

Title	旅館・ホテル宿泊者の口コミデータによる顧客満足度要因分析
Sub Title	
Author	阿部, 璃加子(Abe, Rikako) 木村, 太一(Kimura, Taichi)
Publisher	慶應義塾大学大学院経営管理研究科
Publication year	2024
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2024年度経営学 第4212号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40003001-00002024-4212

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

慶應義塾大学大学院経営管理研究科修士課程

学位論文（ 2024 年度）

論文題名

旅館・ホテル宿泊者の口コミデータによる顧客満足度要因分析

主 査	木村 太一
副 査	井上 哲浩
副 査	林 高樹

氏 名	阿部 璃加子
-----	--------

論文要旨

指導教員	木村太一 先生	氏名	阿部璃加子
(論文題名) 旅館・ホテル宿泊者の口コミデータによる顧客満足度要因分析			
(内容の要旨) 本研究は、旅館・ホテル業における顧客満足度の要因を明らかにし、宿泊施設の競争力向上に寄与する戦略を立案することを目的とする。特に、口コミデータを活用した定量的な分析に着目し、顧客がどのような要素に満足し、どのような点に不満を抱くのかを探る。自社である南三陸ホテル観洋の戦略立案に役立てることを視野に入れつつ、宿泊業全体の傾向や競合他社の宿泊者の動向を把握し、差別化のための示唆を得ることが本研究の狙いである。 データとしては、宿泊予約サイトじゃらん.net に投稿された口コミを対象とし、2023 年 5 月から 2024 年 10 月までの約 2,700 件の口コミデータを分析に用いた。分析手法には、LDA（潜在的ディリクレ配分法）を用いたトピック分析と、順序ロジット回帰分析を採用した。LDA によって口コミから接客・食事・部屋・家族との時間といった主要トピックが抽出され、それぞれのトピックが評点（星の数）に与える影響を順序ロジット回帰分析により検証した。 その結果、接客や食事に関する口コミが高評価に直結する傾向が見られ、一方で部屋や設備に関する不満が低評価につながる事が明らかになった。また、年齢が高くなるほど評価が厳しくなる傾向が認められ、特に 30 代から 60 代の顧客が低評価をつけやすいことが判明した。さらに、価格帯が低～中程度のプランを利用した顧客の方が、高価格帯の顧客よりも高い評点を付けやすい傾向があった。 本研究の結果を基に、南三陸ホテル観洋における戦略として、星 4 の顧客を星 5 に引き上げる施策の実施が求められる。特に、顧客満足度に直結する食事と接客の強化を継続しつつ、部屋や設備の維持・改善を図ることが必要である。また、自社だけでなく競合他社の口コミデータを引き続き分析することで、宿泊業全体のトレンドを把握し、自社の強みをさらに伸ばす方針が重要である。 本研究は、宿泊施設の経営におけるデータ活用の可能性を示すとともに、定量的な分析を通じて顧客満足度向上に向けた実務的示唆を提供するものである。			

目次

1. はじめに	4
1.1. 研究の背景.....	4
1.2. 研究の目的.....	6
2. 先行研究	8
2.1. 宿泊予約サイトの口コミと顧客満足度に関する研究.....	8
2.2. 宿泊予約サイトの口コミに対するテキストマイニングの研究	8
2.3. レビューに対するテキストマイニングの研究.....	9
2.4. 本研究の貢献と新規性	10
3. 研究方法	11
3.1. 対象データ	11
3.2. 分析手順	12
3.3. 分析手法	13
3.3.1. 潜在的ディリクレ配分法(LDA).....	13
3.3.2. 順序ロジット回帰分析.....	15
4. 分析結果	18
4.1. トピック指数の作成.....	18
4.1.1. 形態素解析の結果.....	18
4.1.2. LDA の結果.....	18
4.2. 順序ロジット回帰分析の結果	20
5. 考察・仮説構築.....	22
6. 戦略立案	26
7. おわりに	32
謝辞.....	34
参考文献.....	35

1. はじめに

1.1. 研究の背景

本研究は、筆者の家業で継承予定であるホテルの今後の戦略を考えるうえで、現状として顧客の有益なフィードバックといえる口コミを用いた定量的な分析が行われていないという課題から着想に至った。家業のホテルである南三陸ホテル観洋は、宮城県南三陸町に位置し、1972年に開業した。2004年には大浴場や露天風呂を備える南三陸温泉を開湯した。2011年の東日本大震災では、10階建ての建物のうち2階までが津波により浸水したが、直後から避難所として町民を受け入れながら営業を再開した。当ホテルが運行している語り部バスは、2024年5月の時点で延べ46万人以上が乗車し、日本観光振興協会と日本旅行協会(JATA)、日本政府観光局(JNTO)が開催した第3回ジャパン・ツーリズム・アワードでは大賞を受賞した。また、全国被災地語り部シンポジウムも毎年開催し、ホテル自体も震災の伝承の場となっている。

当ホテルでは、顧客の声を収集する方法として、紙の宿泊者アンケートとじゃらん.net や楽天トラベル、一休.com、るるぶトラベルなどの宿泊予約サイトに寄せられた口コミを用いている。口コミについては、ネット課がこまめに確認し、口コミに対して返信をしたり、指摘があった箇所については各担当部署に共有し、改善を求めたりしている。顧客の声を把握しているのは、ネット課および紙のアンケートを確認する上層部のみであり、レストランや宴会を担当するサービス課や清掃を担当する業務課に全て届いているわけではない。また、ネット課や上層部も顧客が何に対して満足しているか、不満を抱いているかは肌感覚でしか把握していない現状にある。

この研究の動機は、南三陸ホテル観洋の今後の戦略を考えるにあたり、現状では行われていない口コミデータという顧客のフィードバックの定量的な分析を行うことで、当ホテルの強みと弱みを知ることである。口コミデータを定量的に分析することは、分析者のバイアスを可能な限り排除した客観的な結果が得られるほか、顧客の感情までも定量的に評価することができるため、感情の傾向を把握することも可能にする。さらに、自社だけでなく、宿泊業全体の口コミをみて傾向を知ること、競争力強化に向けた戦略にもつながる。

さらに、化粧品や衣料品、電子機器などの製品と同様に宿泊予約サイトの口コミも閲覧者に影響を与えることが報告されている。中国の大手ネット旅行会社の口コミを対象とした研究によると、サイトでの評価が10%上昇すると、オン

ライン予約が5%以上増加することがわかった(Ye, Law and Gu, 2009)。加えて、化粧品や衣料品などの機能やスペックなどの具体的な情報が購入意思決定において優先される検索財に比べ、実際に体験するまで品質を評価しづらいホテルやレストランなどの経験財はレビューの影響が非常に大きい。つまり口コミで宿泊客から良い評点をもらうことは、顧客獲得に繋がる重要なマーケティング施策となる。

現在、宿泊施設の口コミデータを用いた研究はいくつか行われている(田邊・後藤,2008; 沢田,2013; 久保田,2016)。これらに共通していることは、宿泊者がつけた評点ではなく、個々がつけた評点を平均した宿泊施設としての評点を用いて高評価の宿泊施設の口コミを研究対象としている。宿泊施設としての評点について、じゃらん.net では「過去投稿された観光スポット・イベントの評点の平均点を毎日算出しております¹。」(図 1)、楽天トラベルでは「過去1年以内にご回答いただいたアンケート結果を対象とした、平均値を出しています²。」とそれぞれ記載されているが、信憑性があるとは限らない。実際に、大手グルメサイト食べログは、評点を決めるアルゴリズムを一部意図的に変更したことにより飲食店から損害賠償を求められ裁判にまで及んだ³。

本研究では、宿泊者個人による評点と口コミ内容を分析対象とすることで、宿泊施設に対する評点を対象とした先行研究と比較して、より直接的に顧客の評価行動に関する洞察を得ることを目指す。また、先行研究では低評価の口コミを対象とするものはなかったため、どちらの評価にも目を向ける点で顧客満足度に関する新たな洞察が得られると考える。なお、本研究の目的は、あくまで各顧客の評価行動の解明であり、各口コミサイトにおける宿泊施設の評点アルゴリズムの解明ではない。

¹ じゃらん,「観光スポット・イベントの評点の見方・説明」,
<https://www.jalan.net/jalan/doc/howto/guidekuchikomi.html>, 2024 年 6 月 4 日アクセス

² 楽天トラベル,「星の見方について」,
https://travel.rakuten.co.jp/share/toukou/explain_ws.html, 2024 年 6 月 4 日アクセス

³ 「『食べログ』に賠償命令 点数評価の公平性に一石 飲食店の不満反映」, 日本経済新聞, 2022 年 6 月 17 日, 朝刊, 13 頁.

クチコミ評点の見方・説明

■ クチコミとは？

じゃらんnetで予約をし、実際に宿泊された方が投稿したコメントです。投稿できるのは会員登録して予約をした方で、投稿期間はチェックアウト翌日から起算して60日以内とさせていただきます。※会員登録せずに予約をした場合、電話予約センター経由の場合は投稿できません。

■ クチコミ評点について

クチコミは「部屋」「料理（朝食）」「料理（夕食）」「風呂」「接客・サービス」「清潔感」「総合評価」について、以下の評点を投稿していただきます。

5. (★★★★★) 大変満足
4. (★★★★☆) 満足
3. (★★★☆☆) 普通
2. (★★☆☆☆) やや不満
1. (★☆☆☆☆) 不満

なお、食事つきでない場合やお風呂がない場合など、該当設備がない・該当サービスを受けていない場合は「評価不能」を選択することになっており、「評価不能」を除いた評価が5件未満の項目については、「―」を表示しています。

- ・国内宿泊では過去1年間に投稿されたクチコミ評点の平均点を毎日算出しております。
- ・投稿件数が5件未満の宿泊施設については、評点は表示していません。

図1 じゃらん.net『クチコミ評点の見方・説明』 筆者作成

1.2. 研究の目的

本研究の目的は、自社ならびに競合他社の高評価および低評価の口コミデータを分析することで、顧客満足度に影響する要因を明らかにし、高い点数を獲得するために今後とるべき戦略を立案することである。この目的を達成するため、本研究では口コミデータを探索的に分析することによって、各顧客の評点に影響を与える要因についての仮説を導き出す仮説探索型の研究を実施する。

宿泊施設のサービスや顧客満足度を扱うにあたり、重要なキーワードとなるのが、日本特有の文化であるおもてなしである。三省堂『現代国語辞典』によれば、おもてなしは「客に対する扱い方、待遇、ごちそう」と定義される。おもてなしに関する大規模な顧客満足度調査としては、Denove and Power (2006) が挙げられる。彼らは、日本全国の宿泊業グループ・チェーン 139 社を調査対象として、①チェックイン/チェックアウト、②客室、③料飲、④施設、⑤サービス、⑥予約、⑦料金の 7 つの要因に項目を分け、アンケート調査を実施している。各項目に関する宿泊客の評価は複数の質問項目を用いて測定され、それをも

とに総合満足度スコアが算出されている。Denove and Power (2006)による調査で注目すべき点は、おもてなしを 7 つの要素から構成される複合的な概念として捉えている点である。つまり、おもてなしを単に接遇だけを意味する概念ではなく、施設や料理、周辺環境を含めた顧客が体験するサービス全般と対象とした広範な概念として捉えているのである。

北中・香坂・伊藤 (2018) も指摘するように、顧客がおもてなしを評価する際は、サービス内容を明確にせず、包括的な評価を実施している。そのため、宿泊施設において素晴らしいおもてなしを提供する旅館・ホテルと評価されても、顧客が具体的に何に対して満足または不満を感じているかは把握しづらい。また、接客はいいが料理が不味いなど、特定の一つの要素に対する不満が、おもてなし全体に悪影響を与えてしまう場合もある。おもてなしという概念が、全体的な顧客サービスを対象とした概念であることを考慮すれば、おもてなしを構成する一つ一つの要素を分析する必要がある。

口コミデータの分析を通して顧客の声を理解することで、顧客が期待するおもてなしといま現在宿泊施設が提供しているおもてなしが合致しているかを知ることができる。提供しているおもてなしが顧客の期待と合っているとは限らない。そのため、旅館・ホテルの経営においては顧客の声を聴き、今まで提供できていなかった顧客が期待するおもてなしを加えたり、提供していたが顧客が望んでいなかったおもてなしの部分を削減したりするよう尽力する必要がある(内藤, 2015)。

また、個々の口コミを分析対象とすることで性別や年齢などの個人の特徴が点数に影響するかも把握することができる。従来の口コミ分析は施設全体の評価や平均点を対象としたものが多く、個々の口コミ投稿者の属性とその内容がどのように評価点数に反映されるかについては十分に分析されてこなかった。

本研究では、今までおもてなしという言葉で括られてきた旅館・ホテルが提供するサービスを宿泊者が投稿した口コミデータを用いて分析することで、おもてなしを構成する一つ一つの要素への反応を具体的に読み取り、顧客満足度の要因を解明するとともに、顧客の属性に合わせた適切なアプローチを模索する。本研究によって顧客満足度向上のためのヒントが得られ、宿泊施設や観光業界に対する実務的な示唆を提供することが期待される。そして結果をもとに、最終的には研究対象となる自社ホテルの戦略を立案する。

2. 先行研究

2.1. 宿泊予約サイトの口コミと顧客満足度に関する研究

田邊・後藤(2008)は、宿泊施設の利用者が投稿した口コミを分析し、顧客満足度に影響を与える要因について考察した。特に、宿泊施設の評価は部屋・食事・風呂・接客・サービスなどの要素を中心に行われており、これらの項目が顧客の満足度向上に寄与していることが明らかにされた。口コミの中で頻繁に言及される内容が高評価を獲得している場合、その要素が宿泊体験の満足度を左右する重要な要因となることを示した。例えば、風呂の質が高く評価された宿泊施設では、全体の満足度が高くなる傾向が確認された。このことから、施設の特定のサービスや設備に対する顧客の評価が、総合的な満足度に直接影響を与えることが示唆されており、口コミの点数は顧客満足度を測る有力な指標となる。田邊・後藤(2008)は、口コミの評価と顧客満足度が密接に関連していることを明確に示し、口コミ分析を通じて顧客の声を可視化し、満足度の向上につなげるための有効なアプローチを提示した。

2.2. 宿泊予約サイトの口コミに対するテキストマイニングの研究

沢田(2013)は、宿泊予約サイトじゃらん.net の口コミデータを対象に、テキストマイニングを用いて宿泊者のモチベーションを分析した。口コミ内容を佐々木(2000)の旅行者モチベーションの基本的な枠組みである5段階分類(緊張解消、娯楽追及、関係強化、知識増進、自己拡大)に基づいて5つの階層に分類し、計量テキスト分析ツール KH Coder を用いて特徴語の抽出や共起ネットワークの分析を行った。その結果、緊張解消に関連する温泉や食事、部屋が口コミで多く言及され、特に高価格帯の宿泊者ほど高次の欲求(自己拡大や関係強化)に関する要素が重視されることが示された。

久保田(2016)は、宿泊予約サイト楽天トラベルの口コミデータを対象に、テキストマイニング手法を用いて宿泊施設の口コミ傾向を分析した。具体的には、Text Mining Studio および R 言語を活用し、頻度分析、クラスター分析、ネットワーク分析、対応分析を行い、顧客評価の特徴を体系的に明らかにした。その結果、口コミで頻出する単語として、部屋、朝食、利用、ホテル、対応などが抽出され、クラスター分析により宿泊施設がビジネス利用とプライベート利用の2つの軸で特徴づけられることが示された。さらに、対応分析においては、ビジネ

スホテルでは立地、フロント、対応が重視される一方、プライベート利用の宿泊施設では美味しい、快適、過ごすといった快適性に関連する要素が顧客満足度に大きく寄与していることが示された。また、具体的な事例として東横イン、ドミーイン、アパホテルの3つのホテルグループを対象に分析を行い、それぞれの強みと課題を明確に示した。テキストマイニングを用いて宿泊予約サイトの口コミを構造化し、顧客が宿泊施設に求める要因を体系的に分析したことで、ビジネスホテルとプライベート利用の宿泊施設が重視すべきサービスの要素が明らかとなり、顧客満足度向上や宿泊施設の戦略立案に対する有用な示唆を提供した。

荒木・小村・平井(2022)は、国内旅行情報サイトの口コミデータを対象に、口コミ内容とホテルのランキングとの関係性を明らかにするための分析を行った。口コミデータにはテキストマイニングの手法を適用し、LDA(Latent Dirichlet Allocation)モデルによる潜在意味解析を用いてトピック指数を作成した。さらに、ランキングを目的変数、口コミのトピック指数を説明変数とし、ホテルや顧客の特徴をコントロール変数とする階層的重回帰分析を実施した。その結果、口コミデータから、風呂・温泉、清潔さ、ウイルス感染対策、スタッフ対応、家族との食事の5つのトピックが抽出され、特に、ウイルス感染対策、スタッフ対応、家族との食事に関する内容がランキングに対して統計的に有意な正の影響を与えていることが示された。荒木・小村・平井(2022)は、口コミ内容という非構造化データをLDAモデルによって構造化し、ランキングとの関係性を明らかにした点で意義がある。

2.3. レビューに対するテキストマイニングの研究

松河ほか(2017)は、授業評価アンケートにおける自由記述の分析を効率化し、その内容を詳細に把握するために、トピックモデルを用いた研究を行った。本研究では、ある大学で9年間にわたって収集された約60,000件の自由記述を対象に、潜在ディリクレ配分法(LDA)を用いて170のトピックを抽出した。各トピックにラベルを付与し、その妥当性を検証することで、トピックモデルによる自動分類が人間の感覚に適合していることを確認した。また、自由記述と授業科目群の情報を紐付け、各科目群におけるトピックの出現割合を明らかにし、科目群ごとの特徴を把握するための可視化を行った。松河ほか(2017)は、自由記述に含まれる情報を網羅的に分析し、従来の手作業による分析では困難だった大量データの処理を実現した点で意義をもたらした。

2.4. 本研究の貢献と新規性

本研究は、家業のホテルの戦略立案に貢献できると考える。先でも述べたように、自社は現状として口コミデータを用いた定量的な分析ができていない。2024年6月時点で、じゃらん.netには4,034件、楽天トラベルには3,919件、一休.comには111件もの口コミが宿泊者より寄せられている。この膨大で貴重なデータを活用しないことは極めて惜しい。しかしながら一つ一つの口コミを読みながら分析するには、時間と手間がかかり、全てに目を通すのは不可能であると同時に担当者の主観が入る可能性がある。そこで、文章やテキストの統計解析を可能にするテキストマイニングを用いた定量的な分析を行うことで、バイアスを可能な限り除いたより正確な結果を得ることができる。より精度が高い結果をもとに、仮説を構築したうえで戦略立案に結びつけることで、顧客が求めるもののおもてなしの提供につながると考える。

また、口コミは競合他社のものも閲覧可能であるため、自社だけでなく他社の口コミも同時に分析対象とすることで、宿泊施設や観光業界全体の傾向を把握することができる。現状として、宿泊予約サイトにて他社の価格変動は随時確認しているものの、口コミまで目を通すことはしていない。旅館・ホテルの経営者が集い、情報交換をする場もあるが、共有されていない情報も多々あると推測される。おもてなしおよび顧客満足度の要因はそれぞれの宿泊施設によって異なるが、顧客の声から競合他社の動向を把握することで市場のトレンドや顧客のニーズをもとに自社の改善点や差別化の要素を探ることができると期待される。

先行研究で顧客満足度と口コミの関係は明らかになっている。筆者が調べた限りでは、宿泊予約サイトの口コミに対するテキストマイニングの研究は、口コミ内容のみで宿泊者のモチベーションについて分析したものがほとんどで、点数への影響を分析に組み込んでおらず、顧客満足度もマイナスの要因も明らかにしているものは少なかった。また、荒木・小村・平井(2022)では、ランキングと口コミ内容の関係性が明らかになったものの、総合点による順位は信憑性に欠け、評点を上げる要因、下げる要因を探るには十分でないと考ええる。さらに、荒木・小村・平井(2022)は、新型コロナウイルス感染拡大により各宿泊施設が感染対策を行なっている期間が分析対象になっていたためコロナ禍の影響が結果に反映されている。新型コロナウイルスが第5類に移行してから宿泊施設で感染対策がみられなくなった。このことから、アフターコロナのデータで個人の評点と口コミ内容の関係性を分析し、顧客満足度の示唆を得ることは新規性があるといえる。

3. 研究方法

3.1. 対象データ

本研究では、宿泊予約サイトじゃらん.net の口コミデータを研究対象とする。じゃらん.net を選定した理由としては、デジタルコンテンツ視聴率を利用した調査で、旅行・交通カテゴリにおいてじゃらん.net は利用者が最も多いという結果が出ているからである(観光経済新聞, 2022)。

じゃらん.net のデータは、10 施設の 2701 件の口コミを含む。これらは 2023 年 5 月 8 日から 2024 年 10 月 31 日までに投稿されたものである。期間の選定理由については、コロナ禍の影響が口コミに反映されないよう新型コロナウイルスが新型インフルエンザ等感染症(いわゆる 2 類相当)から季節性インフルエンザと同等の位置付けとなる 5 類感染症に移行した 2023 年 5 月 8 日を起点とし約 1 年半のデータとする。研究対象となる 10 施設は、じゃらん.net が特集した売れ筋ランキングの東北エリアで 101 室以上の部門にランクインした 10 施設とする(表 1)。このランキングには自社である南三陸ホテル観洋もランクインし、競合となる近辺で同規模の他社が集まっているため、このランキング特集に掲載されている施設を対象とした。ランキングを取り上げるものの、順位や施設ごとの総合評点はデータとして用いず、あくまで個人の評点と個人の口コミデータを使い分析をする。データの収集はウェブスクレイピングを用いる。分析方法は、大きく 2 つの手段からなる。まず、口コミ内容を反映したトピック指数を作成する。その後、トピック指数を説明変数、個々の口コミの評点(星の数)を被説明変数とした順序ロジット回帰分析を実行する。

表 1 データ概要

口コミ	2,701 件
	〈10 施設内訳〉
	① ホテル松島大観荘
	② スパリゾートハワイアンズホテルハワイアンズ
	③ 仙台秋保温泉ホテル瑞鳳
	④ TAOYA 秋保
	⑤ 秋保グランドホテル
	⑥ ホテルリステル猪苗代ウイングタワー

	⑦ メルキュール宮城蔵王リゾート&スパ
	⑧ 南三陸ホテル観洋
	⑨ アパホテル〈仙台駅五橋〉
	⑩ 会津・東山温泉 御宿 東鳳(オリックスホテルズ&リゾーツ)
期間	2023 年 5 月 8 日～2024 年 10 月 31 日
	i. 口コミ内容を反映したトピック指数の作成
分析方法	ii. トピック指数を説明変数、口コミの評点を被説明変数 とした順序ロジット回帰分析の実行

3.2. 分析手順

分析にあたり、統計ソフトウェアの R を使用する。使用する分析手法は潜在的ディリクレ配分法(以下 LDA とする)と順序ロジット回帰分析である。i.として口コミデータに対してテキストマイニングを行い、口コミ内容を反映した変数(トピック指数)を作成する。トピック指数に関しては、3.3.1 で後述する。具体的には、i.-1)収集した口コミデータに前処理を施し、i.-2)ストップワードを設定した上で、i.-3)口コミに対するテキストマイニング(形態素解析、LDA)を行い、トピック指数を作成する。

i.-1)の段階で、口コミデータに対して記号と絵文字の除去を行う。口コミは自由記述形式のため記号や絵文字が頻出する。これらは、投稿者の感情が反映されることが想定されるが、本研究では、口コミの文章として記載されている内容に着目して分析を行うことに主眼を置いて、テキストマイニングを行う前に記号および絵文字を除去する。

i.-2)では、テキストマイニングを行うにあたり、ストップワードを設定する。ストップワードとは、文章中で頻繁に使われるものの、意味の解釈や分析には寄与しにくい単語のことを指す。具体的には、助詞・助動詞、また・しかしなどの接続詞、私・あなたなどの代名詞、なる・できるなどの補助動詞、いく・見るなどの頻出する一般動詞・形容詞である。ストップワードを使用する主な目的は、テキストマイニングにおけるノイズを削減し、分析の精度を高めることにある。文章中で頻繁に出現する助詞や助動詞、接続詞などは、文の構造には必要であるものの、分析対象のテーマや重要な情報を反映しているわけではない。これらの単語を除去することで、より意味のある単語の抽出が容易になり、重要なキーワードが際立つようになる。加えて、重要な単語が頻出することで、出現傾向がより鮮明に浮かび上がり、テーマや特徴の把握が容易になる。したがって、ストッ

ブワードの使用は、テキストマイニングの結果の信頼性を高め、対象となるデータの本質的な意味を効果的に抽出するための重要なプロセスである。

i.-3)で、口コミデータに対してまず形態素解析を行う。形態素解析とは、「テキストを単語に分割し、各単語に品詞を割り当てる処理のこと」(鍛冶, 2013)を指す。その後、形態素解析を行った口コミデータに対してギブスサンプリング法でLDAを実行する。これにより、どのようなことを話題にしているかなどの口コミデータが持つ意味を量的に把握する。

ii)では、LDAによって算出したトピック指数を説明変数、投稿者個人とホテルの特徴をコントロール変数、個人の評点(星の数)を被説明変数として分析を実施する。目的変数である評点(星の数)が1~5の順序尺度であり、結果の解釈が容易であることから、分析モデルには順序ロジットモデルを採用する。順序ロジット回帰分析では、回帰係数やオッズ比を通じて「特定の要因が高評価や低評価のどちらに影響するのか」を明確に示すことができる。例えば、特定のトピックに関連する係数が正であれば、そのトピックが高評価につながりやすいことが示唆される。最後に、回帰分析の結果をもとに口コミ内容と個人の評点の関係と顧客満足度要因を結びつけて考察を行う。

3.3. 分析手法

3.3.1. 潜在的ディリクレ配分法(LDA)

LDAは、文書に複数のトピックが共存するという前提に基づいた確率的トピックモデルであり、各トピックが一定の確率で文書内に現れると仮定し、その確率分布を推定する手法である(図2)。

文書はトピックの割合が異なる組み合わせから成り立つと考えられ、例として、「文書Aはトピック1が70%、トピック2が30%を占める」一方で、「文書Bはトピック1が10%、トピック2が90%を占める」といったようにトピックごとの比率で表すことができる。トピック自体は特定の単語群によって特徴づけられる。具体的に、ニュース記事を「政治」と「スポーツ」の2つのトピックに分類すると、政治では、首相・議会・政策などの単語が多く現れ、スポーツでは試合・チーム・サッカーといった単語が頻出することが考えられる。さらに、一部の単語は複数のトピックで共通して使用されることもある。

LDAは、文書内にどのトピックがどの程度含まれているかを確率的に推定し、同時に各トピックに関連付けられた単語群を抽出することで、文書とトピック、および単語との関係を解明するモデルである。

本研究の分析には R の Package “topicmodels” を使用し、LDA の結果を踏まえ、トピック指数を算出する。算出式は(1)の式の通りである（荒木・小村・平井, 2022）。

$$X_n = A_n T_n \quad \cdot \cdot \cdot \cdot \cdot (1)$$

$$A_n = [a_{i,j,n}]$$

$$T_n = [t_{j,n}]$$

n: トピック連番(=1,2,3,4)

i: 口コミ連番(=1,2,⋯,2,701)

j: 各トピックにおける出現上位単語(=1,2,⋯,10)

本研究では、 T_n をトピック n における単語の出現確率ベクトルとし、その中から確率値が大きい順に 10 個の単語を抽出したベクトル (j 行 $\times$ 1 列) と定義する。また、 A_n はトピック T_n で抽出された 10 個の単語について、各口コミにおける出現頻度を計算して得られるベクトル (i 行 $\times$ j 列) である。

さらに、 A_n に T_n を乗じることでトピック n におけるトピック指数 X_n を算出する。トピック指数とは、各口コミの単語出現頻度に対して、その口コミがトピック n をどの程度含んでいるかを重み付けした指標とみなすことができる。また、トピック指数の計算にあたっては、各トピックにおける出現頻度上位 10 単語に限定して算出しており、この 10 単語を j として設定している。

テキスト	トピック			
	1	2	3	4
何回か宿泊していますが、ここに泊まる理由は風呂と飯につきます 朝も夜もどちらも間違いない仕上がりです。妻と無口になるくらいカニを食べまくり、焼いた笹かま、寿司、ステーキ、天ぷら、とほぼ食べたものがとにかくあります。妻とはなんか美味しいもの食べようと思ったら高級なレストランよりも選択肢として瑞宝が入ります。また食べたくなったら来ます	0.96	0.01	0.01	0.01
夕食も朝食も種類が豊富で、とても美味しかったです。ただ楽しみにしていた牛タンが硬くて噛みきれず、一枚しか食べられなかったのが残念でした。部屋は綺麗で眺めも良かったです。お風呂は夜と朝で入れ替わりましたが、片方は露天風呂がいくつかあって、脱衣所もとても広く良かったです。機会があればまた行きたいです。	0.94	0.02	0.02	0.02
川の音がいいというので本館に宿泊しました。川は木々に遮られ、見えなかったが、川の流れる音に癒された。本館の風呂はまでも別館の風呂や食事会場まで行くのに長かった。	0.94	0.02	0.02	0.02
なんやかんやで、5回目の利用です、とにかく温泉が最高に良い。眺めも最高。泉質最高。リビ間違いなしです又伺います。	0.92	0.03	0.03	0.03
お風呂は空中に浮いている様でとても良かった。お食事バイキングでしたが、品数も多く、味も良く満足しました。ただサラダの種類がもう少しあったら嬉しいです。是非また利用したいと思います。	0.92	0.03	0.03	0.03

図2 LDAを用いたトピック解析の全体像

3.3.2. 順序ロジット回帰分析

順序ロジット回帰分析とは、順序尺度のデータを目的変数とする場合に用いられる回帰分析手法である。順序尺度とは、データが順位付けされたカテゴリであり、1位、2位、3位や「低い」「中程度」「高い」のように順序関係はあるが、隣接するカテゴリ間の間隔が等しいとは限らないデータを指す。

宿泊予約サイトにおける星の数(1~5)の評価データは、1から5まで順序があるものの、各星の間の満足度の差が等間隔であるとは言えない。このような場合、通常の線形回帰分析ではデータの性質を適切に捉えることができないため、順序ロジット回帰分析が適切な手法とされる(山本,2015)。

順序ロジット回帰では、各カテゴリ(例えば「星1」から「星5」)がある基準を超える確率を推定することで、順序関係を維持しながら分析を行う。モデルの結果として得られる回帰係数は、説明変数が目的変数に与える影響の方向性と強さを示し、オッズ比として解釈することができる。具体的に、特定のトピックや属性が星の数に与える影響が正の場合、その要因が評価を高める方向に寄与していることを意味する。

本手法は、順序付けされた評価データの背後にある要因を統計的に明らかにすることができるため、顧客評価や満足度を対象とした研究において有用である。

本研究では、口コミ内容から作成したトピック指数を説明変数とする。コントロール変数については、投稿者個人の特徴として、性別、年齢、利用したプランの価格帯、旅行相手の有無と、ホテルの特徴として部屋数の合計 5 つの変数を作成する(表 3)。これらは評点のつけ方に影響を与えている可能性があるためコントロール変数として設定した。旅行相手の有無については、口コミ投稿で家族旅行や友達旅行、出張利用など誰との旅行かを記入する項目がある。この項目をもとに旅行相手の有無をダミー変数化する(表 2)。

表 2 旅行相手の有無に関する変数

旅行の種類	変数
一人旅	1
家族旅行	0
子連れ旅行	0
出張	1
団体旅行	0
夫婦旅行	0
友達旅行	0
恋人旅行	0

表 3 その他の変数の概要

性別	男性=1、女性=0
年齢	10代=1、20代=2、30代=3、40代=4、50代=5、60代=6、70代=7、80代=8
利用したプランの価格帯	5,000円ごとに7段階に分け、それぞれダミー変数とした
旅行相手の有無	1人(旅行相手なし)=1、2人以上=0
部屋数	各ホテルの総部屋数

男女や年齢によって生活習慣や好みが異なることが考えられるため、口コミ

の内容や表現方法にも差が生じる可能性がある。そこで、本研究では性別や年齢をコントロール変数として分析に含めた。また、宿泊プランの価格帯によって顧客の期待水準や満足度が変わることが予想されるため、価格帯もコントロール変数として設定している。さらに、宿泊施設の規模、すなわち部屋数の違いによって満足度に影響が及ぶ可能性があることから、部屋数もコントロール変数として投入した。

次に、順序ロジット回帰分析のモデルを示す。式は(2)の通りである。

$$\begin{aligned} \text{logit}(P(Y_1 \leq k)) \\ = \alpha_k - (\beta_1 \text{sex}_i + \beta_2 \text{age}_i + \beta_3 \text{price}_i + \beta_4 \text{travelpartner}_i \\ + \beta_5 \text{numberofrooms}_i + \beta_6 \text{topicindex1}_i + \beta_7 \text{topicindex2}_i \\ + \beta_8 \text{topicindex3}_i + \beta_9 \text{topicindex4}_i) \end{aligned} \quad \dots \dots \dots (2)$$

Y :星の数(1,2,3,4,5)

k :星の数の閾値

(2)式に基づき、順序ロジット回帰分析を実行し、推定された $\beta_1 \sim \beta_9$ について統計的な有意性を確認する。有意であった場合には、係数の値と符号に着目し、口コミ内容が個人の評点にどのような影響を及ぼしているかを解釈する。モデル全体の適合度について、順序ロジット回帰分析では通常決定係数 R^2 が直接計算できない。そのため本研究では、順序ロジット回帰分析のモデル全体の適合度を評価するために、擬似 R^2 (Pseudo R-Squared) を用いた検定を実施する。

4. 分析結果

4.1. トピック指数の作成

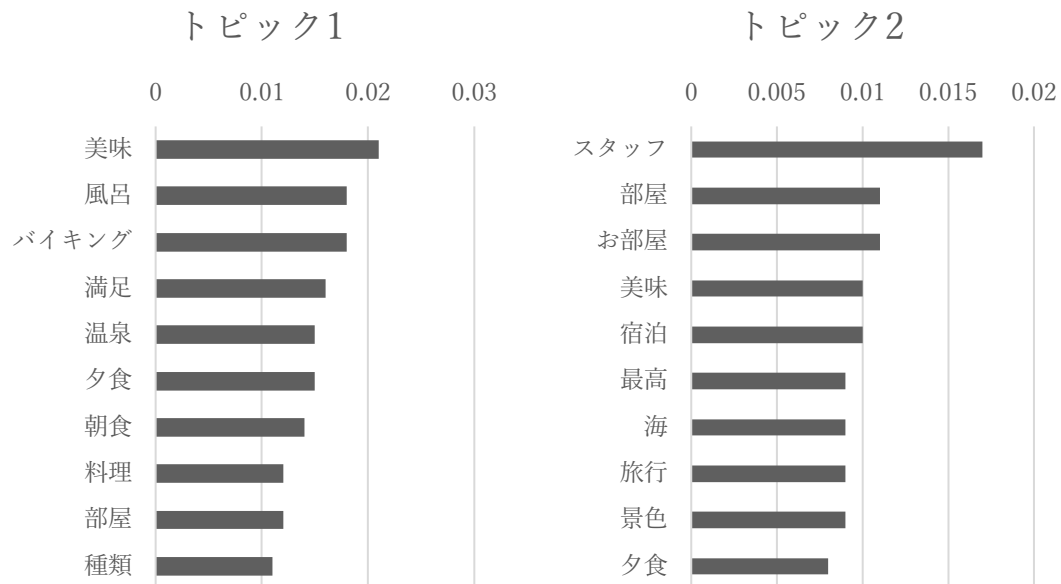
4.1.1. 形態素解析の結果

2,701 件の口コミデータに対して形態素解析を実行した。その結果、単語の種類数は 9,801 個、文字数の平均値は 157.085 文字、最大文字数は 995 文字、最小文字数は 37 文字であった。

表 4 形態素解析の結果

口コミデータ数	2,701 件
単語種類数	9,801 個
文字数平均値	157.085 文字
文字数最大値	995 文字
文字数最小値	37 文字

4.1.2. LDA の結果



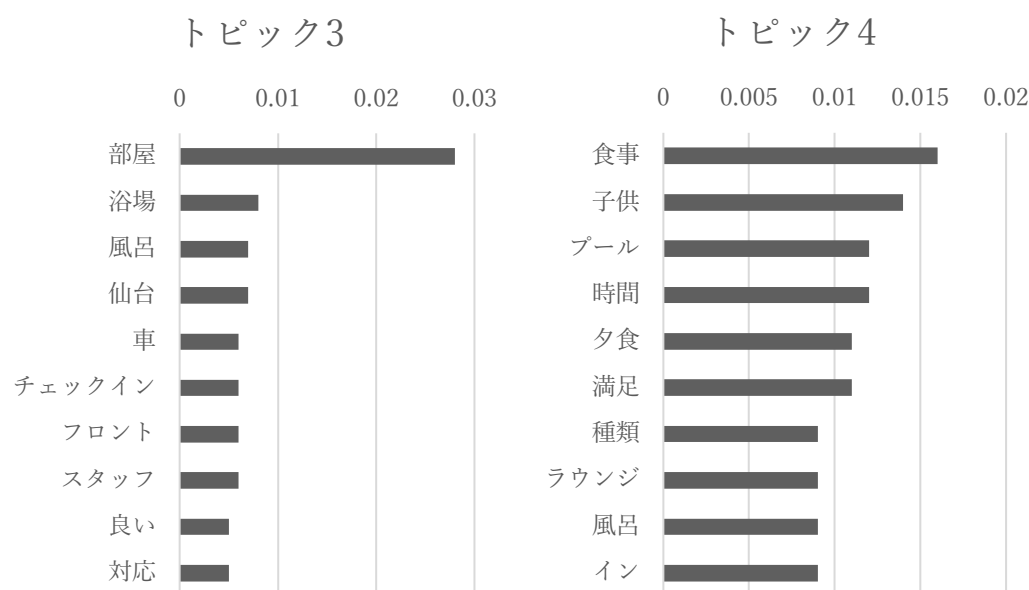


図3 各トピックにおける上位 10 単語

推定されたトピック番号（図に示すトピック 1、トピック 2、トピック 3、トピック 4）は、それ自体に特定の実質的な意味を持たない。そのため、各トピックがどのような特徴やテーマを持つかを明らかにするためには、トピック内で出現確率の高い単語に注目し、それらの単語からトピックの意味を考察する必要がある。この考察に基づき、各トピックに適切な名称を付ける（トピックの命名）ことで、トピックモデルの出力をより具体的かつ解釈可能な形にする。

トピック数の設定については、佐藤(2018)を参考にする。佐藤(2018)は、トピック数の選択方法として複数のアプローチを挙げた。第一に、未知データに対する予測力を高める観点から Perplexity を指標として用いる方法がある。Perplexity はモデルの予測精度を示す指標であり、この値が小さいほどモデルの予測精度が高いことを意味する。したがって、Perplexity が最小となるトピック数を選択することで、汎化能力の高いモデルを構築することができる。第二に、モデルの対数尤度が最大となるトピック数を選択する方法がある。この方法では、モデルがデータを最もよく説明できるトピック数が選ばれるため、データの特徴を適切に捉えることが可能となる。第三に、抽出されたトピックの解釈可能性を基準にトピック数を決定する方法も示されている。トピックモデルによって抽出されたトピックが分析者にとって理解しやすく、実際のデータの特徴と整合性が取れている場合、そのトピック数は適切であると判断される。この方法は、モデルの統計的な指標だけでなく、分析結果の解釈のしやすさを重視する点

で特徴的である。本研究では、解釈可能性に基づいてトピック数を選択する方法を選び分析した。トピック数に関して、本研究では解釈のしやすさとモデルの精度を考慮し、4つに設定した。理由としては、口コミデータ内の話題が最も明確に分類されるトピック数であると判断したからである。

図3から、トピック1は、美味・風呂・バイキング・満足・温泉などの言葉が上位に位置し、食事と風呂や温泉に関連するテーマであると解釈できることから、「食事(バイキング)・風呂(温泉)」と命名した。トピック2は、スタッフ・部屋・宿泊・最高などの単語が多く含まれており、宿泊体験における接客の質や、スタッフの対応に関連している。これを踏まえ、トピック2には「接客」命名した。トピック3は、部屋・浴場・風呂などが上位にきていることから「部屋(館内設備)」と命名した。トピック4は、食事・子供・プール・時間などの単語が抽出された。子供という単語とプールは家族連れの利用が多いことから家族で過ごす時間に関連するテーマであると解釈し、「家族との時間」と命名した。

このうえで、(1)の式に基づいて、トピック指数1(食事・風呂)、トピック指数2(接客)、トピック指数3(部屋)、トピック指数4(家族との時間)を作成した。

4.2. 順序ロジット回帰分析の結果

表5 順序ロジット回帰分析の結果

	Value	Std. Error	t value	p value
topic1	0.8149119480	0.208605820	3.9064679	9.365506e-05
topic2	1.9849179810	0.212979233	9.3197724	1.165901e-20
topic3	-2.977785435	0.240252173	-12.3943875	2.802789e-35
性別	0.0724111449	0.080745132	-0.8967865	3.698329e-01
年齢2	0.0909584794	0.180739915	0.5032562	6.147842e-01
年齢3	-0.7437784613	0.112901601	-6.5878469	4.462505e-11
年齢4	-0.6119062500	0.094086407	-6.5036627	7.838757e-11
年齢5	-0.9480604103	0.089003523	-10.6519425	1.707639e-26
年齢6	-1.0840597350	0.099179331	-10.9302989	8.257593e-28
年齢7	-1.1562666005	0.162262080	-7.1259200	1.033878e-12
年齢8	-0.0297192689	0.014066403	-2.1127838	3.461928e-02
~5,000 円	2.0429535870	0.694995998	2.9395185	3.287227e-03

5,001~10,000 円	1.4226256257	0.262334509	5.4229450	5.862502e-08
10,001~15,000 円	0.6280377645	0.215051832	2.9204018	3.495803e-03
15,001~20,000 円	0.7231571642	0.20985948	3.4459112	5.691373e-04
20,001~25,000 円	0.8036010182	0.226000555	3.5557480	3.769052e-04
25,001~30,000 円	0.8761227088	0.256730413	3.4126175	6.434217e-04
旅行相手	0.2149771973	0.167215278	1.2856313	1.985718e-01
部屋数	-0.0008334808	0.000387605	-2.1503354	3.152869e-02
1 2	-5.4564246319	0.262842475	-20.7592956	1.010368e-95
2 3	-3.9169897476	0.221284395	-17.7011567	4.107826e-70
3 4	-2.3213480866	0.210475667	-11.0290568	2.767523e-28
4 5	0.0343995210	0.208795991	0.1647518	8.691393e-01

表 6 McFadden の擬似 R²検定の結果

擬似 R ² (Pseudo R-Squared)	0.1157415
--------------------------------------	-----------

先で作成したトピック指数を用いて順序ロジット回帰分析を実行し、口コミ内容および投稿者の特徴と評点の関係性を検証した。分析の結果、モデル全体の適合度を示す McFadden の擬似 R²は 0.1157 であり、一定の説明力を持つモデルであることが確認された(表 6)。

トピック指数に関しては、トピック 1 およびトピック 2 が評点に対して 0.1% 水準で統計的に有意な正の値を示した。これに対し、トピック 3 は有意な負の値が確認された。

投稿者の属性については、性別は有意ではなかった。一方、年齢カテゴリ 3 (30 代) から年齢カテゴリ 7 (70 代) にかけては 0.1%水準で統計的に有意な負の係数が見られ、カテゴリ 8 (80 代) については 1%水準で統計的に有意な負の係数が見られた。また、価格帯については、すべてのカテゴリで統計的に有意な正の係数が確認された。旅行相手の有無については、統計的に有意な結果は確認されなかった。ホテルの特徴については、部屋数が評点に対して 5%水準で有意な負の影響を示した。カテゴリの閾値 (切片) については、星 1 から星 4 までの境界はすべて統計的に有意であった。しかし、星 4 と星 5 の境界 (4|5) については、有意な結果は得られなかった。

以上の結果により、口コミ内容や投稿者の属性と評点の関係性が明らかとなった。

5. 考察・仮説構築

本研究では、宿泊予約サイトの口コミデータをもとに順序ロジット回帰分析を行い、口コミ内容や投稿者の属性と評点の関係性を明らかにした。分析の結果を解釈・考察し仮説を立案する。

まず、食事・風呂・接客といったサービスや体験に関する言及が高評価につながることを確認された。特に、接客に関する言及が頻繁に見られる口コミは、星5や星4という高評価を得る可能性が非常に高いことが示された。このことから、顧客が宿泊施設を評価する際には、スタッフの対応やサービスの質が大きな決定要因となることがうかがえる。これは、接客に対する満足度が宿泊体験全体の印象を左右しやすいことを反映していると考えられる。宿泊施設においては、スタッフの接客態度やサービスの質を向上させることが、顧客満足度を高める可能性が高い。

一方で、部屋や館内設備に関する言及は、低評価に結びつきやすいことが明らかになった。特に、部屋の清潔感や大浴場の設備など、物理的な環境に関する不満が低評価の原因となることが多い。これは、設備面での不備や不快感が宿泊体験に直接的な影響を与えやすいことを示している。宿泊施設においては、部屋のメンテナンスや館内設備の維持管理に注力することで、顧客からの評価を向上させることができると考えられる。

また、投稿者の年齢によって評点に差が生じることも判明した。年齢が高くなるにつれて低評価をつける傾向が強く見られ、特に30代の投稿者においてその傾向が顕著であった。その理由として、ライフステージの違いが宿泊施設に対する評価基準や期待値に影響を与えているからであると考察する。30代は仕事や家庭生活が多忙であることが多く、限られた時間や予算の中で宿泊施設を利用するケースが増える。このため、宿泊体験に対する期待が高く、接客や設備に対して細かい点まで気にする傾向があると考えられる。さらに、子育て世代にとっては、子供連れの旅行が多く、施設の安全性や利便性に対する要求が高まることも低評価につながる要因となる可能性がある。その一方で、80代の年齢層では、低評価を下す傾向は他の年齢層と比較して弱まっており、有意ではあるもののその影響は小さかった。これは、宿泊施設に対する期待値が相対的に低くなり、旅行そのものを楽しむことに重点を置く姿勢が影響している可能性がある。

さらに、価格帯と評点の関係については、低価格帯から中価格帯の宿泊プランを利用した顧客の方が高評価をつける傾向が強いことが確認された。これは、低価格帯の宿泊プランでは価格以上の満足感が得られやすく、期待値を上回る体

験が評価に反映されやすいためであると推察される。一方で、高価格帯のプランにおいても高評価を得る傾向は見られたが、その度合いは低価格帯ほど顕著ではなかった。高価格帯では顧客の期待値が非常に高く、わずかなサービスの不備や期待との差異が評価に影響を与える可能性がある。

しかしながら、口コミ投稿者の性別および旅行相手の有無は、宿泊施設の評点に有意な影響を与えなかった。この結果は、宿泊施設の評価基準が性別や旅行スタイルに依存しにくいことを示唆している。宿泊体験における評価項目は、食事の質、部屋の快適さ、スタッフの対応など、性別を問わず共通して重視される要素が中心である。近年の宿泊施設では、すべての顧客に対して均一なサービスを提供することが求められており、特定の性別や旅行形態に特化した対応を行う機会は少ない。これにより、性別や同行者の有無による評点の差が生じにくくなっていると考えられる。また、宿泊施設の口コミは個人の体験に基づく評価が中心であり、旅行相手の存在よりも自分がどれだけ満足したかが評点に反映される傾向が強い。さらに、多くの施設では、一人旅から家族旅行まで幅広い顧客層に対応するプランが整備されており、同行者の違いが顕著な影響を与えることは少ないと推察される。

施設の規模は、顧客評価に影響を与える要因の一つであった。大規模な宿泊施設では、サービスの質や個別対応が行き届かないことが影響し、低評価を受けやすい傾向が示された。規模が大きくなるほど、顧客一人ひとりに対するきめ細かな対応が難しくなることが原因と考えられる。

また、評点が星4と星5の境界において曖昧さが存在することも示唆された。これは、口コミ投稿者が星4と星5を選ぶ際に明確な基準を持たず、主観的な感覚で選択している可能性があることを示唆している。

これらの結果を踏まえ、以下の6つの仮説が導き出された。

- 仮説1：食事と接客の質が宿泊施設の評点に大きく影響する
- 仮説2：部屋や館内設備に関する不満が低評価を引き起こす
- 仮説3：顧客のライフステージによって宿泊施設に求める要素が異なる
- 仮説4：低価格帯から中価格帯の宿泊プランが高評価を受けやすい
- 仮説5：施設規模が大きくなると低評価を受ける可能性が高まる
- 仮説6：星4の顧客は星5をつける潜在的な可能性を持つ

仮説1については、分析の結果、食事・風呂、接客といった項目が宿泊施設の評点に対して有意に正の影響を与えることが明らかとなったことから導き出し

た。特に接客に関するトピックが含まれる口コミは、高評価を得る傾向が顕著であった。このことから、宿泊施設においてはスタッフの接客態度や食事の質を向上させることが、顧客満足度を高める可能性が高いと考えられる。

仮説 2 については、部屋や館内設備に関するトピックは、評点に対して有意に負の影響を与えることが確認されたことから設定した。特に、清掃状況や部屋の老朽化などが指摘される口コミでは、低評価がつけられる傾向があった。これにより、宿泊施設における部屋や設備のメンテナンス、清潔さの維持が顧客の満足度向上に重要であると考えられる。定期的な施設の改修や清掃体制の見直しが、評点の向上につながる可能性がある。

仮説 3 については、年齢が高くなるほど低評価をつける傾向が見られ、特に 30 代から 60 代にかけてその傾向が強かったことから導いた。この結果から、ライフステージごとに宿泊施設への期待やニーズが異なり、これが評価に反映されることが示唆される。ライフステージの違いが宿泊施設の評価基準に与える影響は大きく、宿泊施設にとっては異なる年齢層ごとのニーズを把握し、それぞれに応じたサービスを提供することが求められる。特に、30 代以上の顧客層に対しては、細やかなサービス提供や設備の維持管理を強化することが高評価につながる重要な要素となると考える。

仮説 4 については、低価格帯および中価格帯の宿泊プランを利用した顧客が高評価をつける傾向が確認されたことから導き出した。宿泊施設においては、価格以上の付加価値を提供することが評点を向上させる鍵となることが考えられる。特に、低価格帯でも質の高い接客や設備を提供することで、顧客の期待を上回る体験が実現しやすくなると推察される。

仮説 5 については、部屋数が多い施設ほど評点が低下する傾向があったことから設定した。これは、大規模施設ではサービスの均質化や個別対応の難しさが影響している可能性がある。したがって、大規模施設においては効率的な運営のもとで顧客との接点を増やす工夫をして、可能な限り顧客ひとりひとりに対するきめ細やかな対応を強化し、サービスの質を向上させることが求められると考える。

仮説 6 については、星 4 と星 5 の境界が統計的に有意ではなく、両者の評価基準が曖昧であることが示唆されたため導いた。この結果は、星 4 をつけた顧客の多くがほぼ満足しており、あと一步で星 5 をつける可能性があるという状況にあることを示している。星 4 の評価は、基本的な満足が得られている一方で、何かが足りず些細な部分で不満が残るといった感覚が影響していると考えられる。このことから、宿泊施設においては、星 4 をつける顧客を星 5 に引き上げるための施策が、評点全体の向上に直結する。星 4 を選択した顧客は、すで

に高い評価を抱いており、サービスや体験の質がわずかに向上することで、次回以降に星 5 をつける可能性が高い。特に、顧客が感動や特別感を得られる場面をつくりだすことで、星 4 と星 5 の間にあるわずかなギャップを埋めることができる考える。満足から感動へと引き上げる体験の提供が、星 5 獲得の鍵となる。したがって、星 4 の顧客を星 5 に最も近い存在と捉え、顧客の期待を少しでも超える施策を講じることが、宿泊施設の評点向上および顧客満足度の最大化に寄与する重要な要素である。

6. 戦略立案

分析結果によって構築した仮説をもとに、家業のホテルである南三陸ホテル観洋の今後の戦略を立案する。立案にあたり自社に注目するため、本研究のデータで、南三陸ホテル観洋に宿泊した投稿者のみのデータを抽出した。

表7 データ概要

データ数	265 件
期間	2023 年 5 月 8 日～2024 年 10 月 31 日 (本研究に同じ)

評点の分布をみると、星 5 が半数以上を占めている。次いで星 4 が多く、全体の約 4 割となっている(図 4)。先でたてた仮説をもとに考えると、これらの顧客はあと一步の改善や付加価値の提供によって星 5 に引き上げられる潜在的な可能性を持つ層である。この層にアプローチをすることで、星 5 をつける顧客が 9 割を超えることになる。よって戦略立案の方向性を、星 4 の顧客層を星 5 に引き上げる施策の構築とする。

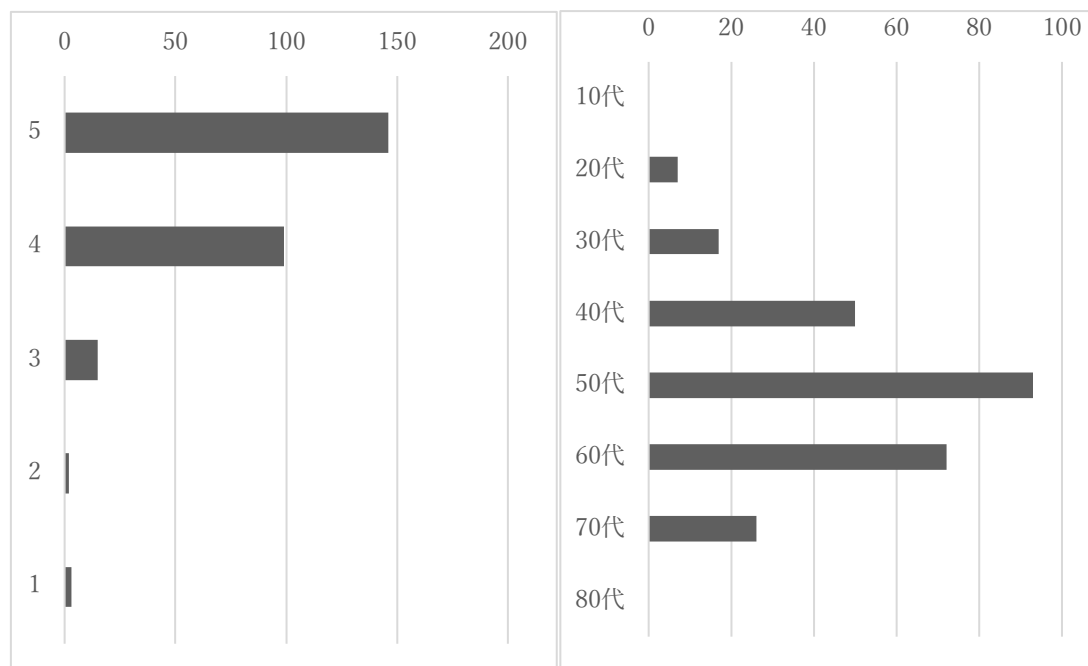


図4 評点(左)と年齢(右)の内訳

施策を考えるうえで、自社の現状の口コミ内容を分析する。先で構築した仮説はあくまで自社と競合をからみた宿泊施設全体の傾向である。そのため、戦略立案においては自社のみの傾向も知り、強みと弱みを必要がある。ここでは共起ネットワークを用いて分析を行う。共起ネットワーク分析は、膨大なテキストデータから単語間の関係性を可視化し、顧客がどのような要素に関心を持っているかを直観的に把握するための有効な手法である。この手法は、単語の出現頻度だけでなく、単語同士がどの程度共起しているかを示し、口コミデータに潜むテーマや特徴を明らかにする。共起ネットワークの最大の利点は、テキストデータにおける頻出単語のつながりを通じて、顧客が特に強く意識している要素や、重要視している項目を可視化できる点である。たとえば、宿泊施設の口コミデータに対して共起ネットワークを適用することで、部屋・食事・風呂・スタッフ対応といったキーワードの中から、どの単語が他の単語と強く結びつき、どのような評価を受けているかが明らかになる。これにより、宿泊体験の中で顧客が最も重視している項目や、評価に強く影響を与える要素を特定することが可能となる。

さらに、ポジティブな単語とネガティブな単語がどのように関連しているかを視覚的に把握できることも、共起ネットワークの重要な特徴である。具体的には、美味しい・快適といった単語が、食事・部屋と共起している場合、それらが高評価を得ていることが示唆される。一方で、古い・狭いといった単語が設備や部屋と共起している場合は、改善が求められている分野であると考えられる。このように、施設の強みと課題を同時に抽出することが可能となり、施設運営やサービス向上のための指針を得ることができる。

加えて、共起ネットワークは改善すべきポイントの優先順位を可視化する手助けとなる。たとえば、特定のネガティブな単語が他の単語と強く結びついている場合、それは顧客が不満を感じている要因が集中している領域であり、優先的に対応すべき課題として浮かび上がる。逆に、ポジティブな単語が強く共起する場合は、施設の競争力をさらに強化する方向性として活用できる。

共起ネットワークを活用することで、顧客の声をより具体的かつ網羅的に理解することが可能となり、サービスの改善や差別化戦略の立案に寄与する。特に、顧客満足度の向上や再訪意向を高めるための施策を検討する際に、感覚的な判断ではなくデータに基づく分析結果をもとに、的確なアプローチを構築することができる。

南三陸ホテル観洋の口コミ内容の共起ネットワークは図5の通りである。

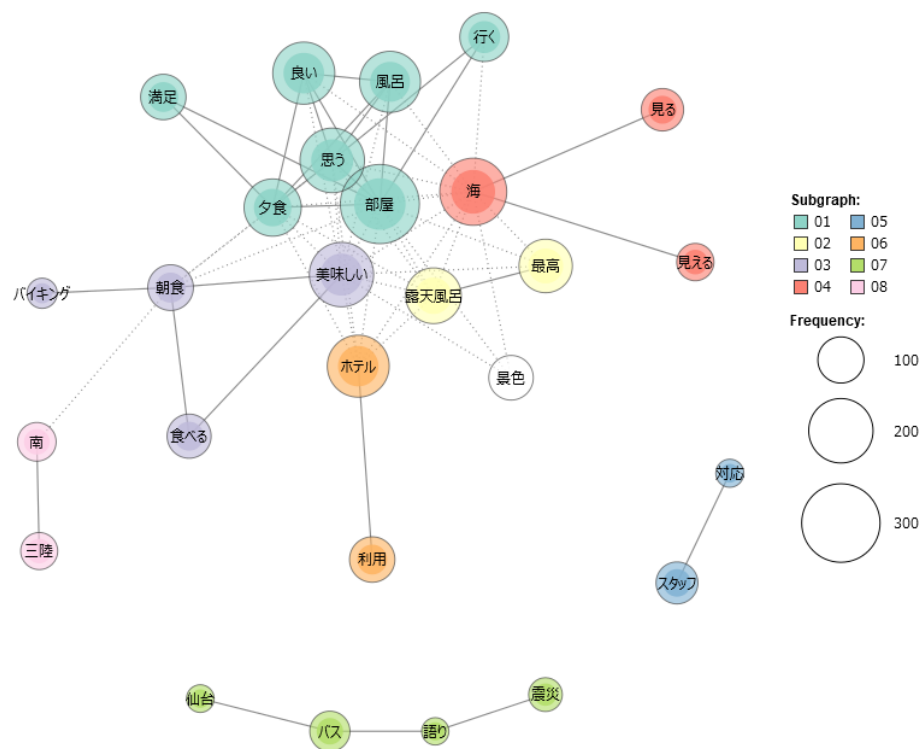


図5 共起ネットワーク分析の結果

共起ネットワーク分析の結果、南三陸ホテル観洋に関する口コミデータから、宿泊者が重視している要素や施設の特徴が明らかとなった。口コミに頻出する単語として、美味しい・部屋・海・朝食などが挙げられた。特に食事の質・露天風呂・海の景観が宿泊者の関心を集めていることが示唆された。

中核となる単語として、美味しい・部屋・風呂・夕食・最高などが頻出しており、これらが宿泊者の満足度に大きく寄与していることが分かる。また、海・見える・最高などの単語群がグループ化されており、海が見える景観が宿泊体験の魅力の一つであることが示されている。

サブグラフの分析からは、青グループでスタッフ・対応などの接客・サービスに関連する単語が強く結びつき、宿泊者がスタッフの接客を重視していることが確認された。また、緑グループでバス・仙台・雪などの交通手段に関する単語が共起し、アクセスや送迎に関する言及が一定数存在していることが示された。さらに、赤グループでは海・見える・景色などの景観に関する単語が目立ち、宿泊施設のロケーションが重要な評価要素となっていることがわかる。紫グルー

プでは朝食・バイキング・食べるといった食事に関する単語が多く、宿泊者が特に朝食バイキングを評価していることが明らかとなった。

この結果から、南三陸ホテル観洋の宿泊者が重視しているポイントは、食事の質・海が見える景観・スタッフの接客対応であることが判明した。一方で、現状の共起ネットワークでは顧客の満足を示すポジティブな単語が中心に配置されており、目立ったネガティブワードの共起は見られなかった。これは、宿泊者が施設の体験に対して概ね満足していることを示しており、食事・景観・スタッフ対応といった強みが顧客の記憶に残りやすいことがうかがえる。

仮説と共起ネットワークの結果をもとに、星4を星5にさせるための施策を提案する。仮説と分析から、顧客が高く評価している食事・風呂・接客の要素が評点に大きく寄与していることが明らかとなった。これらの強みを維持しつつ、さらなる満足度向上を図ることが、星5評価獲得に直結すると考えられる。

食事に関しては、先行研究および本研究の結果から宿泊施設の評価に強く影響する要素であることが示された。特に朝食のバイキング形式は顧客から好評を得ており、リピーター獲得にも寄与している。しかし現状では、夕食の献立は年10回ほど変更される一方で、朝食メニューの変更は年数回にとどまっている。南三陸ホテル観洋には2か月に1度の頻度で訪れる県内顧客が多く存在しており、リピーター層に対して新鮮な体験を提供し続けることが求められる。そのため、朝食メニューも定期的に変更し、季節感を反映させた多様なラインナップを用意することで、訪れるたびに新しい食事体験を提供することが可能になる。また、朝食バイキングではピーク時に席が確保できないケースが見受けられ、これが顧客体験に影響を与えている可能性がある。口コミには直接反映されていないものの、混雑の緩和は潜在的な顧客満足度向上につながる要因となりうる。この対策として、コロナ禍で実施していた朝食時間の事前予約制を繁忙期や週末に再導入することを提案する。これにより混雑が分散され、より快適な食事環境が提供されることで評点の向上が期待される。

風呂に関しては、共起ネットワーク分析の結果から風呂・露天風呂・景観に関する言及が多く、特に南三陸ホテル観洋のインフィニティ温泉は太平洋を一望できる景観という特徴が顧客の満足度を高める重要な要素であると示された。現在、大浴場は日帰り入浴終了後およびチェックアウト後に清掃が行われ、週1回は浴槽の湯を入れ替える体制が整っている。しかし、露天風呂は屋外に位置しているため蜘蛛の巣や落ち葉などが発生しやすく、景観を損なう原因が多数存在している。特に蜘蛛の巣は気づきにくく、清掃が行き届かないケースが見受けられる。これに対しては、従業員が昼間の空き時間に大浴場を実際に利用し、顧客目線で点検を行う仕組みを導入することを提案する。現在、

社員寮に住む従業員は業務終了後に大浴場を利用できるが、これを昼間にも適用することで、景観や清掃状況に関する気づきが増えることが期待される。実際に従業員が利用することで、細部の汚れや不備を見つけやすくなり、より徹底した維持管理が可能となる。

接客に関しては、本研究の結果から、食事や風呂以上に評点への影響が大きいことが明らかとなった。宿泊施設の規模が大きいため、スタッフと顧客との接点が限られていることが課題として挙げられる。接客の質を維持しつつ、顧客とスタッフが接触する機会を増やすことが、顧客満足度の向上に寄与すると考えられる。現在、南三陸ホテル観洋では夕食後にロビーでコンサートを開催するほか、屋上での星空観賞イベントスターパーティを実施しており、これらは顧客から高い評価を得ている。特にスターパーティは家族連れの参加が多く、施設の名物イベントとなっている。今後は、地域資源を活かした体験型イベントをさらに増やすことが求められる。具体的には、夏休み期間に実施する自由研究体験や地元特産品であるわかめの種付け体験など、地域の特色を活かしたワークショップを開催することが効果的であると考えられる。これらのイベントは宿泊者に特別な体験を提供するだけでなく、家族や子どもたちの思い出作りにもつながり、顧客満足度を高める重要な要素となる。また、大規模な宿泊施設では個別対応が難しいため、イベントを通じてスタッフと顧客が直接交流できる機会を増やすことで、施設全体のホスピタリティが向上し、結果的に評点の向上が期待される。

以上の施策を通じて、南三陸ホテル観洋の強みである食事・風呂・接客をさらに強化し、顧客満足度の向上を図ることで、星4評価の顧客を星5へと引き上げることが可能となる。本研究の分析結果からも示されたように、宿泊者はこれらの要素を重視しており、特に食事の質・露天風呂からの景観・スタッフの接客対応が宿泊体験の満足度を大きく左右していることが明らかである。

星4の顧客層は、施設に対して高い評価を持ちながらも、一部の要素において満点に至らない潜在層である。この層に対してきめ細やかなサービスを提供し、些細な不満を解消することで、星5評価への転換が期待できる。特に、50代・60代の宿泊者が全体の約6割を占めることから(図4)、この層のライフステージについて理解を深め、ニーズに応じた施策を展開することは、自社の強みをより一層引き出し、競合との差別化にもつながる。

共起ネットワーク分析を通じて得られた結果は、宿泊者がどのような点を重要視しているのかを視覚的に明らかにし、戦略立案に具体的な示唆を与えるものとなった。今後も継続的に口コミデータを収集・分析し、変化する顧客のニ

ーズを把握しながら、サービスの改善と新たな付加価値の創出に努めていくことが求められる。

このように、データに基づく戦略立案は、感覚的な意思決定を補完し、持続的な競争優位性を確保するための重要なアプローチである。南三陸ホテル観洋のさらなる成長と顧客満足度の向上に向けて、今回の分析を活かした実践的な施策の展開が期待される。

7. おわりに

本研究では、宿泊予約サイトに投稿された口コミデータを分析対象とし、宿泊者の評価や特徴と宿泊施設の評点の関係性を明らかにした。しかし、口コミデータは投稿者による主観的な情報に基づくものであり、旅行の目的などの属性情報は投稿時の自己申告に依存しているため、本研究にはいくつかの限界が存在する。年齢や性別に関しては、じゃらん.netの登録情報から自動的に反映される仕組みであるが、旅行の目的や感想は投稿者自身の判断で記載されるため、必ずしも正確に反映されているとは限らない。投稿者が実際の旅行目的を意図的に異なるものとして記入したり、記入ミスが生じたりする可能性があることから、口コミデータの信頼性には一定のリスクが伴う。これにより、旅行目的に関連する分析結果の精度や解釈に影響が及ぶ可能性がある。

また、口コミを投稿する動機には個人差があり、特に強い印象を抱いた宿泊者が投稿する傾向がある。これにより、非常に高評価または低評価といった極端な評価に偏りやすいバイアスが生じる可能性がある。加えて、満足度が高くても投稿の必要性を感じない層や、低評価を避けるために意見を控える層の存在により、必ずしも宿泊者全体の意見が反映されているわけではない。

さらに、口コミ内容は主観的な意見が中心であり、同じ体験をしても評価が異なることがある。特に、食事が美味しいやスタッフが親切といった感覚的な要素は個人差が大きく、同じサービスでも評価が分かれる可能性がある。このように、宿泊体験の客観的な評価が難しい点も、分析結果の普遍性に影響を与える要因となりうる。

本研究では口コミデータをもとに順序ロジット回帰分析を実施し、評点に影響を与える要因を検証したが、定性的な要素が数値化しづらいという限界がある。口コミの中には、投稿者の感情や文脈が含まれており、単語の出現頻度や共起関係だけでは完全に把握しきれない部分が存在する。

口コミの評点に影響を与える要因をより詳細に検討するために、投稿者の属性と口コミ内容の交互作用を組み込んだ順序ロジット回帰分析を試みた。しかし、モデルの収束が困難となる問題が発生し、クロスタームの詳細な検討が難しかった点も本研究の限界として挙げられる。具体的には、特定のカテゴリに属するデータ数が極端に少ない場合や、特定の変数が0または1に偏ることで、推定結果が不安定となる現象が確認された。特に、本研究においては、価格帯の変数を複数のダミー変数として扱ったため、多重共線性の問題が発生し、すべての価格帯を回帰モデルに投入すると冗長となることが判明した。このため、価格帯

の中から基準となるカテゴリを設定し、他のカテゴリとの相対的な比較を行う必要があったが、それでもなお収束の問題が生じた。また、トピック指数と年齢、トピック指数と価格帯といった複数の交互作用項を同時に投入した場合、モデルの自由度が減少し、計算が不安定になる傾向が見られた。特に、データの偏りによって、一部の交互作用項において推定値が極端に大きくなる、あるいは係数が適切に収束しないといった問題が生じ、適切なモデル構築が困難であった。したがって、本研究では交互作用の詳細な分析には限界があり、今後の研究においては、データの偏りを考慮したサンプルの再構築や、より適切なモデルの選択を通じて、交互作用の影響を精度高く検討する余地がある。

そして、本研究では一部の宿泊施設に特化した口コミデータを分析対象としているため、結果の一般化には限界がある。地域や施設の規模、顧客層の違いによって評価基準は大きく異なるため、他の宿泊施設や異なる地域で同様の結果が得られるとは限らない。

これらの限界を踏まえ、本研究の結果は、口コミデータを活用した一側面の分析として位置づけ、あくまで学術的な貢献よりも家業である南三陸ホテル観洋の今後の戦略を考えることに重きを置いた研究であると認識する必要がある。今後の研究では、宿泊施設が収集した顧客アンケートやフィードバックといった、より精度の高いデータを併用することで、分析の信頼性を向上させることが求められる。

本研究を踏まえて、顧客満足度向上のためには、顧客の声をより直接的かつ具体的に収集する仕組みが必要であると考えられる。特に、宿泊予約サイトの口コミデータは投稿者の主観やタイムラグが影響するため、リアルタイムかつ正確な顧客の声を補完するデータ収集が求められる。この課題に対応するために、フロントでタブレット端末を活用し、チェックアウト手続き後に簡易的なアンケートを実施する。また、音声入力機能を活用することで、顧客が直接意見を述べられる環境を整備する。音声入力の利用によって、文章の記入に不慣れな顧客や高齢者層も簡単にフィードバックを提供できる。これらの仕組みにより、従来の口コミデータに加えて、チェックアウト直後の顧客の生の声を定量・定性的に収集することが可能となる。リアルタイムで得られたデータは、宿泊施設のサービス改善やマーケティング戦略の立案において重要な示唆を与えると期待される。加えてサイトの口コミだけではなく紙のアンケートについても顧客のフィードバックとして今後分析を実施していく。

謝辞

本研究を進めるにあたり、多くの方々のご指導とご支援を賜りました。ここに深く感謝の意を表します。

まず、主査の木村太一先生には、テーマ決定の段階から一貫して親身にご指導いただきました。研究の方向性を見失いそうになった際にも、先生の温かい励ましと的確な助言が大きな支えとなり、最後まで論文を完成させることができました。

また、副査の井上哲浩先生には、マーケティングの視点からの貴重なアドバイスに加え、分析手法についても多大なるご指導を賜りました。先生の幅広い知見は、本研究の質を大きく向上させることに寄与しました。

さらに、副査の林高樹先生には、データ分析手法に関して具体的で実践的なアドバイスをいただきました。先生のご助言により、分析の精度を高めることができました。

本研究は、多くの方々のご協力のもと成り立っております。改めて心より感謝申し上げます。

参考文献

英語文献

- [1] Denove C. and Power J. D.(2006), *Satisfaction: How Every Great Company Listens to the Voice of the Customer*, Portfolio.
- [2] Ye, Q., Law, R., and Gu, B. (2009): *The impact of online user reviews on hotel room sales. International Journal of Hospitality Management*, 28(1):180-182.

日本語文献

- [3] 荒木匠平・小村亜唯子・平井裕久 (2022) .「旅行情報サイトにおけるランキングと口コミ内容の関係性」『日本経営工学会論文誌』,(73):15-26.
- [4] 一休.com ホームページ, <https://www.ikyu.com>(2024 年 6 月 6 日アクセス).
- [5] 鍛冶伸裕(2013).「日本語形態素解析とその周辺領域における最近の研究動向」『知能と情報(日本知能情報ファジィ学会誌) vol.25』,(6):174-183.
- [6] 観光経済新聞,「【データ】1位じゃらん net、2位楽天トラベル、8位 JTB 日本観光振興協会『2022 年観光関連サイト推計閲覧者数ランキング』」 <https://www.kankokeizai.com/> (2024 年 6 月 6 日アクセス).
- [7] 北中英明・香坂千佳子・伊藤宗彦(2018).「ホテル利用客の満足度に対するテキキストマイニングによる分析」『経営情報学会 全国研究発表大会要旨集』,120-123.
- [8] 久保田貴文(2016).「旅行情報サイトにおけるホテル口コミデータのテキキストマイニング」『経営・情報研究 多摩大学研究紀要』,(20):149-156.
- [9] 佐々木土師二(2000).『旅行者行動の心理学』関西大学出版部.
- [10] 佐藤圭(2018).「マーケティング研究におけるトピックモデルの適用に関する一考察」、『経営研究』,68(3):302-308.
- [11] 沢田史子 (2013) .「宿泊予約レビューデータを利用した宿泊者モチベーションの分析方法の提案」『北陸学院大学・北陸学院大学短期大学部研究紀要』,(6):227-283.
- [12] じゃらんホームページ, (<https://www.jalan.net>(2024 年 6 月 6 日アクセス).
- [13] 田邊亘・後藤正幸 (2008) .「宿泊施設の戦略構築を支援するユーザーレビュー分析に関する一考察」『武蔵工業大学環境情報学部情報メディアセンタージャーナル』,(9):91-101.

- [14] 内藤耕(2015).『サービス産業労働生産性の革新理論と実務』 旅行新聞新社.
- [15] 松河秀哉・大山牧子・根岸千悠・新居佳子・岩崎千晶・堀田博史(2018).「トピックモデルを用いた授業評価アンケートの自由記述の分析」,『日本教育工学会論文誌』,41(3):233-244.
- [16] 山本勲(2015).『実証分析のための計量経済学』 中央経済社
- [17] 楽天トラベルホームページ, <https://travel.rakuten.co.jp>(2024 年 6 月 6 日アクセス).