]の事前学習済みモデルを使用する.BERT は深層学習された事前学習済みの言語モデルで,転移学習することにより任意のタスクに応用することができる.本研究では先述の訓練データ・テストデータで転移学習させることとする.東北大学乾研究室が作成している BERT の日本語版事前学習済みモデルを利用する.[11]東北大学の乾研究室による日本語版の BERT 訓練モデルは日本語の Wikipedia を用いて訓練させたものである.このモデルを使用するため Python の自然言語処理ライブラリの HuggingFace の Transformers を使用する.Huggingface には各国の研究機関や企業が作成した BERT の事前学習済みモデルが用意されている.

機械学習モデルの評価指標には正解率,精度,再現率,F 値を用いる.使用するマシンの都合上,バッチサイズは 16,エポック数は 10 とした.また機械学習で扱うテキストデータに行った前処理は形態素解析のみで,従来のテキスト分析の際に行われるストップワードやノイズの除去といった前処理は行っていない.これは BERT モデルを活用して Twitter のツイートの作成者の性別を予測する研究[12]において,テキストの前処理を行わなかったモデルが最も良い精度を出した.この結果に基づき本研究においてもテキストに形態素解析以外の前処理は行わないこととした.訓練データ数は約 13 万件でテストデータ数は約 3 万 3 千件である.

## 7. 結果と考察

### 7.1 評価指標による比較

またテストデータにおける評価指標の結果を表 1 に表す.その際既存の感情分析との比較を行うため ML-ASK との比較を行う.ML-ASK の予測結果については,不満をネガティブなものと仮定した.本研究においては訓練データ・テストデータのラベルは,不満であるものには「1」,不満ではないものには「0」を与えている.よってテストデータに対する評価指標において ML-ASK の判定結果が「Negative」「Mostly\_Negative」で不満ラベル「1」と一致,もしくは「Positive」「Mostly\_Positive」「None」でテストデータのラベル「0」と一致しているものを正解とした.以下が評価指標の結果をまとめたものである.

表 3 評価指標による比較

	正解率	精度	再現率	F 値
ML-ASK	0.58	0.75	0.25	0.38
BERT	0.94	0.89	0.99	0.94

表 3 によると BERT を利用するモデルの方が ML-ASK の感情分析と比較して全ての指標でより良い結果を得た.

### 7.2 テストデータに対する判定結果の考察

テストデータの不満ラベルが付与されているものに対して BERT が正解し ML-ASK が不正解だったものは全部で 1 万 2 千件ほど存在し,ML-ASK の判定の内訳は「None」が 1 万件,「Mostly\_Positive」が約 100 件「Positive」が 1300 件であった.このことから感情分析においては文脈を理解する BERT を用いることが適していると言える.

表 4 は BERT が正解して ML-ASK が不正解だったものを一部抜粋したものである.

表 4 BERT が正解し ML-ASK が間違えたもの

有料貸出のとか風圧が弱くて使いにくいことが多い。 持参した方が良いかなと思う。
冷蔵庫があんまり冷えていなかった。 ちょっとのスペースに冷凍庫も置いて欲しいです。
〇〇ホテル、有名だから安心してたけど清掃状態は悪いし、 部屋の割に値段は高くいいところがなかった。
駐輪場からホテル玄関まで遠回りしてしまった。夜だとわかりにくい 就寝時電気が明るい。。。
もう少しちょうどよく調整できるようにしてほしい!
冷房の温度が低すぎる。寒すぎて楽しさ半減でした。
スイートルームは広いのに定員2人のところが多い。 我が家は3人家族なので泊まれず。泊まれるようにしてほしい。

表 4 には不満による「欲しい」「ほしい」と言いた要望や願望などが含まれているものがあり,これらの文言を適切に抽出することによりユーザーの不満に対する適切な不満の予防策につながることを期待できる.

### 7.3 楽天トラベル宿泊データに対する適用

作成した分類モデルを使用して,楽天トラベルから提供された楽天トラベルの宿泊施設に対するレビューに対して不満抽出を試みた.分類モデルにより不満と判定されたテキスト群の特徴を把握するため,TF-IDF 値を算出した.その際,全ての宿泊レビューコーパスと抽出された不満レビューコーパスとで比較を行った.

とでより有効な不満の予防策につながる事が期待できる。

表 5 全ての宿泊レビューと機械学習により抽出された不満レビューに対する TF-IDF 値(名詞)

不満 コーパス		楽天トラベル ホテルレビュー	
部屋	0.466262	部屋	0.466204
利用	0.373038	利用	0.377143
宿泊	0.269940	ホテル	0.267015
ホテル	0.268959	宿泊	0.262510
朝食	0.252898	朝食	0.252622
風呂	0.218992	風呂	0.217824
食事	0.195561	食事	0.191472
対応	0.140528	対応	0.143003
お部屋	0.128712	お部屋	0.127210
スタッフ	0.124146	スタッフ	0.122907

名詞のみでの TF-IDF 値の算出を行ったところ不満テキストとホテルレビュー全体で表 5 のような結果が得られ、違いはあまり見られなかった。

表 6 全ての宿泊レビューと機械学習により抽出された不満レビューに対する TF-IDF 値(動詞・形容詞・形容動詞)

不満 コーパス		楽天トラベル ホテルレビュー	
ほしい	0.584756	せる	0.387474
欲しい	0.376004	いただく	0.368200
買う	0.267691	広い	0.275932
入る	0.184100	頂く	0.256928
多い	0.182499	泊まる	0.204613
置く	0.180273	感じる	0.188492
高い	0.160422	おる	0.186873
入れる	0.138637	出来る	0.181233
少ない	0.134592	入る	0.167922
悪い	0.132084	近い	0.166878

名詞を除く動詞・形容詞・形容動詞による TF-IDF 値を算出したところ表 6 のように上位単語に大きな違いが見られた。機械学習により抽出された不満コーパスの上位単語に注目すると、「欲しい」「ほしい」といった単語が抽出されており、ユーザーが不満に感じたサービスや状態に対する要望や願望が述べられていることが推測される。このことにより、「ほしい」「欲しい」の単語について、その前後の文脈を理解するこ

表 7 実際に抽出された不満テキスト

部屋の壁が薄く、隣の部屋の音が気になった事と、風呂の排水が悪かったので、なんとかしてほしい。朝食は美味しかったので、良かったです。
綺麗でサービスも良く満足でした。出来れば朝食の種類を増やして欲しい。
久々に泊まったが部屋は普通。朝食のレストランが地下なので遠い。またメニューが変わらないので連泊だと美味しいが飽きてくる。
思ってたほど良くなかった。フロントの対応が少し悪かった。
朝食時混雑がひどく、食材も品薄であった事が残念です
喫煙ルームしかなくて利用したが、臭いがきつすぎて耐えられなかった。周辺に何も無い立地。
冷房の効きが悪かった。
朝食が少し高かった。
思っていたよりも朝食バイキングの品数が少なかった。個人の感覚ですが、また朝食付きプランで利用したいとは思いません。

さらに抽出された不満テキストを考察することにより具体的な不満傾向の把握や不満の予防策を検討する。表 7 は機械学習モデルにより不満と判定されたテキストを一部抜粋したものである。表 7 や表 5 から読み取れるように朝食の良し悪しが不満に繋がりやすい傾向があるということが推察される。また宿泊施設の清潔感に言及するレビューも数多く見られた。これらのことから不満の予防策として朝食サービスの充実さや施設の清潔感を意識する必要があると言える。本研究では施設ごとにコーパスを分けていないため、宿泊施設ごとに不満抽出を行い、考察を行うことでより具体的かつ効果的な不満の予防策を提案できると考えられる。

## 8. 今後の展望

本研究では宿泊業の不満緩和方策について不満情報を抽出するため楽天トラベルのレビューデータを用いたテキストからの不満情報抽出を試みた。その結果従来の感情極性辞書を利用するような感情分析による不満把握より、機械学習技術を利用し文脈から感情分析を行うことでより精度の高い結果を得ることができた。今後の展望としては従来、欧米の小売業を調査対象に行われてきたサービスリカバリ研究と比較し、日本の宿泊業に固有の不満の予防策を見出すことや、宿泊施設のレビューデータにとどまらず、Twitter などの SNS データを活用することにより宿泊業以外の有効性の検証が求められる。例えば観光地などで旅行者の不満を抽出し分析を行うことにより宿泊施設のようなサービス事業者ではなく観光地側（観光まちづくり組織や住民）における不満緩和方策のあり方の検討に寄

与できる可能性がある.また本研究においては既存の感情分析との比較の際に ML-ASK のみを利用するにとどまったが,他の既存の感情分析が使えるツールを利用し同様に精度比較を行うことも求められる.

#### 謝辞

本研究では,国立情報学研究所の IDR データセット提供サービスにより株式会社 Insight Tech から提供を受けた「不満調査データセット」を利用した.本研究では,国立情報学研究所の IDR データセット提供サービスにより楽天グループ株式会社から提供を受けた「楽天データセット」([https://rit.rakuten.com/data\\_release/](https://rit.rakuten.com/data_release/))を利用した.また本研究は科研費(21K12484)の助成を受けたものです.

#### 参考文献

- [1] 武谷 慧悟(2016) サービス・リカバリーにおける説明と顧客満足 -メッセージ要因と送り手要因に関する理論的検討-, 商学研究科紀要 (82), 21-46, 2016-03-25,早稲田大学大学院商学研究科
- [2] Kelly S. W,K. D. Hoffmann, and M. K. Davis,"A Typology of Retail Failures and Recovery," Journal of the Academy of Retailing, vol.69(4),pp.429-452.1993
- [3] 斎藤一,横川祥司,"感情語辞書と位置情報付きツイート分析に基づいたアプリケーション「Emonavi」の観光利用の検討”,DEIM FORUM 2016
- [4] 東北大学乾研究室,日本語評価極性辞書
- [5] 吉田伊武貴, 倉田陽平,"旅先における失敗リスクを把握可能にするための機械学習を用いた失敗談ツイート抽出方法の構築と静岡県内観光地での適用” 第7回とうかい観光情報学研究会, pp.1-4, オンライン発表.2021
- [6] 吉田伊武貴, 倉田陽平," 失敗談ツイートのテキスト分析による北海道における観光失敗リスクの把握”, 観光情報学会第22回研究発表会講演論文集,pp.25-28,オンライン発表.2021
- [7] Kensuke Mitsuzawa, Maito Tauchi, Mathieu Domoulin, Masanori Nakashima and Tomoya Mizumoto. FKC Corpus: a Japanese Corpus from New Opinion Survey Service. In proceedings of the Novel Incentives for Collecting Data and Annotation from People: types, implementation, tasking requirements, workflow and results, pp.11-18, Portorož, Slovenia, May 2016.
- [8] Yu Suzuki," Filtering Method for Twitter Streaming Data Using Human-in-the-Loop Machine Learning” Journal of Information Processing, Vol. 27, p. 404-410.2019
- [9] 楽天グループ株式会社 (2014): 楽天データセット. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.2.0>
- [10] Devlin, Jacob, Chang, Ming Wei, Lee, Kenton, Toutanova, Kristina.“BERT: Pre-training of Deep Bidirectional Transformers for

Language Understanding. Association for Computational Linguistics” Human Language Technologies”, Volume.1,pp.4171-4186 2018

- [11] 東北大学 乾研究室 日本語 BERT 訓練済みモデル <https://github.com/cl-tohoku/bert-japanese>
- [12] Esam Alzahran, Leon Jololian.” How Different Text-preprocessing Techniques Using The BERT Model Affect The Gender Profiling of Authors”, CoRR,2021