

# 宿泊施設評価データを用いた 宿泊施設の受賞要因分析

中央大学大学院 理工学研究科 ビジネスデータサイエンス専攻  
博士前期課程2年 マーケティングサイエンス研究室  
三村 崇彰



1. 研究背景
  2. 研究目的・方法
  3. 使用するデータ
  4. 分析
    - i. 分析概要
    - ii. 数値データの分析結果
    - iii. ユーザー投稿文の分析結果
  5. 考察
  6. まとめと今後の課題
- 参考文献

### 【目的】

宿泊施設の評価データを用いて、宿泊施設が賞を受賞する要因・受賞施設の特徴を明らかにする

- 旅行者の主観的な評価と宿泊施設が行っている受賞に向けた取り組みが、どのように受賞に影響を与えているか分析する
- 宿泊施設が受賞を目指す上で重視すべき内容を明らかにする

### 【方法】

1. 数値評価データを用いて、総合評価・受賞に与えている要因を分析
2. ユーザ投稿文を用いて、旅行者が何を評価したのか、受賞の要因を分析

## 使用するデータ

使用するデータ：楽天グループ株式会社から提供された楽天トラベルのデータ[8]

- ユーザー評価データ（約353万件）・・・2012年から2019年までに登録された評価データ  
データカラム：投稿者ID, 投稿日時, 施設ID, プランID, プラン名, 部屋種類, 部屋名, 目的, 同伴者, 評価1（立地）, 評価2（部屋）, 評価3（食事）, 評価4（風呂）, 評価5（サービス）, 評価6（設備）, 評価7（総合）, ユーザ投稿本文, 施設回答本文

user_21	2019-07-06 22:28:17	162643 3954015.0	【朝食付き】スタンダードプラン<宮崎名物を取り入れた和洋食バイキング>	◆禁煙ダブルルーム15平米◆	レジャー人	4 5 5.0 5.0 5.0 5.0 5	部屋は広くて綺麗、お風呂も非常に滑らかな肌触りで満足しました。また宮崎に行ける機会があれば是非...	この度は「天然温泉日向の湯 ドーミーイン宮崎」をご利用いただきまして、誠に有難うございます。...
---------	---------------------	------------------	-------------------------------------	----------------	-------	-----------------------	--	---

- 施設名データ（約2.9万施設）・・・施設IDと施設名を結びつけるマスターデータ  
データカラム：施設ID、施設名

施設ID	施設名
87	ベルビューガーデンホテル関西空港
89	東京ビジネスホテル
90	ホテルたてしな
91	ホテルニューオータニ大阪

### 温泉大賞

ユーザー投票によって、日本全国から高評価を得た温泉宿と温泉地を決定する（主催 ビッグローブ株式会社）

#### 選定理由

1. 2008年から毎年開催されており、受賞の**バックナンバー**も掲載
2. 2012年以降**同じ基準で公表**（そのため、本研究では2012年以降の結果を使用）
3. ユーザー投票により決定されており、**受賞プロセスが明確**かつ利用者目線の評価が行われている

本研究では、温泉地部門における各年の東西各上位10施設を受賞施設として分析を実施

### <データの前処理>

1. 分析対象にする賞を温泉大賞にしたため、利用目的が**レジャー**である件数の割合が**90%以上の宿泊施設**を分析対象とした
2. データにカラム[award]を追加 (0 : 非受賞, 1 : 受賞)

### <分析>

1. 数値評価データの分析 (Pythonを使用)
  - i. 重回帰分析
  - ii. ランダムフォレスト (アンダーサンプリングを実施)
2. ユーザー投稿文の分析 (**Text Mining Studio**を使用)
  - i. 単語頻度解析
  - ii. 評判抽出
  - iii. 特徴語・特徴表現抽出



## 分析：数値評価データの分析

### <数値評価データの分析>

5段階の数値評価データを用いて、総合評価・受賞に与えている要因を分析

### <分析内容>

i. 重回帰分析

目的：総合評価に影響を与える評価項目を明らかにするため

ii. ランダムフォレスト（アンダーサンプリングを実施）

目的：数値データのみでの分類を実施し、受賞要因の特定を行うため

## 分析：重回帰分析

### <目的>

総合評価に影響を与える評価項目を明らかにするため

説明変数：各評価項目（立地、部屋、食事、風呂、サービス、設備）

目的変数：総合評価

	coef	std err	t	P> t
const	-0.1540	0.025	-6.071	0.000
location	0.0989	0.005	19.031	0.000
room	0.1848	0.005	40.361	0.000
bath	0.2365	0.004	52.735	0.000
meals	0.1920	0.005	35.911	0.000
service	0.2873	0.005	55.400	0.000
facilities	0.0500	0.005	9.340	0.000

受賞宿泊施設（決定係数：0.692）

	coef	std err	t	P> t
const	-0.0378	0.003	-11.655	0.000
location	0.0926	0.001	129.030	0.000
room	0.1950	0.001	272.478	0.000
bath	0.2536	0.001	374.302	0.000
meals	0.1148	0.001	172.499	0.000
service	0.3111	0.001	398.262	0.000
facilities	0.0636	0.001	82.447	0.000

非受賞宿泊施設（決定係数：0.734）

図2：重回帰分析の結果

風呂とサービスの評価が総合評価に与える影響が相対的に大きいと言える

## 分析：ランダムフォレスト（アンダーサンプリング）

### <目的>

数値データのみでの分類を実施し、受賞要因の特定を行うため

### <方法>

1. 宿泊施設ごとに受賞・非受賞に分け、評価項目（立地，部屋，食事，風呂，サービス，設備）の**評価の平均値**を算出
2. 非受賞施設を受賞施設の2倍の数にするために，**アンダーサンプリング**を実施
3. 4つに分割し、クロスバリデーションを実施

### <結果概要>

表1：ランダムフォレストの結果

<b>Accuracy (mean)</b>	<b>0.75</b>
<b>F1 Score (mean)</b>	<b>0.61</b>
<b>Precision (mean)</b>	<b>0.62</b>
<b>Recall (mean)</b>	<b>0.60</b>

## <混同行列>

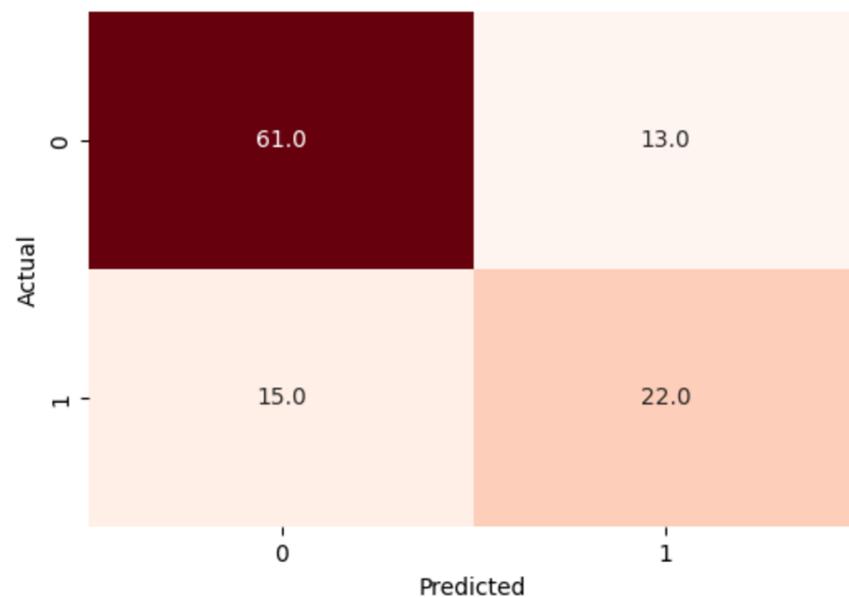


図3：ランダムフォレストの結果（混同行列）

## <特徴量重要度>

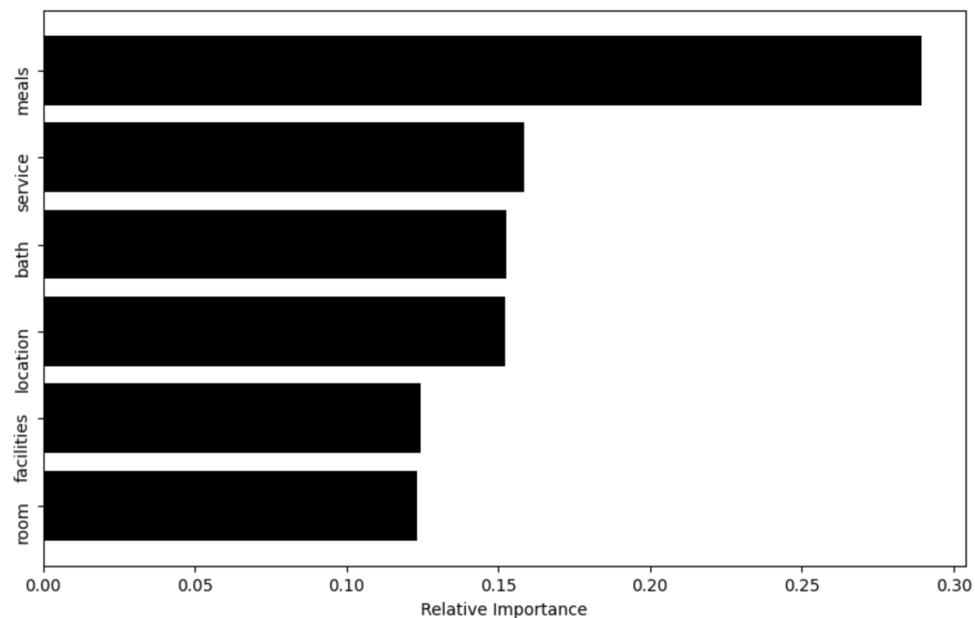


図4：ランダムフォレストの結果（特徴量重要度）

評価項目：**食事**の特徴量重要度が最も高い

## 分析：ユーザー投稿文の分析

### ＜ユーザー投稿文の分析＞（Text Mining Studioを使用）

受賞宿泊施設の評価ポイントや宿泊施設の強みを分析するために**受賞宿泊施設のユーザー投稿文のみを用いて**，を用いて，分析を実施

#### ＜分析内容＞

i. 単語頻度解析

目的：どのような単語が多いかを知ることを通じて本データの特徴を把握するため

ii. 評判抽出

目的：受賞施設の口コミにおける好評語や不評語を分析するため

iii. 特徴語・特徴表現抽出

目的：受賞施設に高評価をつけている人の口コミを用いて，特徴的な評価内容を分析するため

## 分析：ユーザー投稿文の単語頻度解析

ユーザー投稿文の単語頻度解析の結果の上位25語

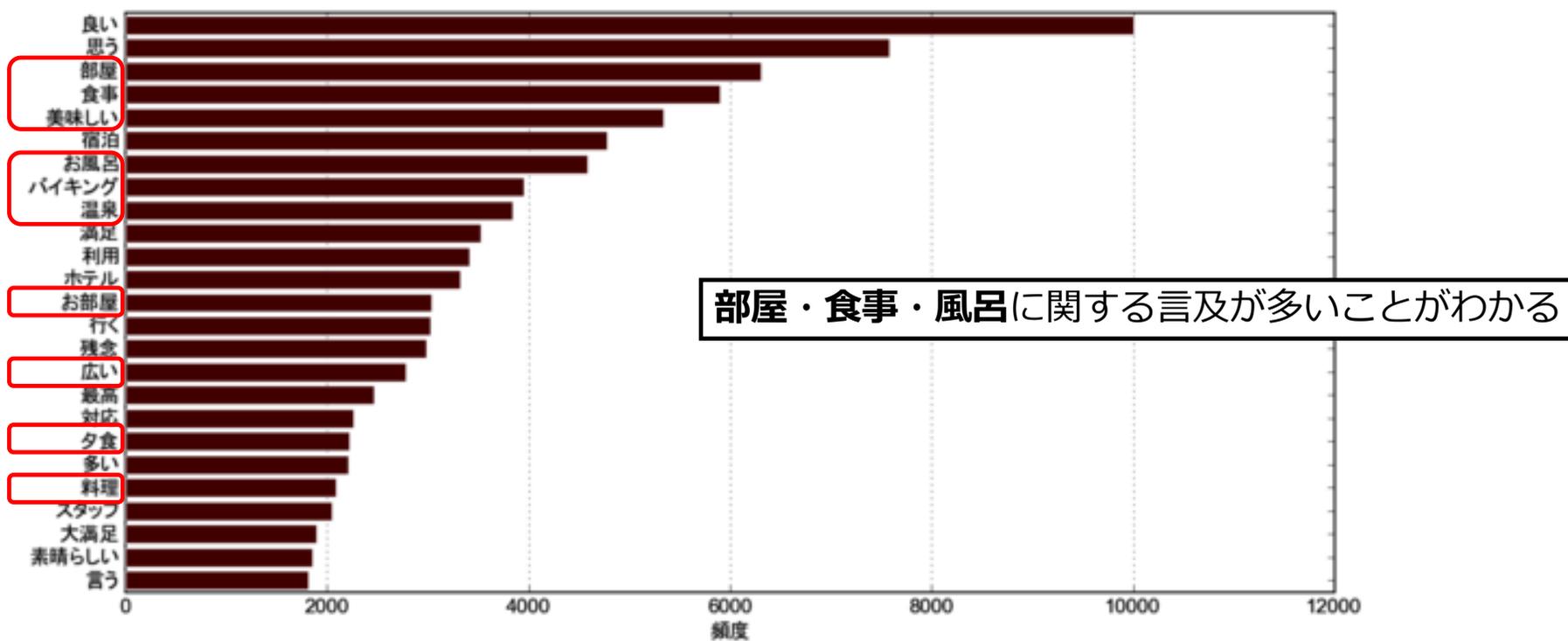


図5：ユーザー投稿文の上位語

## 分析：評判抽出（好評語順）

### 評判抽出

いいイメージで語られる言葉・悪いイメージで語られる言葉を抽出  
ユーザー投稿文の評判分析を実施（好評語順）

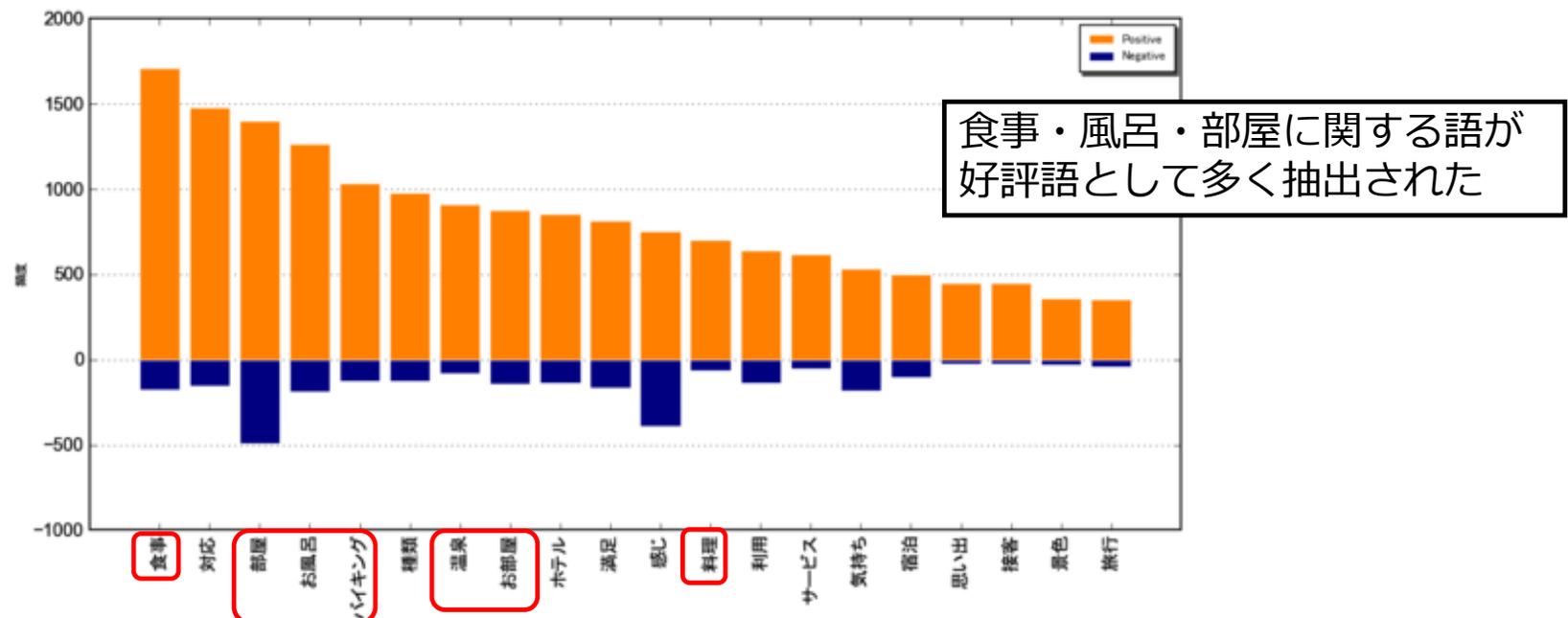


図6：ユーザー投稿文の好評語の評判抽出の結果

## 分析：評判抽出（不評語順）

ユーザー投稿文の評判分析を実施（不評語順）

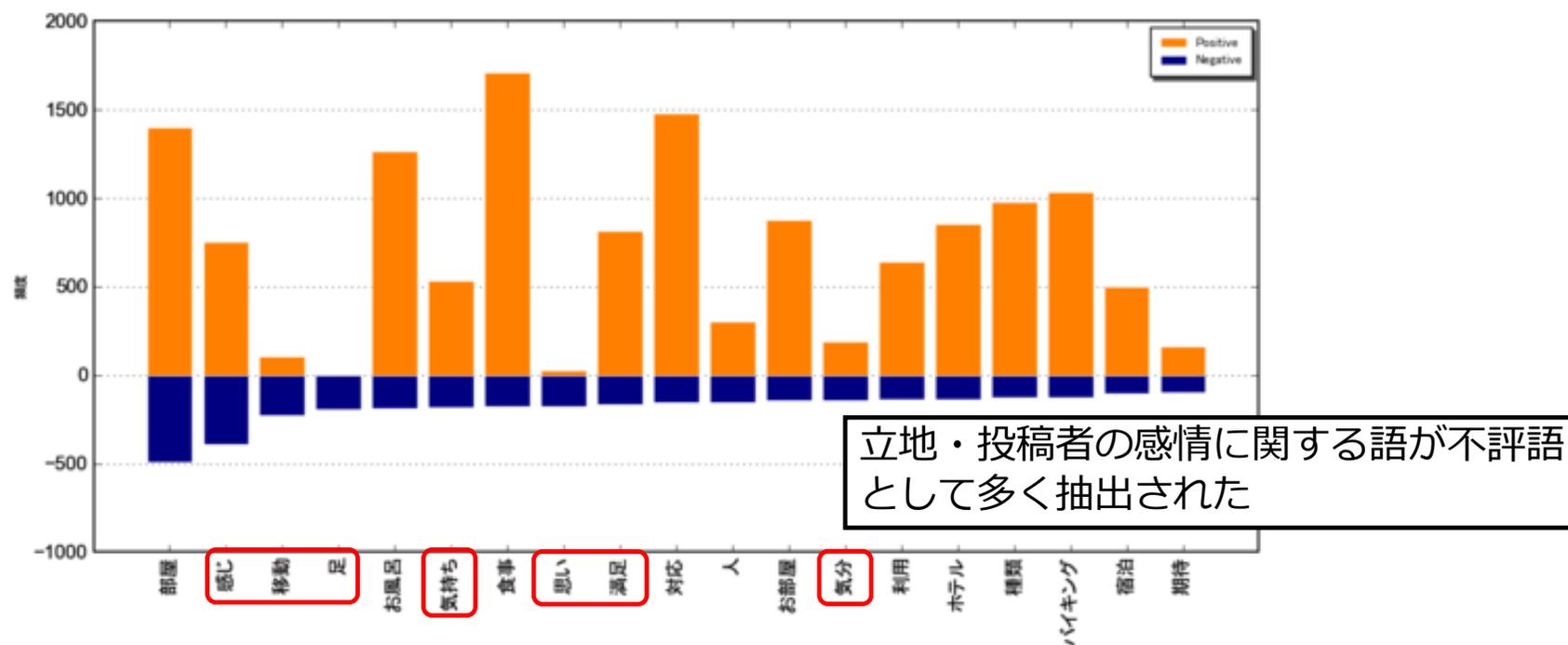


図7：ユーザー投稿文の不評語の評判抽出の結果

## 分析：特徴語の抽出

### <特徴語の抽出（高評価）>

データに付随する属性(今回は総合評価の値)ごとに特徴的に出現する単語を抽出

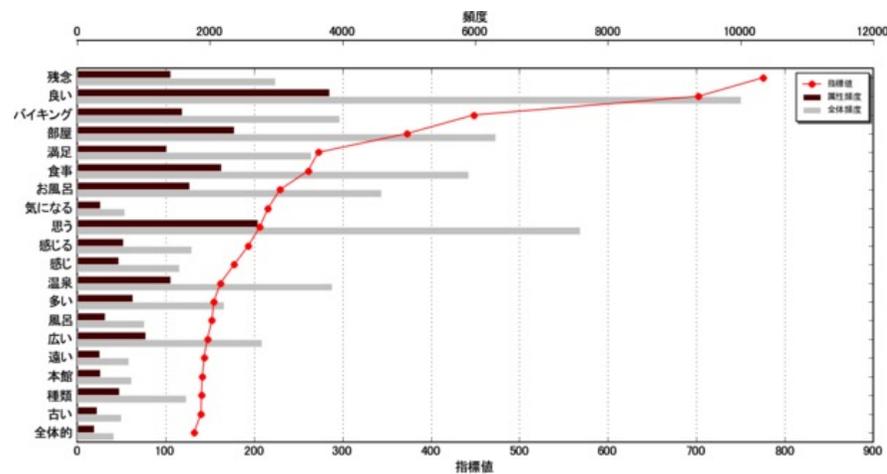


図8：ユーザー投稿文の特徴語抽出の結果  
(総合評価5)

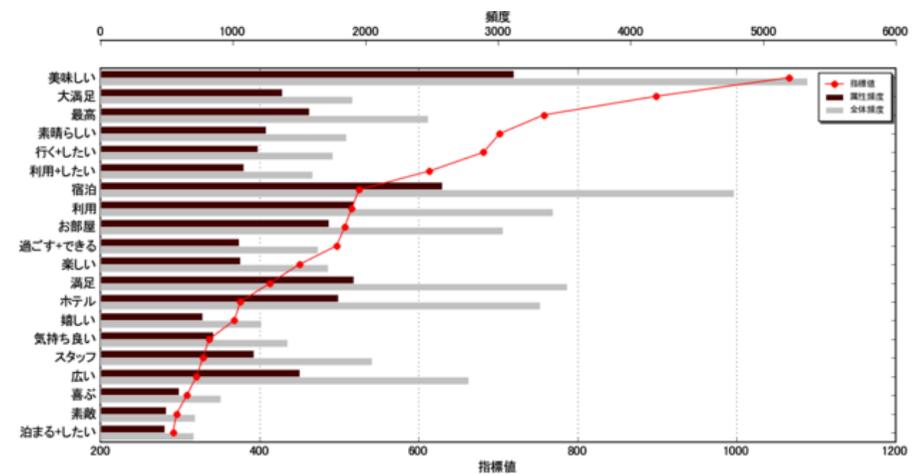


図9：ユーザー投稿文の特徴語抽出の結果  
(総合評価4)

## <特徴語の抽出（低評価）>

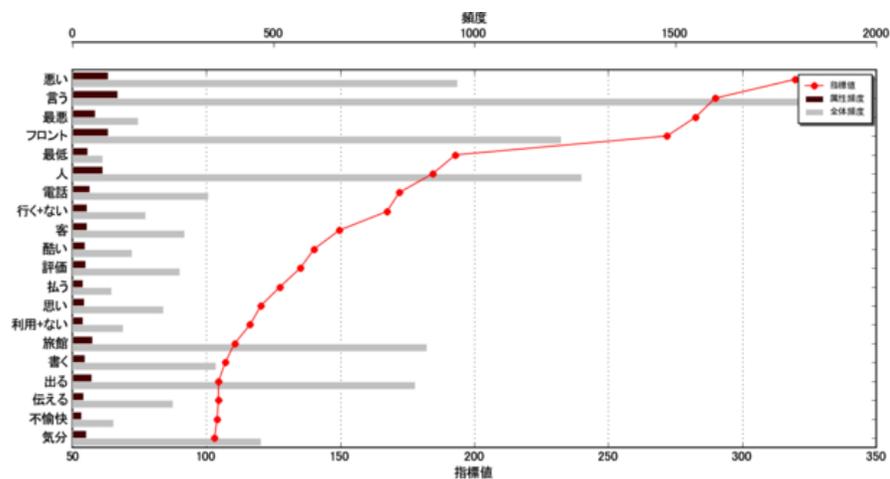


図10：ユーザー投稿文の特徴語抽出の結果  
(総合評価1)

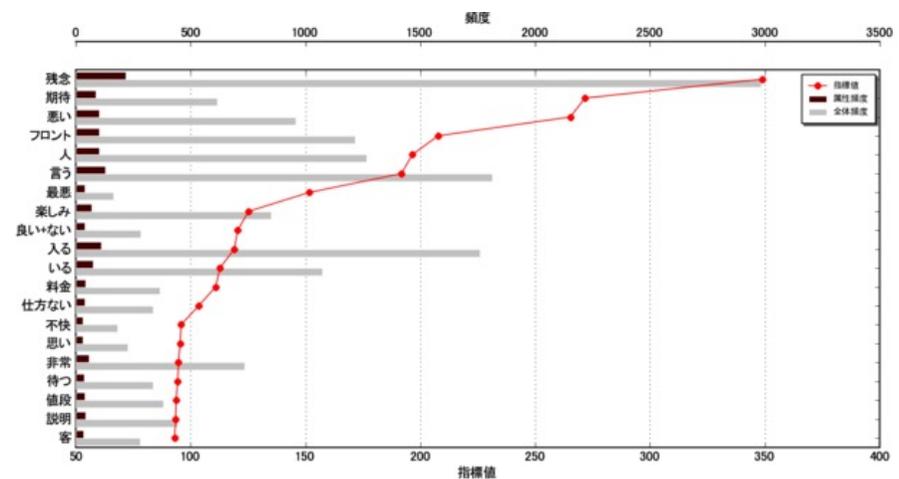


図11：ユーザー投稿文の特徴語抽出の結果  
(総合評価2)

投稿者の気持ちに関する語が多く抽出されたc

## 分析：特徴表現の抽出

### <特徴表現の抽出（高評価）>

データに付随する属性(今回は総合評価の値)ごとに特徴的に出現する係り受け表現を抽出

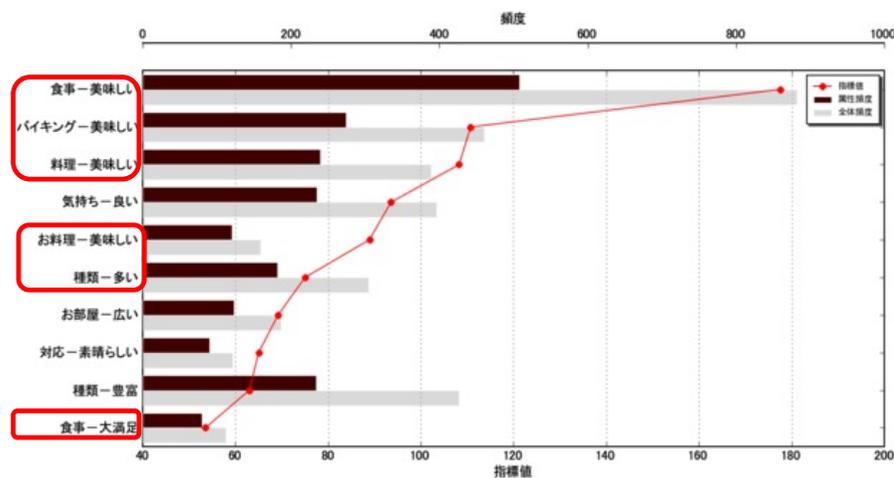


図12：ユーザー投稿文の特徴表現抽出の結果  
(総合評価5)

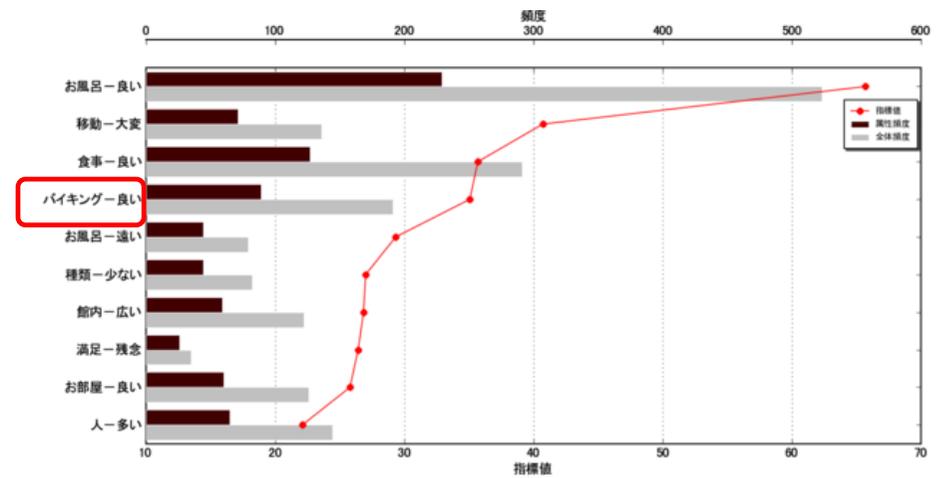


図13：ユーザー投稿文の特徴表現抽出の結果  
(総合評価4)

食事に関する語が多く抽出された

# 分析：特徴表現の抽出

## <特徴表現の抽出（低評価）>

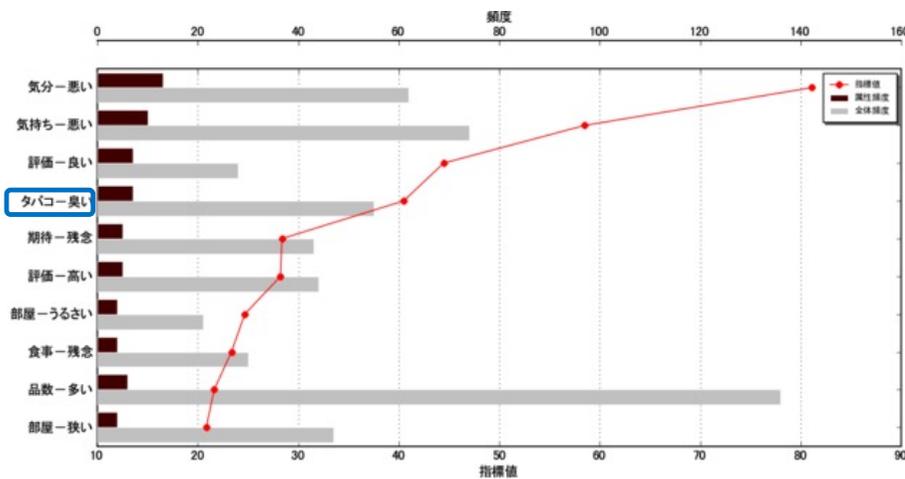


図14：ユーザー投稿文の特徴表現抽出の結果  
(総合評価5)

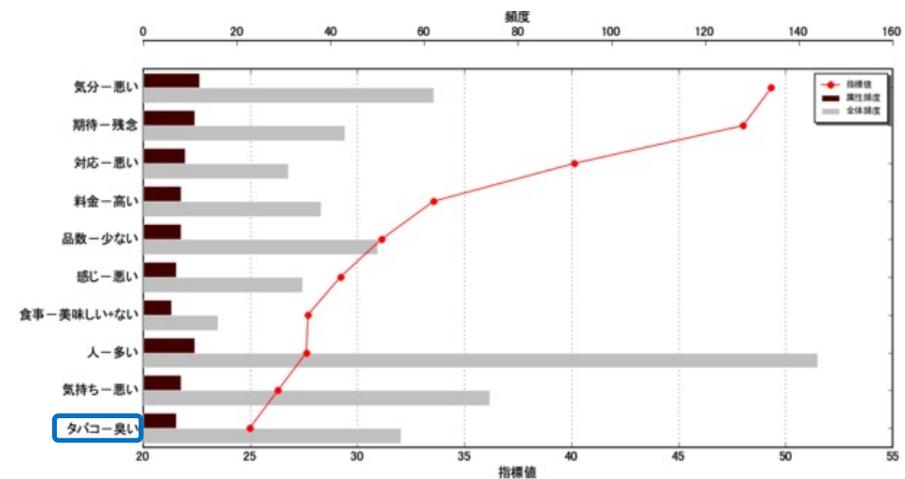


図15：ユーザー投稿文の特徴表現抽出の結果  
(総合評価4)

他の結果にはなかった“タバコ臭い”が抽出された

## 考察：重回帰分析

- ▶受賞施設・非受賞施設のどちらの結果も**風呂・サービス**の係数が相対的に高い
  - ⇒**風呂・サービス**は受賞・非受賞に関わらず顧客を満足させることができ、総合評価に与える**影響が大きい**と考えられる
  
- ▶受賞施設・非受賞施設のどちらの結果も**立地・設備**の係数が相対的に低い
  - ⇒**立地・設備**は受賞・非受賞に関わらず、総合評価に与える**影響が小さい**
  
- ▶受賞施設と非受賞施設では**食事の係数の差**が最も大きい
  - ⇒**食事**が受賞・非受賞を分ける**一番の要因**となっていると考えられる

## 考察：ランダムフォレスト

### ▶ **食事**の評価が分類に与える**影響が大きい**

⇒食事は見た目・味など**評価しやすく**、**好みも分かれる**ため数値で評価しやすいと考えられる  
多くの人を満足させることができる食事を提供することは受賞につながると言える

### ▶ **風呂・立地**の評価が分類に与える**影響が小さい**

⇒実際に利用した人自身が**風呂（温泉）の差を感じる**ことが**難しい**ことが影響していると考えられる  
温泉を目的とする時、宿泊施設や温泉地が目的地になることが多く、**立地を考慮した選択は行われにくい**  
と言える

➤ **食事**（“食事”，“バイキング”，“料理”）・**風呂**（“お風呂”，“温泉”）・**部屋**（“部屋”，“お部屋”）に関する語が好評語として多く抽出された。また，好評語として抽出された**“種類”**についても**食事の種類や風呂の種類**について述べられてるものが多くあった。

⇒ 食事・風呂・部屋が好評の内容として述べられ，受賞の要因になったと考えられる

また，**種類を多くすることによって，多くの人を満足させる**ことができ，ポジティブな投稿が増えることにつながりやすい

➤ **立地**（“移動”，“足”）・**投稿者の感情に関する語**（“感じ”，“気持ち”，“思い”，“満足”）が不評語として多く抽出された

⇒ 温泉地は必ずしも**交通の便がいいところにあるとは限らない**ため，受賞宿泊施設であったも不評を述べられる要因となった

受賞宿泊施設に対して，**期待しすぎたこと**によって**不評を述べた投稿**があった可能性が考えられる

## 考察：特徴語・特徴表現の抽出

- ▶総合評価5をつけている投稿文に対して特徴表現の抽出を行った結果では、**食事について述べた表現が上位10表現中5つを占める**など特徴表現の抽出においても食事に関する語句が多く抽出された
  - ⇒評判抽出の結果と同様に、食事については高評価をつけた人が多く述べており、**高評価をつける要因**になったと考えられる
  
- ▶**低評価をつけた特徴表現**の結果で抽出された表現の中で、“**タバコ臭い**”が特徴的なものとして現れた。
  - ⇒投稿者が宿泊中に**タバコの臭いに不快感**を、非喫煙者にとって宿泊施設やその部屋の**清潔さや快適さ**に対する**否定的な評価**が述べられていると考えられる。

- **風呂・サービス**は受賞・非受賞に関わらず顧客を満足させることができ、利用した人自身が**風呂（温泉）の差を感じる**ことが難しいことが影響していると考えられる
- **立地・設備**は受賞・非受賞に関わらず、温泉地は必ずしも**交通の便がいいところにあるとは限らない**ため総合評価に与える**影響が小さく**、**立地を考慮した選択は行われにくい**と言える
- **食事**は、見た目・味など**評価しやすく**、**好みも分かれる**ため数値で評価しやすいため、受賞・非受賞を分ける**一番の要因**となっていると考えられる。また、**種類を多くすることによって**、**多くの人を満足させる**ことができ、受賞につながると言える
- 受賞宿泊施設に対して、**期待しすぎたこと**によって**不評を述べた投稿**があった可能性が考えられる
- 宿泊中に**タバコの臭い**に**不快感**を、非喫煙者にとって宿泊施設やその部屋の**清潔さや快適さ**に対する**否定的な評価**が述べられていると考えられる。

### <まとめ>

- **食事**は評価しやすい項目であり、**受賞を分ける一番の要因**と明らかになった
- 風呂（温泉）の差を感じるものが難しく、**風呂の内容が評価に与える影響**は低い
- また、食事や風呂の**種類を充実**させることによって多くの人を満足させやすくなり、**受賞に繋がりやすい**
- **立地**は**受賞に影響を与えない**評価項目である

### <今後の課題>

食事や風呂の**種類の充実**は、宿泊施設の広さや部屋数、収容人数など**宿泊施設の規模**によって充実させられる度合いが変わることが想定されるため、それらを考慮した分析を行う必要があると言える

- [1]室岡祐司,「ツアーグランプリ」受賞商品の特徴と旅行商品の差別化要因の探索的研究, 第37回日本観光研究学会全国大会学術論文集 (2022年12月) pp.135-140
- [2]荒木匠平、小村亜唯子、平井裕久, 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, Vol.73 pp.15-26
- [3]ビッグロブ株式会社, 温泉大賞, <https://travel.biglobe.ne.jp/onsen/award/>
- [4]落合夏美、川合康央, 宿泊施設のユーザーレビューによる地域分析, 情報処理学会第83回全国大会, pp.528-529
- [5]田邊亘、後藤正幸, 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, 情報メディアセンタージャーナル2008.4 第9号, pp.91-101
- [6]山本雅人、大内東, 宿泊施設の分類システムと自己評価, 人工知能学会誌 26巻3号, pp.272-279
- [7]企業表彰が株価・人気ランキングに与える影響, 日本労働研究雑誌 No.538, pp43-58
- [8]楽天グループ株式会社 (2020): 楽天トラベルデータ. 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.2.2>

# Appendix

## VIFの値

重回帰分析を行った各変数間のVIF値は以下となる

Feature	VIF
const	39.720187
location	1.342653
room	1.839917
meals	1.493636
service	1.907146
facilities	2.146440

非受賞宿泊施設

Feature	VIF
const	52.853892
location	1.324424
room	1.594919
meals	1.350239
service	1.909366
facilities	1.984643

受賞宿泊施設

以上の結果からどちらも多重共線性がなく、係数の推定は正しく行えていると言える