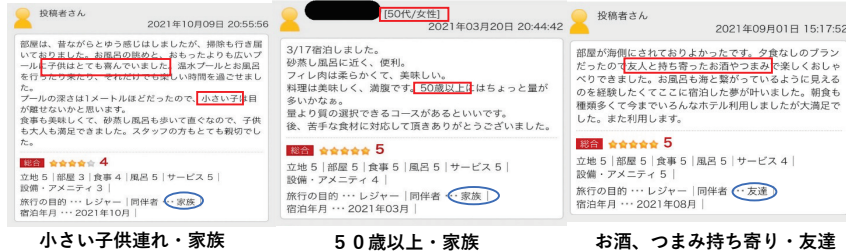


研究背景

同伴

家族
恋人
友達
仕事仲間
一人
その他

ホテルレビューの同伴者欄とレビューに着目し、それぞれの宿泊施設の客層を推定する

使用するデータ

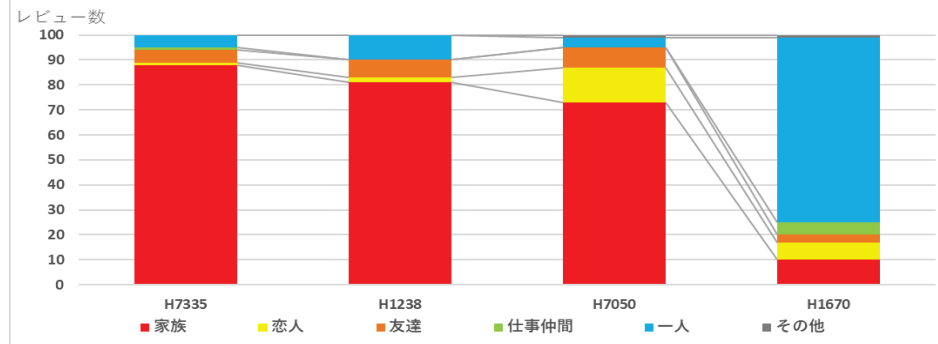
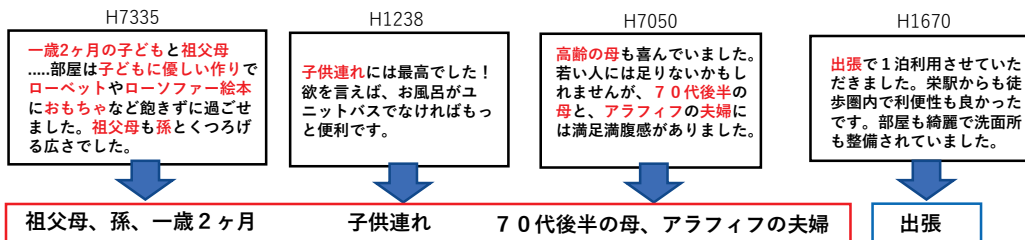


図1：各ホテルの同伴者の内訳



参考文献

[1] 楽天株式会社(2017,2018,2019): 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 (楽天トラベルデータセット)
info:doi/10.32130/idr.2.2

[2] 観光を対象としたレビューからの耳より情報抽出, 阪井奎伍, 瀬本明代, 情報処理学会研究報告, Vol.2015-DBS-162 No.13, 2015

[3] 商品推薦のための商品レビューの極性分析に基づく特徴語抽出手法, 吉田朋史, 北山大輔, DEIM Forum 2015 B2-1

提案手法

1. 客層に関する特徴語の抽出

①各宿泊施設から100件ずつレビューと同伴者を集計

②同伴者欄が「家族」のレビュー文に対し、形態素解析して名詞のみを抽出
(品詞情報に非自立、代名詞、副詞可能、接尾のいずれかを含む名詞は対象外[3])

宿 / に / いくつも / お世話 / に / なり / まし / た

名詞 助詞 名詞 名詞 助詞 動詞 助動詞 助動詞

一般 格助詞 代名詞 名詞 名詞 格助詞 自立

③TF-IDFで客層に関する
特徴語調査③Okapi BM25で客層に関する
特徴語調査

表1：TF-IDF値を使用

特徴語	H7335	H1238	H7050	H1670
温泉	0.007	0	0.017	0
ディズニー	0	0.063	0	0
バイキング	0.010	0.001	0	0.002
子供	0.010	0.015	0.001	0
繁華街	0	0	0	0.007
ビジネス	0	0	0	0.010
挨拶	0	0.001	0.007	0
浴場	0.005	0	0.007	0
料理	0.006	0.001	0.006	0
夕食	0.005	0.0003	0.006	0
屋敷	0	0.005	0	0

赤：重要度が最も高い
黄：TF-IDF値が0.005以上

・客層に関する特徴語

子供、祖父母、パパ、ママ、親子、赤ちゃん 中年、夫、ビジネス
(子供連れ) (大人)表2:TF, IDF, DL値を使用
(Okapi BM25)

特徴語	H7335	H1238	H7050	H1670
ケーキ	0	0	0.002	0
懐石	0	0	0.002	0
中年	0	0	0.001	0
夫	0	0	0.001	0
祖父母	0.003	0	0	0
パパ	0.002	0	0	0
ビジネス	0	0	0	0.016
繁華街	0	0	0	0.011
風俗	0	0	0	0.005
ディズニー	0	0.054	0	0
屋敷	0	0.006	0	0
ママ	0	0.003	0	0
親子	0	0.002	0	0
学年	0	0.002	0	0
赤ちゃん	0	0.002	0	0

赤：重要度が最も高い
黄：TF-IDF値が0.001以上

2. 客層推定

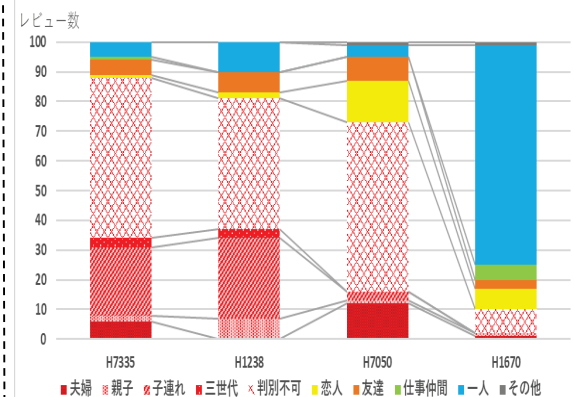


図2：各ホテルの「家族」の内訳

ただ、小さい子供さん達が多く、夕飯時に泣いたりして、落ち着かなかったです。私は、母と二人旅だったので もう少しゆっくりと食事がしたかったです。(H7335)

完全に大人向け。
当日、小さな子供連れではなく静か。(H7050)

ゴンドラで迎えに来てくれた中年の男の方。
細かいところで気配りが足りない。(H7050)

・扱う特徴語の見直し
・着目するレビュー、同伴者

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/)を利用しました。
心から感謝申し上げます。