

株式会社NTTデータ数理システム 学生奨励賞応募論文

宿泊サイトのレビュー評価値と口コミの信頼性

-新たな評価項目を定義する-

法政大学工学部創生科学科

呉研究室 久貝航平

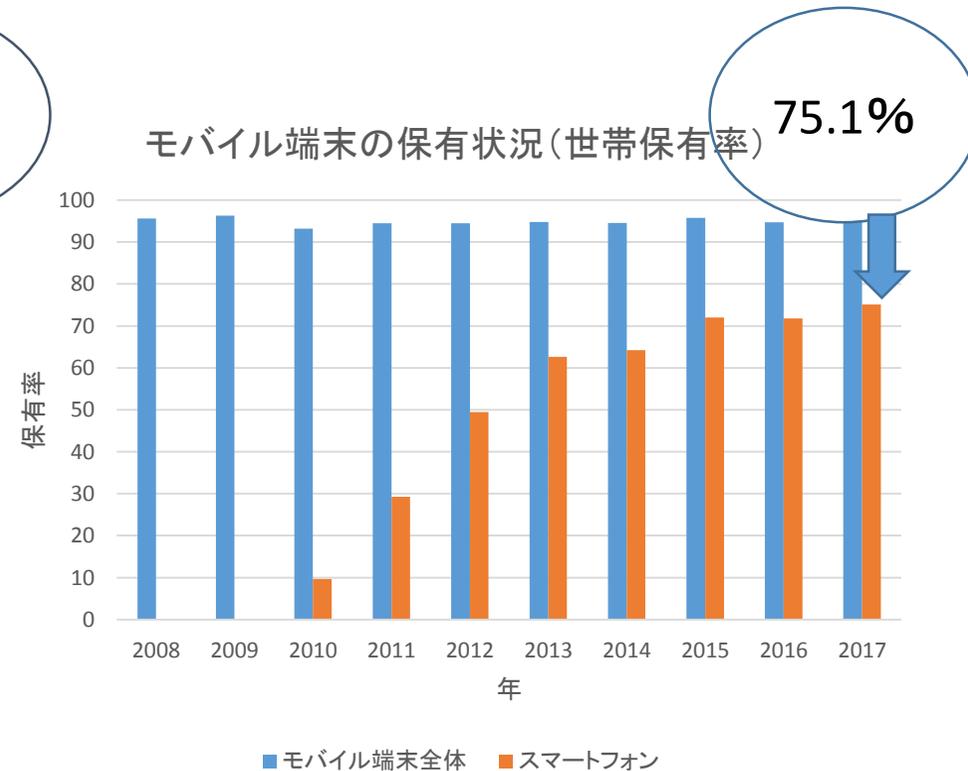
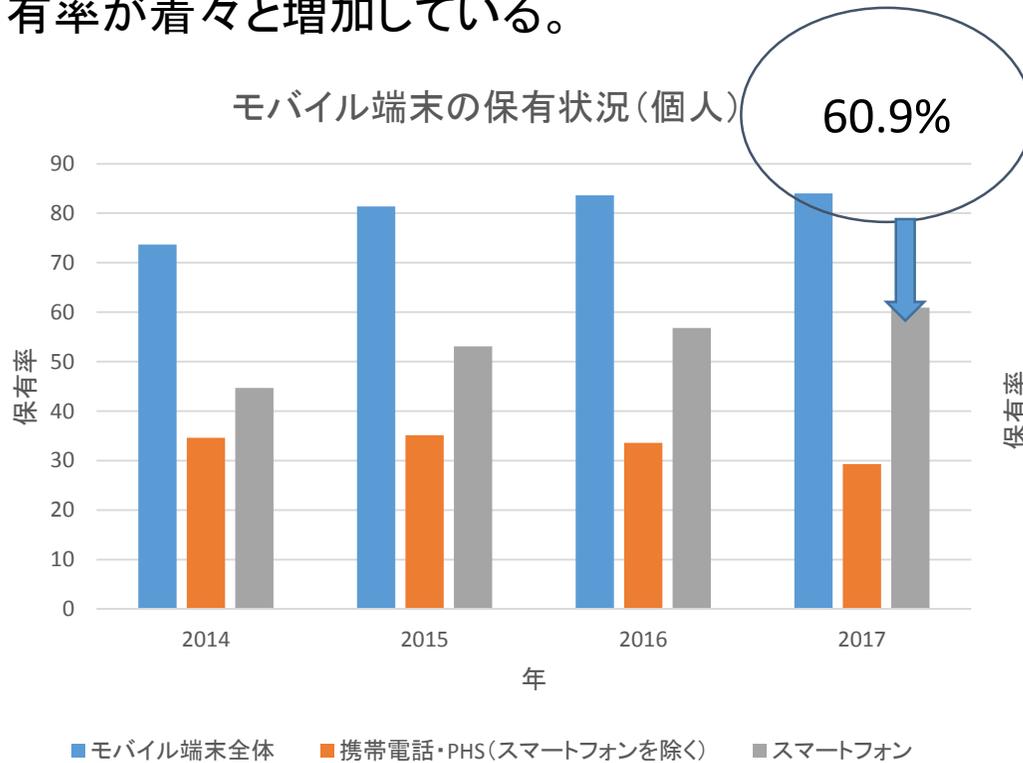
目次

- 1.研究背景
 - 2.研究目的
 - 3.本研究のデータ収集
 - 4.分析手法
 - 5.分析結果
 - 6.考察
 - 7.結論
 - 8.今後の課題
- 参考文献

1. 研究背景

1-1. インターネット利用率・スマートフォン保有率の現状

- 2017年、インターネットの利用率(個人)が8割を超えた。(総務省, インターネットの利用状況)(注①)
- そして総務省の情報通信機器の保有状況を見ると, モバイル端末全体及びスマートフォンの保有率が着々と増加している。



1-2. マーケティングにおける口コミの重要性

インターネット, モバイル端末の保有率の増加に伴い

オンラインショッピングサイトや,
旅行やレストランなどの予約サ
イトの増加

自由に書き込むことのできるSNSなどの
ソーシャルメディアやユーザー投稿型の
評価サイトが数多く出現

購入, 予約前に実物
などを見ることができ
ない

判断基準が必要

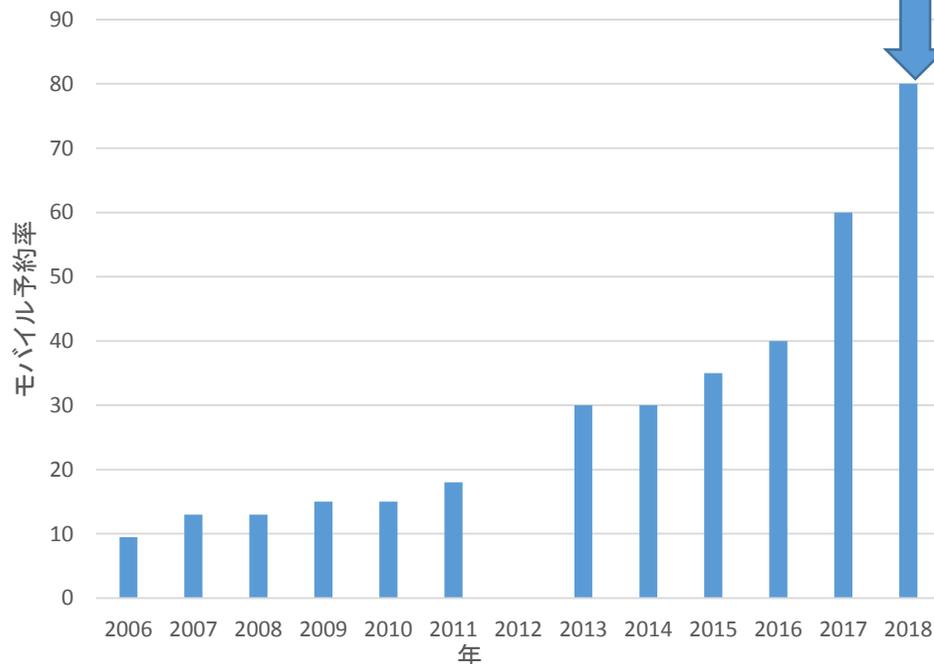
- 1, 消費者の信頼を得やすく購買行動の手助け。
- 2, 口コミから情報を抽出し, マーケティング活動に役立てる.
サイト側は購入者・利用者に口コミを求め, 掲載する

1-3. 宿泊予約サイトの実態調査

- ・観光経済新聞の宿泊予約サイト実態調査によると旅行業界においても宿泊予約サイトを用いた予約が増加している。(注②)

80%のモバイル予約率

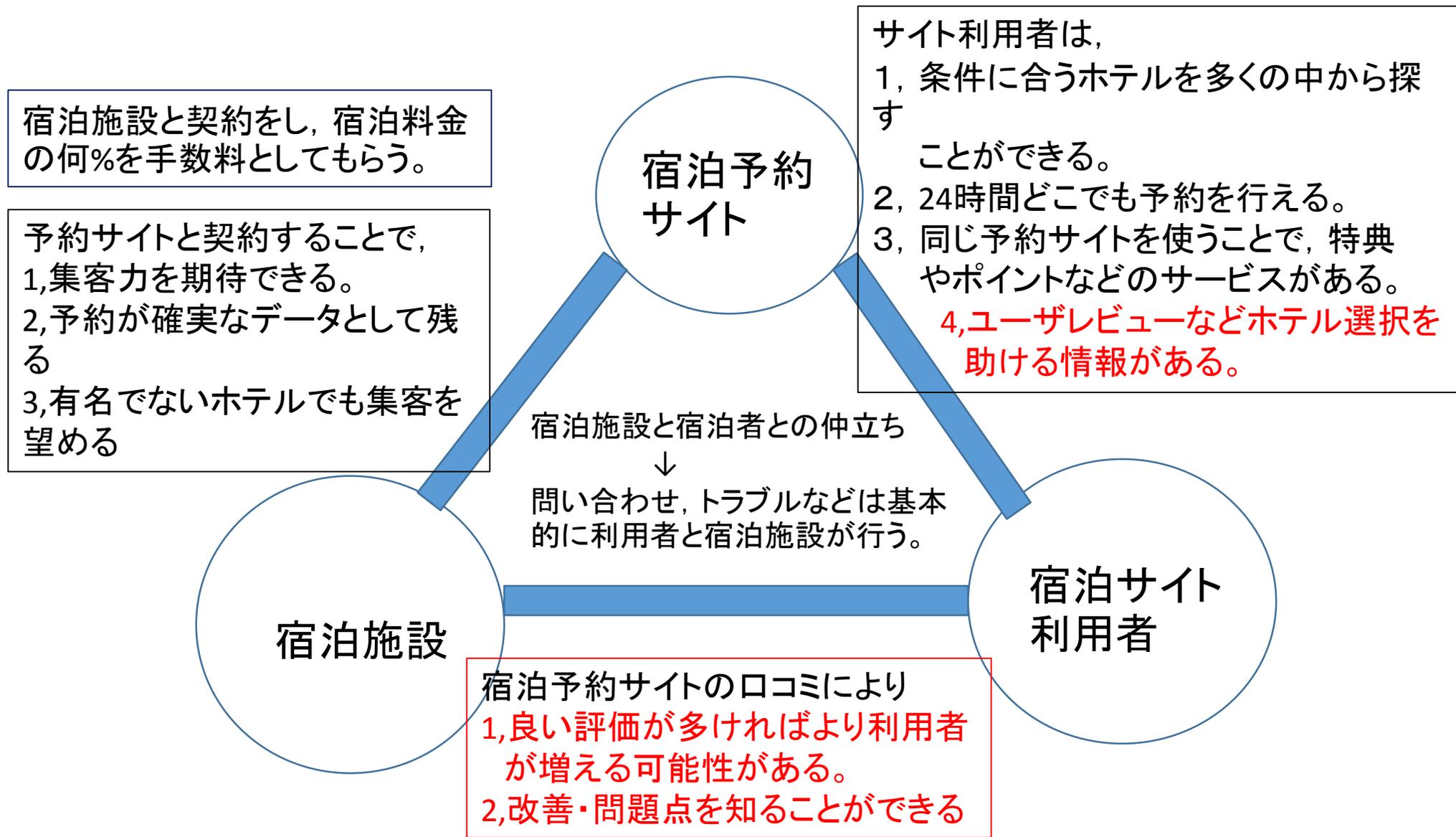
楽天トラベルにおけるモバイル予約率



ホテルの善し悪しの判断基準がサイト利用者から求められている。

宿泊予約サイトは評価項目と口コミを宿泊者に求め掲載することで、サイト利用者にホテル予約の手助けを行っている

1-4. 宿泊予約のEC化における三者関係



2.研究目的

本研究では、以下の二つを目的とする。

- ①楽天トラベルとじゃらんの評価値と口コミを用いて既存の評価項目の信頼性を確かめ、
- ②口コミの分析を行うことで既存にない重要項目を抽出し新たな評価項目はないかを検討する。

楽天トラベルとじゃらんnetの評価項目

楽天トラベル	総合	サービス	立地	部屋	設備・アメニティ	風呂	食事
じゃらんnet	総合	部屋	風呂	料理(朝食)	料理(夕食)	接客・サービス	清潔感

本当にこの評価項目は重要なのか？
他に重要項目はないのか？

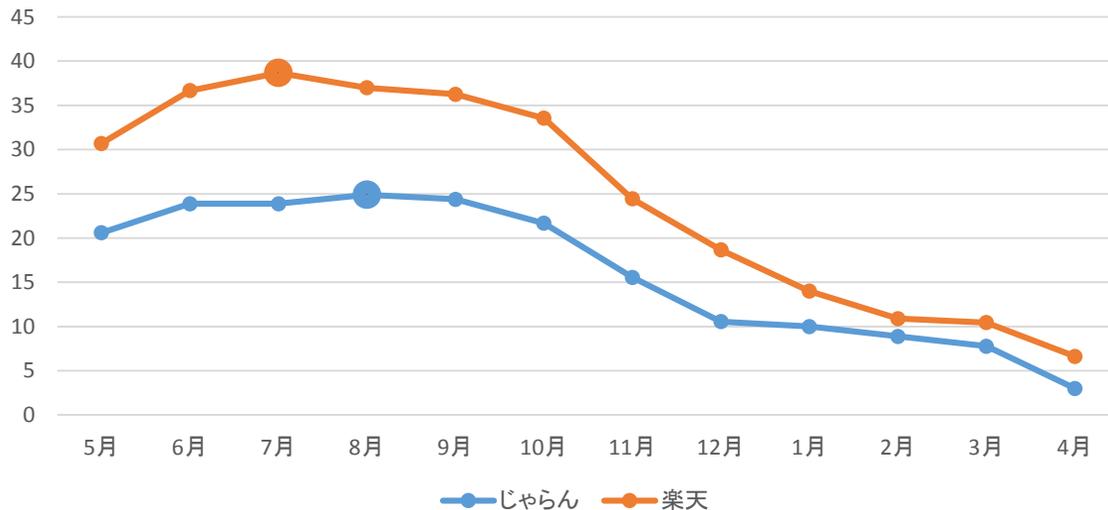
3.本研究のデータ収集

①渋谷駅周辺のホテルを楽天トラベル・じゃらんnetより検索を行う。

両宿泊サイトに掲載されているホテルを抽出する為に、掲載されている宿泊施設件数を月ごとに最大値を調べる。

収集期間は2019年5月~2020年4月までの一年間分とした。

ホテル掲載数



楽天トラベルは7月、じゃらんnetは8月に掲載数が最大になることが分かる。

上記より、最も多くなった時のホテルを書き出し、両サイトに掲載されているホテルの抽出を行った。

- ・ 楽天トラベル・じゃらんnetより共通のホテルに対するレビューの収集を行った。
収集するデータ

2018年7月～2019年6月までの過去一年間のすべてのレビュー

楽天トラベル

ホテルデータ	ホテル評価項目				テキストデータ	ホテル属性データ	
ホテル名	総合	サービス	立地	食事	口コミ	旅行の目的	同伴者
総レビュー数	部屋	風呂	設備・アメニティ		ホテルからの返信	年代	性別

じゃらんnet

ホテルデータ	ホテル評価項目				テキストデータ	ホテル利用者属性データ	
ホテル名	総合	部屋	風呂	朝食	口コミ	目的	宿泊価格帯
総レビュー数	夕食	接客	清潔感		ホテルからの返信	年代	性別

両サイトともに食事項目のデータが少ないため、今回は分析対象としない。

一年間の件数が極端に少ないホテルをある為、今回は総レビュー数及び総口コミ数が最も多い10件のホテルを抽出し、今回の分析対象とする。

分析対象ホテル10件

楽天トラベル→303件

じゃらんnet→254件 のホテルレビューデータ

サクラ・フルール青山	渋谷エクセルホテル東急
JR東日本ホテルメッツ 渋谷	オリンピックイン渋谷
渋谷グランベルホテル	東急ステイ渋谷（道玄坂上）
セルリアンタワー東急ホテル	渋谷東武ホテル
ホテルユニゾ渋谷	東急ステイ渋谷（新南口）

4.分析方法

4-1 既存の5段階評価項目における分析

今回、評価項目の評価値の分析にはVisual Mining Studio(VMS)を使用した。

①相関分析, 回帰分析

楽天トラベル・じゃらんnetの評価値をそれぞれ相関分析をすることで、項目間の相関関係を調べる。

次に、相関分析より項目間に関係がありそうなものについて単回帰分析による分析を行う。



相関関係及び、単回帰分析の結果から総合評価及び項目間の関係性の分析を行う。



②主成分分析

楽天トラベル・じゃらんnetの評価項目で主成分分析を行うことで「次元の縮約」を行う。



「次元の縮約」を行い、全体のばらつきを最もよく表す主成分を抽出することで、評価項目の中でも重要な項目を導き出す。

評価項目から重要項目になりうる項目を導き出す

4-2 口コミにおける分析

- ・ 楽天・じゃらの口コミから書かれている内容を10項目に分類・整理を行った。

10項目の分類を決める際に、宿泊サイト11サイトの評価項目を参考に評価項目の選定を行った。

今回使用する10項目

部屋・客室	風呂・水回り	食事	サービス(アメニティ)	接客
清潔感	立地	施設・設備	コスパ・金額	快適さ・満足度

部屋, 客室	風呂・水回り	食事	サービス (アメニティ)	接客
綺麗だった	お風呂が良かった	料理の味	部屋のアメニティについて	従業員の態度
広々としている	ゆったりとくつろげる	朝食が選べる	アメニティが充実している	心遣いしてくれた
隣の部屋がうるさかった	お風呂が広い	食事の量	水のサービスがある	対応が良い, 悪い
少し部屋が狭かった	風呂のアメニティ (石鹸や, シャンプー, タオルなど)	品数	消臭剤や, リセッシュが欲しい	親切な対応
部屋に関するネガティブな意見	シャワーに関する記述	提供スピード	部屋の備品が揃っていた	チェックイン, アウト時についての記述
ホコリが残っていた	風呂の設備	レストランからの景色		
	洗面所, トイレに関する記述			
立地	施設・設備	コスパ・金額	快適さ・満足度	清潔感
目的の場所から近かった	駐車場	コストパフォーマンス	快適に泊まれた	綺麗だった
駅近	設備が整っていた	金額に見合ったホテル	安心できる	清潔感があった
場所がわかりやすかった	コーヒーマシンが設置してあった	コスパ重視	満足だった	においや汚れなど気になるところがなかった
利便性	セキュリティ	料金設定について	非常に良かった (あまりに良かった)	汚れがない
アクセスが良い	館内に関する記述	希望に価格に見合っているか	気持ちよく宿泊ができた	
歩いていける範囲	老朽化, 少し古い			
周辺施設の有無	可愛い, 新しい, おしゃれ			
	入口が暗い, ドアが重い			
	建物に関する記述			



上記のようなある程度基準となる項目を定めた。

この評価項目10項目において口コミの中に記述がある場合は1,記述がない場合は0として,【口コミ件数×評価項目列】の0-1行列を作成した。



【口コミ件数×評価項目列】の0-1行列でコレスポンデンス分析を行う

①

「次元の縮約」を行い,各成分の特性を解釈する。

その際に累積寄与率と散布図の値から**口コミから読み取れる重要項目になり得る可能性**のある項目を分析する。



コレスポンデンス分析から読み取れた項目を踏まえ,楽天トラベル・じゃらんnetの口コミをText Mining Studioで分析を行う。

4-3 口コミにおける分析

- Text Mining Studioを用いて、係り受け分析及び評判抽出分析をすることで口コミから重要項目を分析する。

係り受け頻度分析

楽天トラベル・じゃらんnetの
口コミをそれぞれを形態素に
切り分け、その間の関連頻度
を抽出する。

より頻出度が多いコメントが
ある項目は何かを分析する

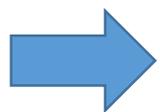
評判抽出分析

同じようにそれぞれの口コミ
を評判抽出分析を行い、ネガ
ティブな意見、ポジティブな意
見がどのくらいあるのかを分
析を行う。

係り受け頻度分析同様に、頻
出度の多いコメント項目は何
かを分析する。



口コミから読み取れる重要項目を選定する



評価項目及び口コミから導き出した重要項目から、**評価項目がどれほどの信頼性があるかを検討し、これまでにない項目があるか分析を行う。**

5.分析結果

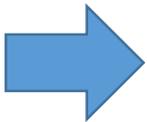
5-1 楽天トラベルの5段階評価項目における相関分析結果

相関分析結果

	総合	サービス	立地	部屋	設備・アメニティ	風呂
総合	1	0.715091	0.520235	0.704895	0.586457035	0.576294
サービス	0.715091	1	0.351842	0.566962	0.592908011	0.505255
立地	0.520235	0.351842	1	0.378198	0.362403449	0.322365
部屋	0.704895	0.566962	0.378198	1	0.58882964	0.587599
設備・アメニティ	0.586457	0.592908	0.362403	0.58883	1	0.559296
風呂	0.576294	0.505255	0.322365	0.587599	0.559295906	1

サービス, 部屋は総合との相関関係が高い。(黄色)

立地はどの評価項目においても相関がないといえる。(水色)



5-2 楽天トラベルの5段階評価項目における総合-部屋・サービスの単回帰分析結果

単回帰分析結果

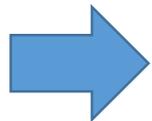
目的変数	説明変数	決定係数	回帰係数	P-値
総合	部屋	0.495086	0.621857	8.18E-44
総合	サービス	0.509617	0.619908	1.33E-45

部屋、サービス共に決定係数より高い相関があると言える。
回帰係数を見ると、二つの項目はどちらも総合との関係性が高い。

5-3 じゃらんnetの5段階評価項目における相関分析結果

相関分析結果

	総合	部屋	風呂	接客・サービス	清潔感
総合	1	0.799579	0.717071	0.68805003	0.770573
部屋	0.799579	1	0.70724	0.561601564	0.704119
風呂	0.717071	0.70724	1	0.557405539	0.701957
接客・サービス	0.68805	0.561602	0.557406	1	0.632267
清潔感	0.770573	0.704119	0.701957	0.63226709	1



総合評価はどの項目に関しても総合との相関関係が高いのが分かる。(黄色)

清潔感に関してもほとんどの項目と相関関係がある。
→清潔感、どこのどの清潔感の評価項目なのかが分かりにくい。

5-4 じゃらんnetの相関が高い項目の単回帰分析結果

単回帰分析結果

目的変数	説明変数	決定係数	回帰係数	P-値
総合	部屋	0.637867	0.758766	1.27E-56
総合	風呂	0.512224	0.693254	1.34E-40
総合	接客	0.471281	0.746425	2.93E-36
総合	清潔感	0.592138	0.693893	3.15E-50
部屋	風呂	0.498164	0.720526	4.53E-39
部屋	清潔感	0.493742	0.668156	1.35E-38
風呂	清潔感	0.49069	0.653822	2.84E-38
風呂	部屋	0.498164	0.694198	4.53E-39

決定係数, 回帰係数どちらも他の項目よりも高い

清潔感は, 部屋と風呂に関して同等くらいの関係性を持つ。

- 1, 全ての項目において, 総合と関係性が高い。それぞれの項目の回帰係数より, 「部屋, 接客, 清潔感, 風呂」の順位付けができる。
- 2, 清潔感の項目は部屋, 風呂どちらに関しても関係性がある。

5-5 楽天トラベルの評価項目における主成分分析結果

主成分名	第1主成分	第2主成分	第3主成分	第4主成分	第5主成分	第6主成分
標準偏差	1.918635	0.867201	0.719559	0.665117	0.62639	0.462917
寄与率	0.613527	0.12534	0.086294	0.07373	0.065394	0.035715
累積寄与率	0.613527	0.738866	0.82516	0.898891	0.964285	1

第一主成分が61%の情報量を持っている累積寄与率が8割を超えた第3主成分までを採用

列名	第1主成分	第2主成分	第3主成分	第4主成分	第5主成分	第6主成分
総合	-0.46137	-0.09644	-0.24527	-0.32701	0.022454	-0.78118
サービス	-0.42101	0.146524	-0.64367	0.00496	0.436571	0.443124
立地	-0.30918	-0.90662	0.165161	0.123943	0.053025	0.192313
部屋	-0.4316	0.146071	0.11316	-0.44529	-0.66743	0.368561
設備・アメニティ	-0.4133	0.208924	0.019604	0.819253	-0.30737	-0.13962
風呂	-0.39648	0.286839	0.696469	-0.09068	0.515899	0.032873

ホテルの総合評価軸 立地軸 風呂-サービス軸

5-6 じゃらんnetの評価項目における主成分分析結果

主成分名	第1主成分	第2主成分	第3主成分	第4主成分	第5主成分
標準偏差	1.935215	0.69811	0.559778	0.531998	0.413777
寄与率	0.749011	0.097471	0.06267	0.056604	0.034242
累積寄与率	0.749011	0.846483	0.909153	0.965758	1

第一主成分が74%の情報量を持っている
累積寄与率が8割を超えた第2主成分までを採用

列名	第1主成分	第2主成分	第3主成分	第4主成分	第5主成分
総合	-0.47694	0.005232	-0.33343	-0.01611	-0.81306
部屋	-0.45266	0.360286	-0.56754	-0.2945	0.506424
風呂	-0.44048	0.376079	0.737829	-0.34487	-0.03494
接客・サービス	-0.40669	-0.85297	0.088393	-0.24204	0.201623
清潔感	-0.4563	0.034325	0.120492	0.857611	0.201484

↓ ↓
ホテルの総合評価軸 接客・サービス軸

5-7 楽天トラベルの口コミにおけるコレスポネンダ分析結果

楽天トラベルでは【290×10列】の0-1行列で行う。

項目	部屋・客室	風呂	食事	サービス（アメニティ）	接客
項目における意見数	161	46	48	46	114
項目	清潔感	立地	施設・設備	コスパ・金額	快適さ・満足度
項目における意見数	11	157	56	33	44

	x	寄与率	累積寄与率
1	0.625559	0.391323	0.142875071
2	0.614479	0.377584	0.280733793
3	0.592292	0.350809	0.408816867
4	0.573193	0.32855	0.52877299
5	0.56083	0.314531	0.643610478
6	0.536231	0.287544	0.748594931
7	0.500861	0.250861	0.840186275
8	0.472043	0.222825	0.92154127
9	0.463565	0.214892	1
10	1.59E-15	2.53E-30	1

累積寄与率8割に達したのが第7主成分であった。

それぞれの軸を解釈すると、

第1主成分 風呂ー快適さ・満足度軸

第2主成分 コスパー快適さ・満足度軸

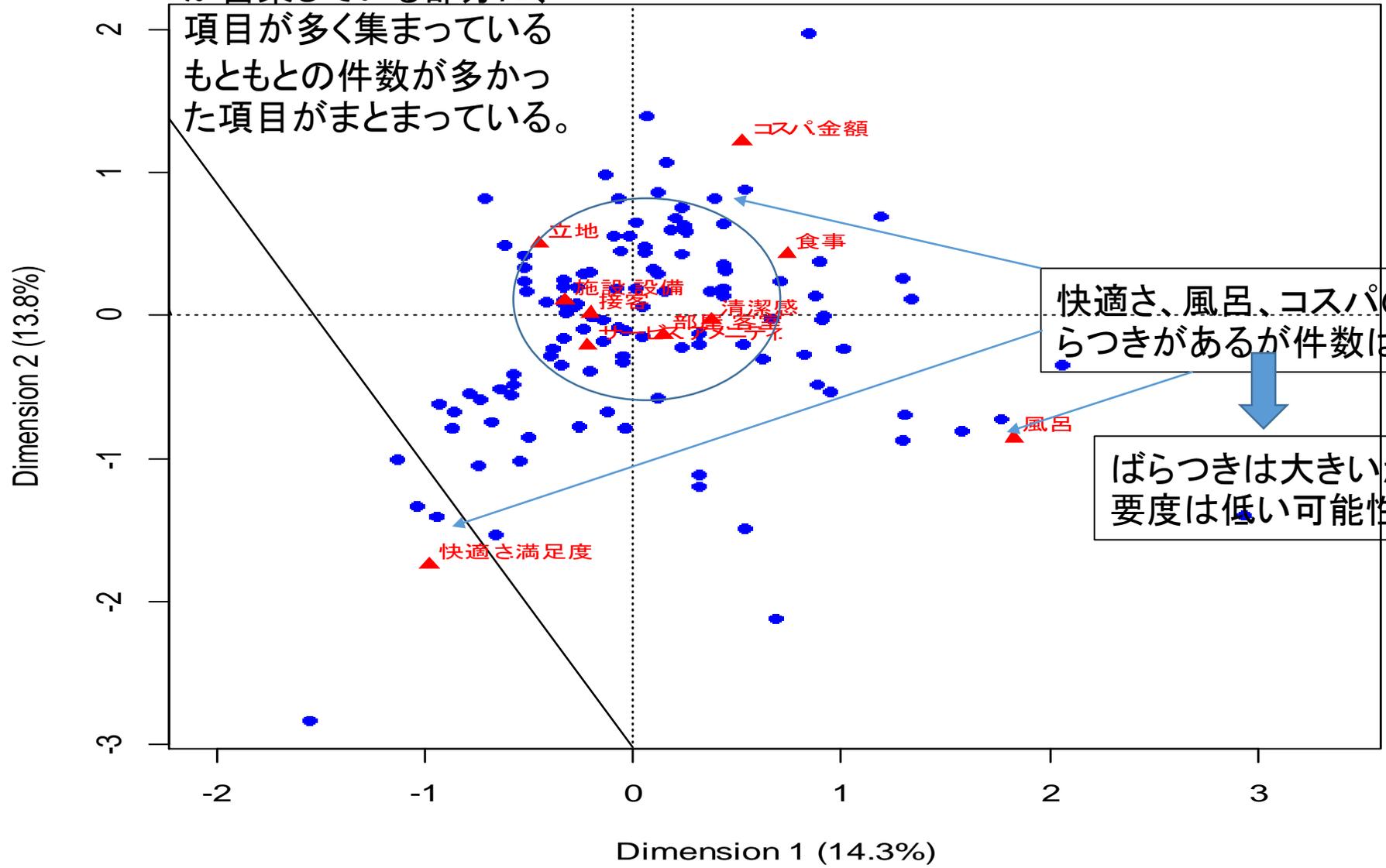
第3主成分 食事軸

第4主成分 接客ーコスパ金額軸

5以降は解釈ができなかった。

0-1行列 コレスポネンダ分析 1-2主成分図

軸にはなっていないが、点
が密集している部分に、
項目が多く集まっている
もともとの件数が多かった
項目がまとまっている。



快適さ、風呂、コスパの項目はばらつきがあるが件数は少ない。

ばらつきは大きいが重要度は低い可能性

5-8 じゃらんnetの口コミにおけるコレスポネンズ分析結果

じゃらんnetは、【247×10列】の0-1行列で行う。

項目	部屋・客室	風呂	食事	サービス（アメニティ）	接客
項目における意見数	170	51	59	60	100
項目	清潔感	立地	施設・設備	コスパ・金額	快適さ・満足度
項目における意見数	53	151	65	43	49

	x	寄与率	累積寄与率
1	0.535969	0.287262	0.145831287
2	0.524586	0.27519	0.285533927
3	0.512283	0.262434	0.418760675
4	0.482915	0.233207	0.537150325
5	0.46383	0.215138	0.646367288
6	0.459539	0.211176	0.753572661
7	0.430849	0.185631	0.847809958
8	0.404516	0.163633	0.930879873
9	0.368991	0.136155	1
10	1.60E-15	2.55E-30	1

累積寄与率8割に達したのが第7主成分であった。

それぞれの軸を解釈すると、

第1主成分 風呂—コスパ・金額軸

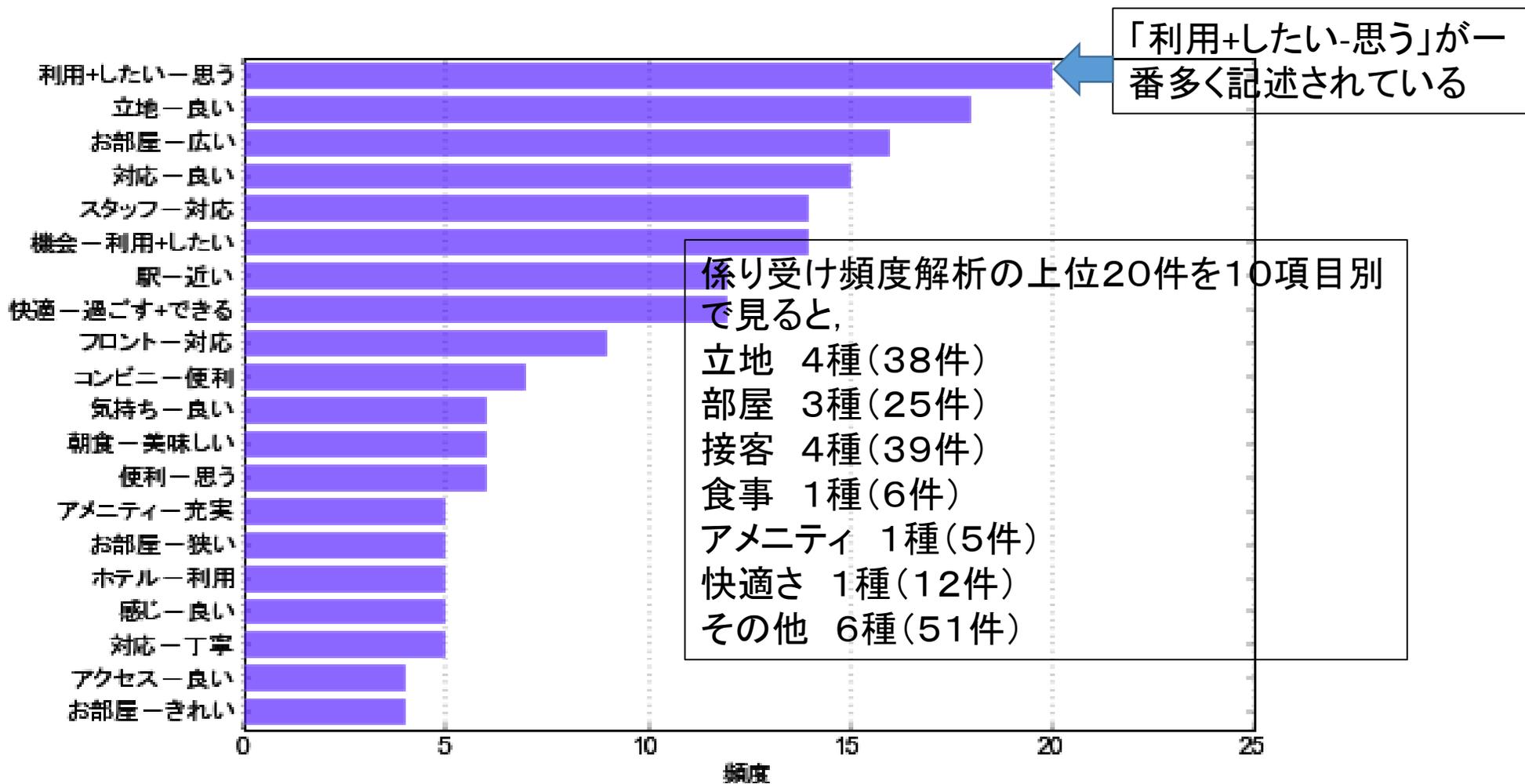
第2主成分 快適さ・満足度—コスパ・金額軸

第3主成分 食事軸

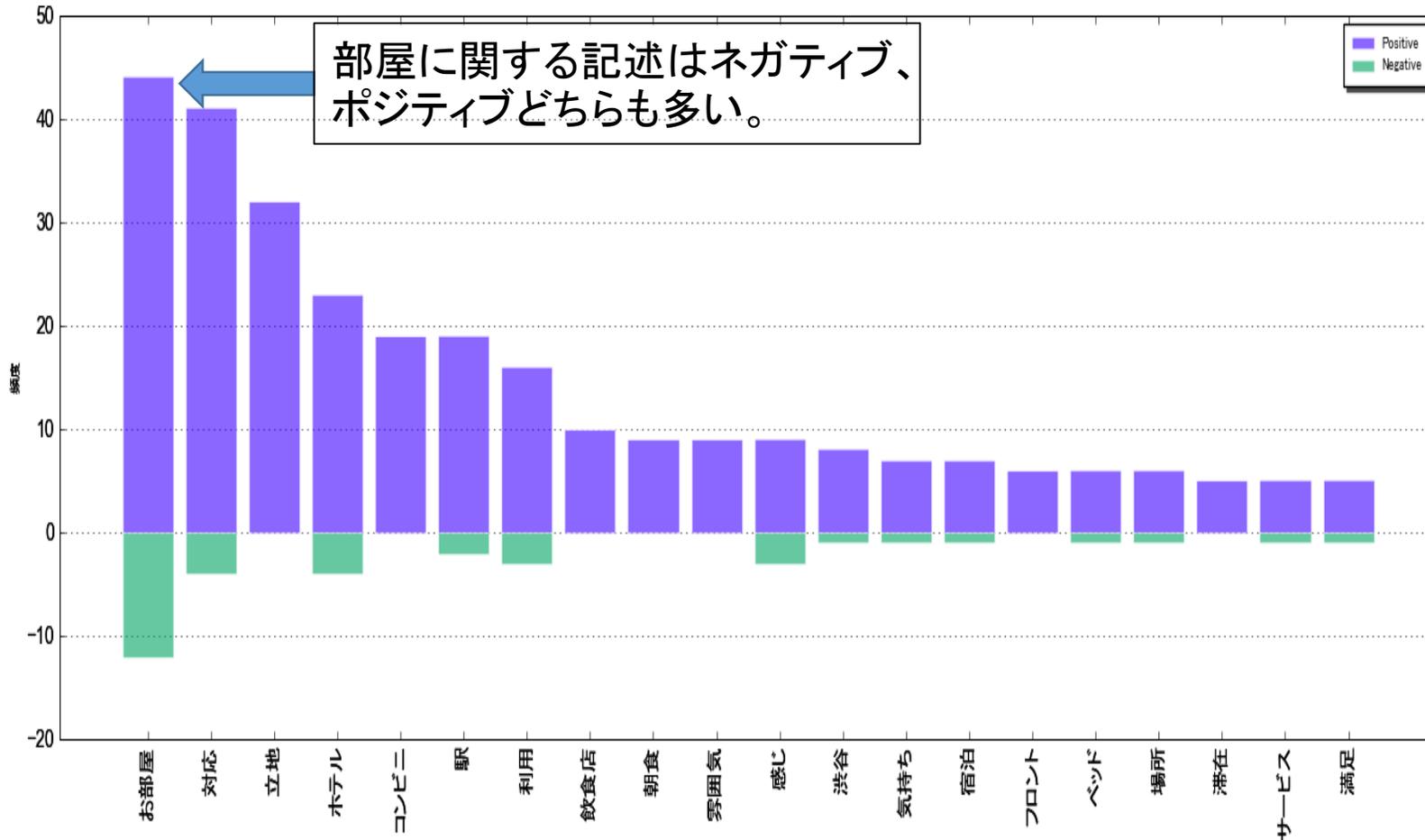
第4主成分 施設・設備—コスパ・快適さ軸

5以降は解釈ができなかった。

5-9 楽天トラベルの口コミにおける係り受け頻度分析結果



5-10 楽天トラベルの口コミにおける評判抽出分析結果



評判抽出において、「部屋」・「接客」・「立地」・「利用」の順番で頻出しているのが分かる。

5-11楽天トラベルの係り受頻度分析・評判抽出結果

部屋に関する記述 全40件

広い	綺麗	狭い	快適	設備	満足	良い
16	8	5	5	2	2	2

→ 部屋の広さに関する記述が多い

接客に関する記述 63件

対応	スタッフ対応	フロント対応	接客
26	23	12	2

→ 接客でも、対応に関する項目が多い

立地に関する記述 全60件

立地	駅からのアクセス	コンビニ近い、便利
23	24	13

→ 駅に関する項目が多い

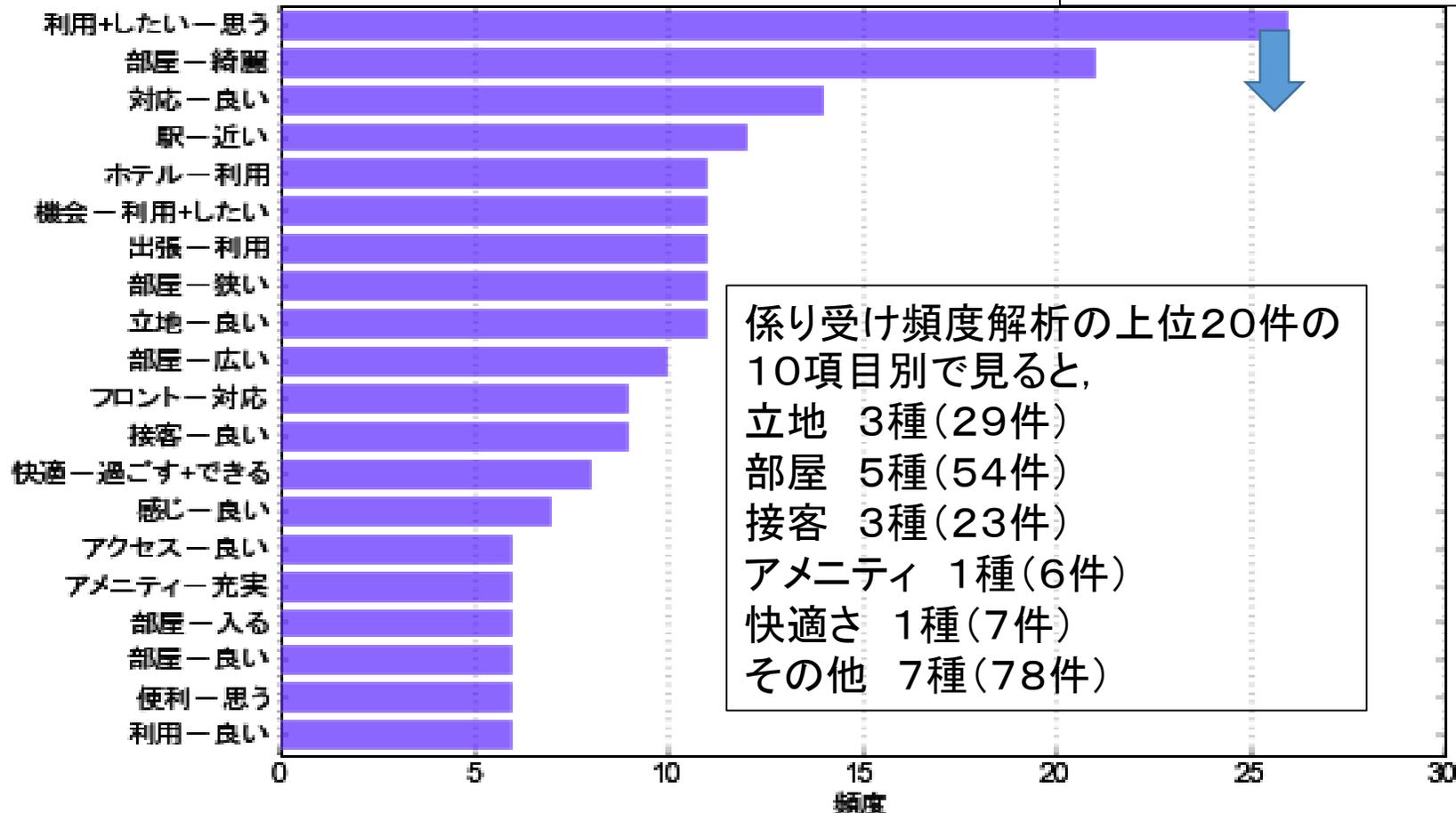
利用に関する記述 26件

また利用したい	利用したくない	利用-おすすめ
22	2	2

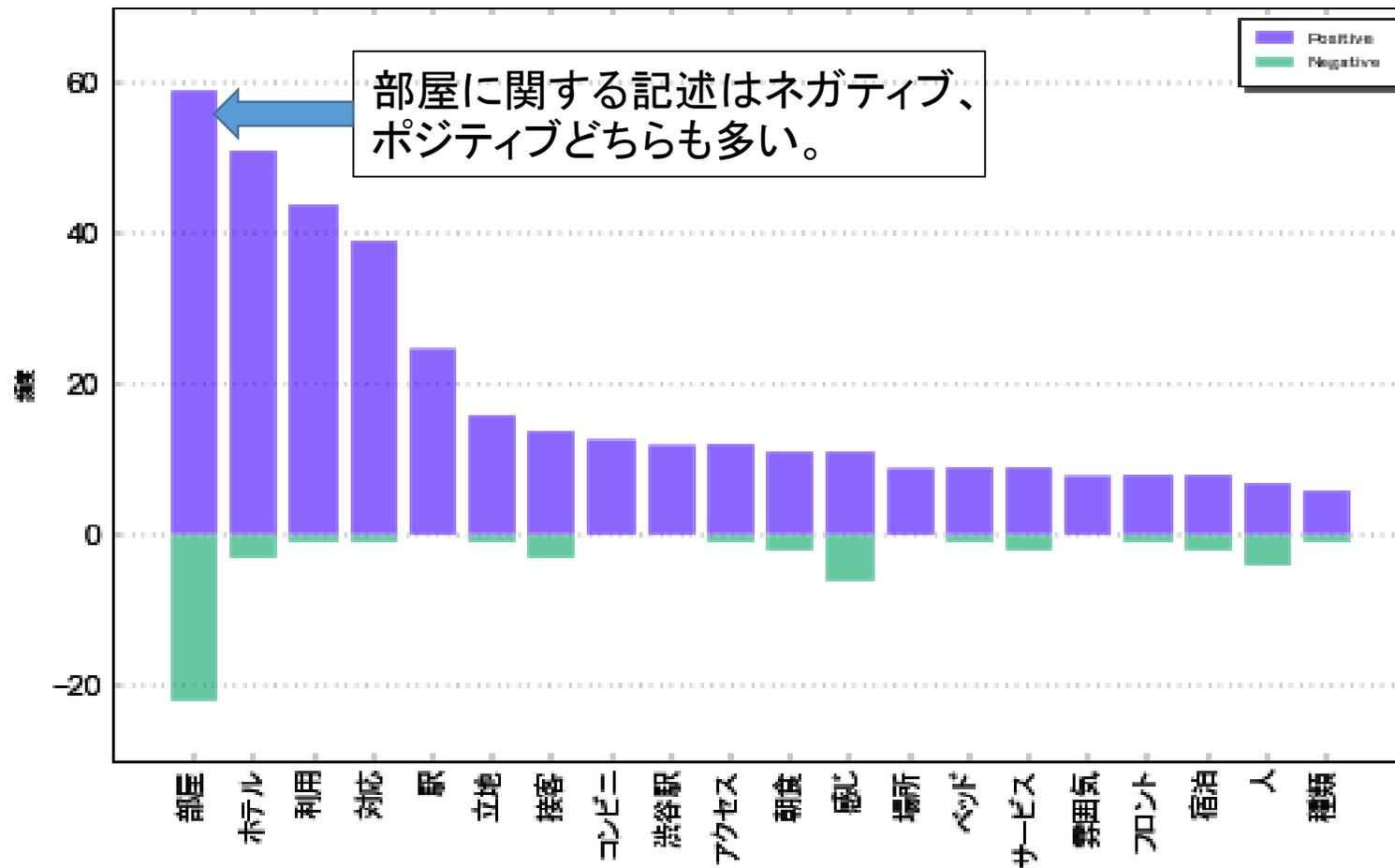
→ 再利用の有無に関する記述が多い

5-12 じゃらんnetの口コミにおける係り受け頻度分析結果

「利用+したい-思う」が一番多く記述されている



5-13 じゃらんnetの口コミにおける評判分析結果



評判抽出において、「部屋」・「利用」・「接客」・「立地」の順番で頻出しているのが分かる。

5-14 楽天トラベルの係り受頻度分析・評判抽出結果

部屋に関する記述 全51件

綺麗	広い	狭い	清潔
21	13	14	3

➡ 部屋の広さに関する記述が多い

接客に関する記述 全48件

対応	スタッフ対応	フロント対応
19	12	17

➡ 接客でも、対応に関する項目が多い

立地に関する記述 全45件

駅からのアクセス	アクセス	コンビニ
32	9	4

➡ 駅からの距離に関する記述が多い

利用に関する記述 全30件

利用したい	おすすめ	目的
28	2	11

➡ 再利用の有無に関する記述が多い

6. 考察

6-1 5段階評価項目の相関分析・単回帰分析結果

- ① **部屋、接客・サービス**は総合との関係性が高く重要項目になる可能性が確認できた

楽天

部屋・サービスは総合との関係性が高い。

➡ 重要度が高い可能性

評価項目同士ほとんどが相関関係がある中で、立地に関してはどの項目とも関係性が低かった。

➡ 他の項目との種類が違う。

じゃらんnet

すべての項目で総合との関係性が高いが、単回帰分析の回帰係数より総合との関係性の順位付けを行った。

1位 部屋 2位 接客・サービス

3位 清潔感 4位 風呂 となった。

清潔感の項目は部屋、風呂どちらにも関係性がある

➡ 清潔感の基準はどこに対してなのか？

6-2 5段階評価項目における主成分分析結果

② 総合、立地、接客・サービスは重要項目の可能性がある

楽天トラベル

- ・ 楽天トラベルにおいては、第1主成分に総合評価、次に立地軸、サービス軸となった。

 立地軸は総合の次に、ばらつきのある項目だとわかり、重要な項目である可能性がある。

じゃらんnet

- ・ じゃらんnetにおいては、第1主成分に総合評価、次に接客・サービス軸となった

①②から、既存の5段階評価項目から分析される重要項目は **総合、部屋、接客・サービス、立地**であると言える

6-3 楽天トラベル・じゃらんnetにおけるコレスポネンス分析結果

- どちらともに第1主成分～第4主成分まで、意見数が少ない項目が軸となっている。
- 第1主成分と第2主成分で作った図からも、成分の軸となっている項目の近くよりも、部屋や立地など件数が多かった評価項目の近くに件数も多くなっていることがわかる。

今回の対応分析の結果からは、項目の意見数に依存しすぎてしまう結果になった。

6-4 楽天トラベル・じゃらんnetの係り受け頻度分析・評判抽出結果

部屋、立地、接客が重要項目と抽出できた

さらに、「また・機会があれば利用したい」などの再利用に関する記述が多く読み取れた。

それにより再利用率の項目が新たに定義できる

部屋、立地、接客の項目を詳しく口コミを見ると、

部屋→部屋の広さ

立地→駅からのアクセス

接客→対応の良さ

と細かく定義することができた。

7. 結論

本研究では、渋谷駅周辺のホテルを対象に行った為、都心のビジネスホテルに関しての評価項目基準が定めることができた。

それぞれ異なる、楽天トラベルの「総合」、「サービス」、「立地」、「部屋」、「設備・アメニティ」、「風呂」、「食事」、じゃらんnetの「総合」、「部屋」、「風呂」、「朝食」、「夕食」、「接客」、「清潔感」の項目を分析し、既存の評価項目の分析から導いた重要項目「**部屋**」、「**立地**」、「**接客・サービス**」の項目は、口コミより導き出せた重要項目に対応していることから**信頼性が高い**といえる。

「**総合**」の項目は、他項目との関係性が高いが、唯一関係性が低い「立地」が重要項目であるため評価項目として入れてもよいと考えられる。

さらに口コミの分析結果より、「**部屋**」、「**立地**」、「**接客**」の項目についてはそれぞれ、「部屋の広さ」、「駅からのアクセス」、「対応」と細かく定義できる。

また、口コミの分析結果から新たに「**再利用度**」という項目を付け加えることが必要であると考え、以上の「**総合**」、「**部屋**」、「**立地**」、「**接客**」、「**再利用度**」の5項目を宿泊サイト側、ユーザ側、ホテル側が共有できる評価項目として提案する。

8. 今後の課題

課題① 今回の研究では渋谷駅周辺を対象にしている為、ビジネスホテルに偏っている。

解決手段としては、無作為にホテルを抽出することで、より幅広いホテルに対して評価項目と口コミの信頼性を分析することができる。

② 口コミを数値化し、コレスポネンス分析を行った結果が、評価項目件数による偏りが出てしまった。

解決手段としては、項目の中でも、より細かく項目を分けることで件数の偏りを減らすことで、コレスポネンス分析から有意な結果が得られると考える。

参考文献

- ・仲谷秀一他 (2018年)「ホテル・マーケティング・ブック」 中央経済社
- ・仲谷秀一他 (2018年)「ホテル・ビジネス・ブック」第2版 中央経済社
- ・岡太彬訓 森口剛 (2010年)「マーケティングのデータ分析 分析手法と適用事例」 朝倉書店
- ・(注①)総務省 情報通信統計
<http://www.soumu.go.jp/johotsusintokei/>(2019/10/1)
- ・(注②)観光経済新聞
<https://www.kankokeizai.com>(2019/10/1)