

同行者と旅行目的に基づく旅館レビュー分類のためのカテゴリの検討

Consideration of Accommodation Review Categories
Based on Travel Party and Purpose坂野 妃菜*
Hina Sakano谷口 雅空*
Masataka Taniguchi鈴木 信太郎*
Shintaro Suzuki工藤 竜矢*
Ryuuya Kudo宍戸 伶雅*
Ryoka Shishido酒井 柊輔*
Shusuke Sakai延澤 志保*
Shiho Hoshi Nobesawa

1. 研究背景

旅館レビューには「子連れでもゆっくりできます」などユーザにとって満足のいく宿泊であったことが伺えるような感想があるのに対して、「小さい子供が多く夕飯時に泣いたりしていて落ち着かなかった」などの不満が残る宿泊であったことを示す感想も含まれている。サイトごとに立地、部屋、食事、風呂、サービス、設備の6カテゴリといったようなレビューカテゴリが分けられている場合があり、レビュー分類を試みる研究は多くある [2, 3] が、一般的な評価項目のみではユーザが宿泊施設を選択する際に自身のニーズに合う宿泊施設かどうかを判定することは難しい場合がある。宿泊施設レビューサイトの同行者項目では、子供連れも高齢者連れもすべて家族としてひとまとめになっている場合もあり、これらの自動分類は容易ではない [4]。そのため本稿では、同行者と旅行目的に考慮した旅館レビューの適切なカテゴリを検討する。

2. 旅館レビューの特徴

楽天トラベルデータセット [1] のうち 2017 年から 2019 年までのデータ 1,497,593 件について、各宿泊施設の客層を調べるため、レビュー数が多い順に上位 24 件の宿泊施設からレビューを 500 件ずつ無作為に抽出し、楽天トラベルの既存の同行者項目である家族、恋人、友達、仕事仲間、1人についてそれぞれの出現数を調べた (図 1)。図 1 は横軸が 24 件の宿泊施設、縦軸

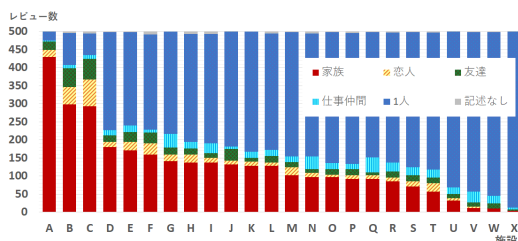


図 1: 宿泊施設 24 件の同行者の内訳

が同行者ごとのレビュー数であり、家族の割合が高い順に宿泊施設を並べている。図 1 では宿泊施設 ID をアルファベットに置き換えて説明する。

図 1 に示した 24 件の宿泊施設のうち、家族の割合が最も高い施設 A は家族が 85.9% であり、このほか施設 B、施設 C でも家族が半数を超える。それに対し家族の割合が最も低い施設 X では家族が 0.8% なのに対して 1 人での宿泊は 97.2% であり、施設 X は 1 人での宿泊を想定した宿泊施設と考えられる。ただし、施設 C

から施設 X までの施設で 1 人での宿泊が半数を超えており、24 件の施設の 80% 超が 1 人向けとは考え辛いことから、1 人での宿泊が複数での宿泊の場合に比べてレビューを書く可能性が高いものと仮定できる。

宿泊施設の種類を検討するため、宿泊施設の特徴を表す語の抽出を考える。レビューの評価属性として名詞に着目している研究は数多い [5, 6]。観光レビューについても名詞を特徴語とする手法が多い [2, 7] ことから、本稿では名詞に着目することでユーザの評価対象が捉えられるものと仮定し、レビューの特徴を表す名詞の抽出を試みる。また、品詞情報に非自立、数、接尾、副詞可能、代名詞を含む名詞は宿泊施設に関係がない語が多いことから抽出対象としない。

図 1 のうち施設 A、施設 C、施設 L、施設 Q、施設 W の 5 件について出現頻度の高い名詞上位 10 語を表 1 に示す。ここで、各語の数値は左がその宿泊施設で

表 1: 宿泊施設 5 件の出現割合上位 10 語

| 特徴語候補 | 施設A | | | 施設C | | | 施設L | | | 施設Q | | | 施設W | |
|-------|------|------|------|------|------|------|------|------|-------|------|------|------|------|------|
| | 全体 | 利用 | 部屋 | 全体 | 利用 | 部屋 | 全体 | 利用 | 部屋 | 全体 | 利用 | 部屋 | 全体 | 利用 |
| 利用 | 2.93 | 2.90 | 部屋 | 4.87 | 3.58 | 部屋 | 5.76 | 3.58 | 朝食 | 5.82 | 2.14 | 利用 | 4.37 | 2.90 |
| 部屋 | 2.73 | 3.58 | ホテル | 4.06 | 2.18 | 利用 | 4.58 | 2.90 | 利用 | 4.53 | 2.90 | 部屋 | 3.99 | 3.58 |
| ホテル | 2.71 | 2.18 | 利用 | 3.89 | 2.90 | 宿泊 | 3.06 | 2.12 | 部屋 | 4.00 | 3.58 | トイレ | 3.95 | 0.32 |
| バイキング | 2.71 | 0.44 | 宿泊 | 3.37 | 2.12 | ホテル | 3.06 | 2.18 | ホテル | 2.72 | 2.18 | 風呂 | 3.80 | 1.51 |
| 食事 | 2.68 | 1.48 | お部屋 | 2.24 | 0.73 | 朝食 | 2.58 | 2.14 | 宿泊 | 2.60 | 2.12 | 朝食 | 3.69 | 2.14 |
| 宿泊 | 2.25 | 2.12 | 夜景 | 1.75 | 0.05 | 立地 | 2.30 | 0.86 | アメニティ | 1.95 | 0.27 | ホテル | 2.60 | 2.18 |
| 満足 | 1.99 | 1.22 | スタッフ | 1.63 | 0.74 | 駅 | 1.98 | 0.87 | 満足 | 1.72 | 1.49 | 宿泊 | 2.24 | 2.12 |
| プール | 1.81 | 0.06 | 満足 | 1.46 | 1.22 | 満足 | 1.57 | 1.22 | バイキング | 1.54 | 0.44 | バス | 2.93 | 0.30 |
| 子供 | 1.59 | 0.43 | 朝食 | 1.25 | 2.14 | 対応 | 1.47 | 1.03 | ベッド | 1.36 | 0.36 | バス | 1.79 | 0.13 |
| 風呂 | 1.54 | 1.51 | 対応 | 1.28 | 1.03 | フロント | 1.47 | 0.94 | 駅 | 1.20 | 0.87 | フロント | 1.77 | 0.94 |

の出現割合 (%), 右が旅館レビュー 35,000 件全体での出現割合 (%) を示す。背景色が濃いほど割合が高い。

表 1 を見ると、施設 A では「プール」や「子供」の出現割合が全体での出現割合に比べて高いことから、施設 A はレジャー施設と子供の話題を中心としたユーザが多い宿泊施設であると考えられる。施設 C では「夜景」の出現割合が全体に対して高く、景色について言及しているユーザが多いことがわかる。施設 A も施設 C も同行者が家族であるユーザの割合が大きいですが、表 1 の特徴語候補の出現状況の相違から、ユーザの同行者が施設 A と施設 C で異なる可能性が考えられる。

施設 L、施設 Q、施設 W は「駅」や「アメニティ」、「トイレ」などの出現割合が全体での出現割合と比べて高い。これらの宿泊施設では同行者が 1 人のユーザの割合が大きい (図 1) ことから、施設 L、施設 Q、施設 W はビジネスホテルであると考えられ、家族の割合が大きい施設 A、施設 C と比較すると、ユーザの話題が異なるのは自然である。

3. 同行者と旅行目的に基づく旅行レビューカテゴリ

3.1. 同行者に関する特徴語の抽出

同行者の違いによるレビューの差異を検証するため、レビュー群から夫婦のみ、子供連れ、幼児連れ、赤ちゃん

* 東京都市大学, Tokyo City University

ん連れ, 高齢者連れの 5 種類に当たるレビューを 500 件ずつ選び, それぞれのレビュー群の名詞の出現状況を調べた (表 2). 夫婦のみレビュー群ではほとんどの語

表 2: 同行者 5 種類の出現割合上位 10 語

| 夫婦のみ | | 子供連れ | | 幼児連れ | | 赤ちゃん連れ | | 高齢者連れ | | |
|------|------|------|------|------|------|--------|------|-------|------|------|
| 特選候補 | 全体 | 特選候補 | 全体 | 特選候補 | 全体 | 特選候補 | 全体 | 特選候補 | 全体 | |
| 夫婦 | 4.02 | 2.75 | 4.47 | 0.43 | 3.02 | 0.13 | 4.94 | 0.17 | 2.91 | 3.58 |
| 部屋 | 2.77 | 3.58 | 3.27 | 3.58 | 2.53 | 3.58 | 2.70 | 3.58 | 2.16 | 2.90 |
| 宿泊 | 2.04 | 1.51 | 3.00 | 2.90 | 2.15 | 2.90 | 2.15 | 0.43 | 2.15 | 1.51 |
| 食事 | 2.03 | 1.48 | 2.71 | 2.18 | 2.10 | 0.43 | 2.09 | 2.90 | 1.98 | 0.10 |
| 利用 | 1.87 | 2.90 | 2.34 | 2.12 | 2.03 | 1.48 | 2.03 | 1.48 | 1.91 | 1.48 |
| 風呂 | 1.68 | 1.49 | 1.74 | 1.51 | 1.96 | 2.18 | 1.80 | 1.49 | 1.77 | 0.13 |
| 満足 | 1.44 | 1.22 | 1.52 | 2.14 | 1.87 | 2.12 | 1.51 | 1.51 | 1.74 | 2.18 |
| ホテル | 1.29 | 2.18 | 1.49 | 0.73 | 1.56 | 1.51 | 1.47 | 0.73 | 1.46 | 1.49 |
| 温泉 | 1.24 | 0.71 | 1.38 | 1.48 | 1.07 | 2.14 | 1.28 | 2.18 | 1.24 | 2.14 |
| 朝食 | 1.24 | 2.14 | 1.21 | 1.22 | 0.99 | 0.73 | 0.97 | 0.74 | 1.07 | 1.21 |

についてレビュー全体との出現割合の差が小さく, 特徴的な語の存在が期待できない. これは, 夫婦のみレビュー群に「朝食が美味しかった」など一般的な内容を書いているレビューが多いことが原因であると考えられる. 子供連れレビュー群では「子供」が出現割合 1 位であり (表 2), 全体での出現割合に比べて高い値を示すことから, 子供連れレビュー群の特徴を表す語に「子供」が含まれることがわかるが, 幼児連れレビュー群と赤ちゃん連れレビュー群でも「子供」の出現割合は高いため, 「子供」は子供連れレビュー群のみがもつ特徴とはいえない. 子供連れレビュー群の特徴語候補は全体と出現割合に大きな差がなく, 子供連れレビュー群を自動推定することは困難と考えられる. それに対して, 幼児連れ, 赤ちゃん連れ, 高齢者連れでは上位 10 語のみでも全体での出現割合と差のある特徴語候補が見られ, これらについては同行者によってレビューの記載内容に特徴があるものと推察できる.

3.2. 旅行目的に関する特徴語の抽出

旅行目的の違いによるレビューの差異を検証するため, レビュー群から家族旅行, 新婚旅行, 卒業旅行, 女子会, ライブ遠征の 5 種類に当たるレビューを 500 件ずつ選び, それぞれのレビュー群の名詞の出現状況を調べた (表 3). 家族旅行のレビュー群では「子供」の

表 3: 旅行目的 5 種類の出現割合上位 10 語

| 家族旅行 | | 新婚旅行 | | 卒業旅行 | | 女子会 | | ライブ遠征 | | |
|------|------|------|------|------|------|------|------|-------|------|------|
| 特選候補 | 全体 | 特選候補 | 全体 | 特選候補 | 全体 | 特選候補 | 全体 | 特選候補 | 全体 | |
| 部屋 | 2.91 | 3.58 | 2.82 | 2.90 | 2.86 | 2.90 | 3.14 | 2.90 | 4.28 | 0.12 |
| 子供 | 2.75 | 0.43 | 2.47 | 3.58 | 2.56 | 3.58 | 2.49 | 3.58 | 3.43 | 2.90 |
| 利用 | 2.55 | 2.90 | 2.24 | 1.51 | 2.26 | 1.51 | 2.26 | 1.51 | 2.72 | 3.58 |
| ホテル | 2.40 | 2.18 | 2.10 | 2.18 | 1.93 | 2.18 | 1.86 | 1.48 | 2.65 | 2.18 |
| 宿泊 | 2.16 | 1.51 | 1.91 | 0.73 | 1.82 | 0.05 | 1.85 | 1.49 | 2.45 | 1.51 |
| 風呂 | 1.77 | 1.49 | 1.82 | 1.48 | 1.57 | 1.48 | 1.79 | 1.51 | 1.42 | 2.14 |
| お部屋 | 1.76 | 0.73 | 1.47 | 1.49 | 1.54 | 1.49 | 1.75 | 2.18 | 1.29 | 1.49 |
| 食事 | 1.42 | 1.48 | 1.35 | 0.74 | 1.31 | 1.21 | 1.42 | 1.21 | 0.93 | 0.73 |
| 朝食 | 1.16 | 2.14 | 1.25 | 0.55 | 1.26 | 2.14 | 1.20 | 0.41 | 1.07 | 0.94 |
| 満足 | 1.10 | 1.21 | 1.20 | 1.21 | 1.21 | 0.73 | 1.18 | 2.14 | 0.95 | 1.48 |

出現割合がレビュー全体と比べて高く, 家族旅行のレビュー群は子供の話題が多いことが考えられる. しかし, 3.1 節の子供連れレビュー群と同様に, 特徴語候補の出現割合は全体での割合と大きな差は見られず, 家族旅行特有のレビューが多いとは考えにくい. それに対して卒業旅行のレビュー群や女子会のレビュー群では全体とは出現割合に差のある特徴語候補が散見され, 特有の話題の存在が期待できる.

3.3. 同行者と旅行目的に基づくカテゴリ

本稿では, 同行者の異なるレビュー群として夫婦のみ, 子供連れ, 幼児連れ, 赤ちゃん連れ, 高齢者連れの 5 種類 (3.1 節), 旅行目的の異なるレビュー群として家族旅行, 新婚旅行, 卒業旅行, 女子会, ライブ遠征の 5 種類 (3.2 節) について, それぞれ特徴的な語の有無を検討した. その結果, 子供連れや高齢者連れなどのように同行者に特徴が見られる場合も, ライブ遠征

や卒業旅行など旅行目的に特徴が見られる場合もある反面, 同行者や旅行目的に明確な特徴が見られない場合もあり, 同行者のみ, 旅行目的のみでのカテゴリ分類は適切ではないと考えられる. 特徴を表す語が少ない原因として, 出現頻度の割合のみではすべてのカテゴリのレビュー群に現れるような一般的な語が多いことが考えられるため, 固有のカテゴリのみで出現する単語を考慮する必要があると考えられる. 本稿ではあらかじめ複数の同行者, 複数の旅行目的を想定してレビュー群を作成したが, 今後は特徴的な同行者, 特徴的な旅行目的の自動推定も含め, 適切なカテゴリの設定について検討する必要がある.

4. まとめ

本研究では, 選択した宿泊施設がユーザのニーズに合わないことが原因で不満の残る宿泊になるユーザがいることから, 宿泊施設を利用した他ユーザの状況や特徴を含む感想を提供することで, ユーザが自身のニーズに合った宿泊施設を選択する際に基準のひとつとして参考にすることができると考える. 一般的な評価項目のみでは, ユーザが宿泊施設を選択する際に, 自身のニーズに合う宿泊施設かどうかを判定することは困難なため, 本稿では同行者と旅行目的に基づいたカテゴリについて検討した.

5 種類の同行者と 5 種類の旅行目的のカテゴリの各レビュー群から出現頻度の割合を求めた結果, 同行者と旅行目的それぞれに基づく特徴が現れるカテゴリがあるのに対し, 特徴を表す語がないカテゴリもあることがわかり, 同行者のみ, 旅行目的のみに基づいたレビューのカテゴリ分類は適切ではない. そのため同行者と旅行目的の自動推定が必要と考えられる.

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楽天グループ株式会社からご提供いただいた「楽天データセット」(https://rit.rakuten.com/data_release/) を利用しました. 心より感謝いたします.

参考文献

- 1) 楽天グループ株式会社 (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. (データセット). <https://doi.org/10.32130/idr.2.2>.
- 2) 伊草 久峻, 島海 不二夫, “宿泊予約サイトにおけるレビュー自動分類,” 第 34 回人工知能学会研究発表会, No.1L4-GS-5-04, pp.1-4, 2020.
- 3) 高師 遥, 阿部 遥佳, 鈴木 信太郎, 大和矢 悠仁, 延澤 志保, “旅館レビューのカテゴリ分類のための重文自動分割,” 情報処理学会第 84 回全国大会, no.6V-08, pp.755-756, 2022.
- 4) 坂野 妃菜, 高師 遥, 鈴木 信太郎, 延澤 志保, “レビューに基づく宿泊施設の客層推定,” NII-IDR ユーザフォーラム, No.S04, 2021.
- 5) 吉田 朋史, 北山 大輔, “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,” DEIM Forum, No.B2-1, pp.1-5, 2015.
- 6) Zhang Bo, 白井 清昭, “レビューテキストの書き手の評価視点に対する評価点の推定,” 言語処理学会第 23 回年次大会, pp.803-806, 2017.
- 7) 阪井 奎伍, 灘本 明代, “観光を対象としたレビューからの耳より情報抽出,” 情処研報, Vol.2015-DBS-162, No.13, pp.1-6, 2015.