

同行者と旅行目的に基づく旅館レビュータグ付け

坂野 妃菜*

谷口 雅空*

鶴島 仁人*

延澤 志保*

*東京都市大学知識工学部

*東京都市大学情報工学部

1 はじめに

1.1 研究背景

旅館レビューには宿泊施設の部屋や食事などの評価項目に関する不満だけでなく、「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」といった利用するユーザの状況が宿泊施設に合わないことによる不満も含まれている。

利用目的に基づくレビュー分類の提案では、飲食店の料理に関するレビューとホスピタリティに関するレビューを学習データとした結果、ユーザが満足する飲食店の特性が利用目的に応じて変化することが示された [1]。情報推薦の分野で重要視されているコンテキストである同行者によってトピックが変化することに着目した研究では、同行者依存のトピックモデルを構築した結果、同行者の種類によって文書の傾向が変化することが示された [2]。旅館レビューを用いて同行者5種類、旅行目的5種類ごとのレビュー群での語の出現頻度と全体の語の出現頻度の差の分布を調査したところ、同行者と旅行目的それぞれの項目に特徴が見られるものがあるということが分かった [3]。この結果から、同行者と旅行目的を用いたレビューのタグ付けはユーザのレビュー検索の支援において有用であると言える。

1.2 旅行要件タグの提案

観光関連旅行 web サイトのうち、2020 年の年間閲覧者数が上位 6 位¹までのサイト (じゃらん net², 楽天トラベル³, トリップアドバイザー⁴, JR 東日本⁵, 全日本空輸 (ANA)⁶, Yahoo!トラベル⁷) に掲載されている特集記事カテゴリを表 1 に示す。ただし特集記事の記載のないサイトは表 1 では割愛する。表 1 の特集記事見出し欄は、各サイトの同等の特集記事をまとめたもので、同等の見出しが 4 サイトすべてに掲載されたもの、3 サイトに掲載されたもの、2 サイトのみに掲載されたものの順に並べてある。表 1 では 1 サイトのみに掲載のあった特集記

表 1: 本研究で提案する旅行要件タグ

旅行要件タグ	特集記事見出し			
	じゃらん net	楽天トラベル	全日本空輸	Yahoo!トラベル
カップル	夫婦旅行	夫婦におすすめ旅行	夫婦・カップル旅行	カップル・夫婦に人気
ペット連れ	ペットと泊まれる宿	ペットとの旅行	愛犬と泊まる上質な宿	ペット泊まれる宿
プール	プールが楽しい宿	プールのある宿	プール付きホテル	プライベートプールが人気
温泉	温泉旅行	全国の温泉宿	温泉旅行	温泉自慢の宿
夏休み	夏休み旅行	夏休み旅行	夏旅行	夏休みの家族旅行
絶景	絶景の宿	絶景スポット 46 選	絶景を望む宿	絶景が話題の宿
キャンプ	自然観光・キャンプ	自然体験・グランピング	グランピング	おすすめのグランピング
子供連れ	小学生が喜ぶ宿	子供連れに人気の宿	子供連れに人気の宿	お子連れに人気の宿
観光	観光スポットが近い宿	駅・空港近く	歴史観光	有形文化財指定の宿
駅近	アクセスしやすい宿	駅・空港近く	駅・空港近く	
三世代	三世代におすすめ	三世代旅行	三世代旅行	
結婚記念	ハネムーン	結婚記念	結婚記念	
赤ちゃん連れ	赤ちゃん大歓迎の宿	赤ちゃん歓迎の宿	新婚旅行	赤ちゃんとの旅行で人気
海水浴	海で遊べる	海水浴のある宿		プライベートビーチが人気
女子旅	女子旅	女子旅		女子旅に人気
学生	農業体験	大学生旅行	大学生におすすめの宿	卒業旅行
味覚狩り	味覚狩り	味覚狩り		フルーツ狩り
バイキング	バイキングが人気	バイキング		
コスパ			コスパが良い宿	コスパが人気の宿
高齢者	シニア旅行		シニアに人気の宿	シニアに人気の宿
ゴルフ			ゴルフ特集	ゴルフにおすすめ
エステ	スパ/エステおすすめの宿			スパ・エステが人気
平日		平日旅行	平日旅行	

事は記載していない。本稿では、表 1 の特集記事見出しを整理し、旅行要件タグとして提案する。

表 2 は本稿で提案する旅行要件タグと楽天トラベル既存カテゴリの対応を調べた結果である。楽天トラベル表 2: 旅行要件タグと既存カテゴリ

旅行要件タグ	楽天トラベル既存カテゴリ			
	旅館レビュー	同行者	目的	評価
カップル	6,164	家恋	遊	サ地室設浴食
ペット連れ	6,101	家恋	遊	サ地室設浴食
プール	3,700	家	遊	サ地室設浴食
温泉	3,668	家恋友職独	遊	サ地室設浴食
夏休み	3,257	家恋友	遊	サ地室設浴食
絶景	2,545	家恋独	遊	サ地室設浴食
キャンプ	1,110	家	遊	サ地室設浴食
子供連れ	9,983	家友	遊	サ地室設浴食
観光	9,967	家恋友独	遊	サ地室設浴食
駅近	6,130	家恋友職独	遊	サ地室設浴食
三世代	4,253	家	遊	サ地室設浴食
結婚記念	3,827	家	遊	サ地室設浴食
赤ちゃん連れ	2,627	家	遊	サ地室設浴食
海水浴	1,888	家恋友	遊	サ地室設浴食
女子旅	821	家友職	遊	サ地室設浴食
学生	563	家友	遊	サ地室設浴食
味覚狩り	10	家	遊	サ地室設浴食
バイキング	9,967	家恋友	遊	サ地室設浴食
コスパ	9,624	友職独	遊	サ地室設浴食
高齢者	4,122	家友	遊	サ地室設浴食
ゴルフ	2,854	家友職独	遊	サ地室設浴食
エステ	948	家友職	遊	サ地室設浴食
平日	249	独	遊	サ地室設浴食

(同行者) 家: 家族, 恋: 恋人, 友: 友人, 職: 仕事仲間, 独: ひとり。(目的) 遊: レジャー, ビ: ビジネス。(カテゴリ) サ: サービス, 地: 立地, 室: 部屋, 設: 設備・アメニティ, 浴: 風呂, 食: 食事。

データセット [4] の 2017 年から 2018 年までのレビュー 200,000 件について対応する旅行要件タグを調べた結果、旅行要件タグにはそれぞれのレビューに付与された同行者情報、目的情報、評価情報 (表 2 下部参照) が複数種類含まれ、かつ旅行要件タグ同士で既存カテゴリが重複していることが分かる。表 1, 表 2 より、特集記事カテゴリを旅行要件としてレビューにタグ付けすることでユーザが同行者と旅行目的でのレビュー検索が可能になると考えられる。

Accommodation Review Tagging Based on Travel Party and Purpose.

Hina Sakano*, Masataka Taniguchi*, Takashi Tsurushima*, and Shiho Hoshi Nobesawa*.

* Faculty of Knowledge Engineering, Tokyo City University

* Faculty of Information Technology, Tokyo City University

¹ 観光経済新聞, <https://www.kankokeizai.com/> 日本観光振興協会「2020 年観光関連サイト閲覧者数」.

² じゃらん net, <https://www.jalan.net/>.

³ 楽天トラベル, <https://travel.rakuten.co.jp/>.

⁴ トリップアドバイザー, <https://www.tripadvisor.jp/>.

⁵ JR 東日本, <https://www.eki-net.com/>.

⁶ 全日本空輸, <https://www.airtrip.jp/air-lp/>.

⁷ Yahoo!トラベル, <https://travel.yahoo.co.jp/>.

2 レビューの旅行要件タグ付け手法

本研究では、特集記事カテゴリから得られた旅行要件タグを用いて旅館レビューにタグ付けを行うことで、同行者と旅行目的を用いたレビュー検索を可能にし、ユーザーの宿泊施設選択の支援を目指す。本研究では旅館レビューを入力し、学習器を用いてレビューへの旅行要件タグ付けを行う。

本システムでは、まず、各旅行要件のレビュー群から特徴語と特徴語の周辺にある語を周辺語として抽出する。形態素解析エンジン MeCab を用いて品詞情報に自立、サ変接続、ナイ形容詞語幹、一般、形容動詞語幹、固有名詞、接続詞的を含む名詞を対象にレビューから抽出し、各旅行要件のレビュー群での語の出現頻度と全旅行要件を含むレビュー群での語の出現頻度の差が 0.003 以上の語を特徴語とする。周辺語は各旅行要件レビュー群での特徴語を含む 1 文中の名詞、動詞、形容詞、形容動詞を対象とする。レビュー群から得られた特徴語と周辺語を素性として用いる。レビューを形態素ごとに分け、1 文中での特徴語と周辺語の出現頻度を求め、得られた出現頻度を各レビューの文書ベクトルとする。文書ベクトルを用いて旅行要件レビュー群の学習を行いレビューを分類する。当てはまる旅行要件タグをレビューに付与していき、タグ付きレビューを出力する。

3 旅行要件タグ付け実験

3.1 実験の概要

本研究では、表 1、表 2 で挙げた旅行要件タグのうち、楽天トラベルの旅館レビューが 500 件未満である味覚狩りと平日を除いた 21 種類の旅行要件タグをタグ付けの対象とする。

楽天トラベルデータよりレビュー 87,499 件を対象に SVM を用いて 2 値分類モデルの学習を行った。この際 10 分割交差検証を行い分類モデルの平均精度を測った。

学習データに偏りがあった場合、正事例と負事例のどちらかに学習結果が引きずられる可能性があると考え、学習データに偏りがないように正事例のデータ数に合わせて負事例のデータ数を同数とした。複数種類のタグの要素を含むレビューは当てはまるタグすべての正事例データに含み、正事例と負事例に同じレビューが重複しないよう、正事例とする対象の旅行要件が含まれないレビューのみを負事例とした。レビュー 87,499 件の文字列平均を調べた結果、1 レビューあたりの平均は 212.32 文字であり、「良かったです。」などの平均文字列より極端に少ない 10 文字以下の短いレビューにはユーザーの状況や特徴が含まれていないため対象としなかった。

3.2 旅行要件タグ付けの評価

楽天トラベルデータセット [4] に収録されているレビュー群から 3.1 節で学習に用いたデータ 87,499 件とは異なるレビューを対象とした。各旅行要件に当てはまるレ

ビュー各 169 件を用いてレビューのタグ付けを行い、各旅行要件において正しく予測できたレビュー数を調べた結果を表 3 に示す。表 3 はそれぞれの旅行要件の混同行列

表 3: レビューの旅行要件タグ付けの混同行列

	旅らん	温泉	ペット	夏休み	キャンプ	プール	カップル	新婚	海外	観光	学生	ハイキング	ユスバ	温泉	ゴルフ	結婚記念	子供	三世	女子	エス		
旅らん	167	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
温泉	0	152	0	1	0	2	2	0	1	0	1	4	0	0	0	1	2	0	0	0	3	0
ペット	0	1	145	0	2	1	3	0	0	0	0	1	0	0	3	1	1	1	0	0	3	6
夏休み	0	0	0	143	2	2	2	0	2	1	1	0	0	1	1	2	2	1	9	0	0	0
キャンプ	1	2	0	1	134	0	2	2	1	4	0	0	2	4	3	3	2	1	2	5	0	0
プール	0	2	0	0	1	125	1	1	4	1	0	3	1	1	1	2	9	3	7	3	0	0
カップル	0	2	3	0	0	1	125	0	0	1	4	3	3	3	1	14	1	0	1	5	1	1
新婚	0	1	0	0	2	0	1	125	1	15	2	2	6	0	1	2	0	1	0	9	1	1
海外	1	0	1	3	4	5	2	0	124	6	0	2	0	2	0	7	1	0	5	6	0	0
観光	0	0	0	0	2	0	0	6	0	123	1	4	1	3	1	8	4	2	5	9	0	0
学生	1	0	2	0	1	1	1	1	0	5	123	2	0	9	0	10	2	2	6	3	0	0
ハイキング	0	0	1	1	0	2	0	5	1	4	0	117	3	11	0	2	4	0	4	14	0	0
ユスバ	0	0	2	0	0	1	2	6	3	13	1	8	107	3	1	2	2	0	4	14	0	0
温泉	0	1	0	0	1	1	2	1	2	4	0	6	0	105	2	18	6	2	5	12	1	1
ゴルフ	0	1	0	0	4	1	4	2	1	12	3	3	1	7	99	11	4	1	7	8	0	0
結婚記念	3	1	1	0	1	0	13	5	1	5	0	6	0	11	1	59	6	1	10	6	1	1
子供	0	0	5	4	1	3	2	5	5	1	4	0	3	1	12	87	8	26	2	0	0	0
三世	1	0	0	0	2	1	1	5	1	9	1	5	0	6	1	16	8	72	12	31	0	0
女子	0	1	0	3	3	6	2	5	1	6	3	5	0	17	1	16	22	10	59	11	2	2
エス	1	3	2	1	2	2	2	5	2	11	5	4	2	2	10	26	4	2	23	57	5	5
エス	0	0	1	0	1	2	3	5	2	14	0	3	2	20	0	20	6	2	7	27	50	0

列であり、行が正解タグのレビュー数、列が予測タグのレビュー数を示す。全旅行要件のうち、各旅行要件に当てはまるレビュー群の半数以上に正しいタグ付けができた旅行要件を調査するため、表 3 ではレビュー数が 169 件に近いほど赤色、169 件の半数である 84.5 件に近いほど青色、0 に近づくほど白色とした。表 3 より学生、三世代、女子旅、エスエといった他の旅行要件と混同されやすいものもあるが、多くのレビューが正しいタグ付けに成功していることが分かる。

4 まとめ

同行者と旅行目的ごとに旅館レビューを分けることで他ユーザーの意見を参考にでき、自身に合った宿泊施設選択の支援になると考えられる。本研究では、同行者や旅行目的を旅行要件タグとして特集記事カテゴリから集計し、レビューのタグ付けを行った。各旅行要件のレビュー群から抽出した特徴語のみでは単語同士の関係が考慮されずに誤判定されるため、特徴語を含む 1 文中の名詞、動詞、形容詞、形容動詞を周辺語として特徴語と周辺語を素性として用いた。各旅行要件の 2 値分類モデルを用いてレビューのタグ付けを行った結果、多くのレビューがタグ付けに成功したことから特集記事カテゴリを旅行要件タグとして用いることは有用であると言える。

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様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] 山西 良典, 藤岡 寛子, 西原 陽子, “擬似コーパスを用いた飲食店レビューの観点の自動分類,” 人工知能学会論文誌, Vol.36, No.1, WI2-A, pp.1-8, 2021.
- [2] 深澤 佑介, 太田 順, “同行者に応じたトピックモデル,” 情報処理学会論文誌, Vol.55, No.1, pp.413-424, 2014.
- [3] 坂野 妃菜, 谷口 雅空, 鈴木 信太郎, 工藤 竜矢, 戸内 伶雅, 酒井 終輔, 延澤 志保, “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出,” 第 21 回情報科学技術フォーラム (FIT2022), no.6D-1, 2022.
- [4] 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>