

旅客施設のレビュー解析に基づく性格情報を用いた嗜好分析

Preference Analysis using Personality Information based on Review Analysis in Passenger Facilities

経営システム工学専攻
20N7100010F 眞保 祐樹

1. 序論

新型コロナウイルスの流行により、国内の旅行市場は大きな打撃を受けている。2020年(令和2年)の日本人の国内宿泊旅行者数は延べ1億6,070万人(前年比48.4%減)、国内日帰り旅行者数は延べ1億3,271万人(前年比51.8%減)と大きく減少した¹⁾。一方で感染拡大以前から、観光業界では団体旅行の減少や後継者不足といった課題を抱えている。観光庁では、令和2年度第3次補正予算にて新たな補助制度の創設や融資制度の大幅拡充を行い、観光施設を再生し地域全体の魅力や収益力を高める働きかけを行なっている²⁾。以上の背景を持つ観光業界においては、今一度ニーズに注目し利用者に寄り添う施設や施策の創出を行い、収益構造をより強固にすることが求められる。現代ではSNSや旅客サイトなどの普及により、利用者との接点は時間や客層を問わず無数に広がっている。従って、最大限にデータを活用することで利用者のニーズを精密に把握することが可能である。これをサービスに十分に反映し、個々の収益構造を大きく改善することで業界全体を復興へ近づけることができる。

本研究では上記に示した課題意識を基に、研究対象を旅客施設として、利用者の性格に注目した嗜好分析を行う。本手法の狙いは、性格という誰もが普遍的に持ちそれぞれが異なる属性値を用いることで、網羅的かつパーソナライズした分析を実現することである。なお、本研究では旅客サイトのレビューテキストから性格情報を推測することで、大数データを用いた分析を可能とした。本分析の目的を、性格特性Big5(開放性、誠実性、外向性、協調性、情動性)と施設特徴の結びつきを可視化すること、性格情報を用いた嗜好分析手法の提案の2点として、それぞれを「旅客施設における利用者の性格情報と嗜好の関係性の分析」、「性格情報を用いた宿泊満足度予測モデルの提案」と題して行なった。これらの結果を基に、嗜好分析における性格の属性値への適用可能性や性格特性を考慮した満足度予測モデルの適用可能箇所について考察し、本手法の有用性を検証する。

2. 研究対象データと分析データセットの作成方法

研究の対象データには、旅客施設の予約に用いられる楽天トラベルのレビューデータ³⁾より、2014-19年の5年分を抜粋して使用した。対象施設は十分なレビューデータ数を保持するレビュー投稿数上位1000施設とした。なお、

好印象の体験を分析対象とするため感情極性を推測できるasari⁴⁾を用いて好印象のデータを抜粋した。また、施設体験の統一化のため食事が未評価のデータは削除した。

各施設の特徴を定義するためにレビューデータより、施設ID、部屋名前、同伴者、評価(立地、部屋、食事、風呂、サービス、設備、総合)、ユーザ投稿本文の11項目を抜粋した。また、レビューデータに含まれていない所在地、利用最低料金、レビュー総数、特色の4項目は楽天トラベル施設情報APIにより追加した。さらに、利用者の性格情報や目的・関心をユーザ投稿本文から推測し追加した。まず、性格情報の推測では、テキストから性格特性Big5を予測するIBM Personality Insight⁵⁾を用いて、利用者の性格特性値を数値ベクトル化した。次に、利用者の目的・関心の推測では、日本語自然言語処理ライブラリであるGinza⁶⁾を用いて固有表現抽出を行い、レビュー文章内の語句に固有表現タグを付与し、施設に付随するタグの付与数を解析した。今回考慮したタグは、「職種名」、「食べ物」、「温泉」、「街」、「施設」、「金額」の6つであり、表現上近いタグをそれぞれに集約した。また、「清潔さ」、「観光名所」のタグを独自に作成し、施設に対する語句を最大限拾えるよう工夫した。以上の手順から、計45の特徴量を持つデータセット(表1)を作成した。各特徴量の取り得る値に関して簡単に説明する。同伴者、部屋に関しては、利用割合や稼働率を算出した。目的・関心は、出現数上位5つのタグに対し1位に5、2位に4、...、5位に1、ランク外に0として配点した。所在地は、データ数の差を考慮するため地方に集約しダミー変数により0,1で示した。

3. 分析方法

3.1. 利用者の性格情報と嗜好の関係性の分析

本分析の目的は、性格特性Big5と施設特徴の関係性を明確化することである。ここでは、データセットにおいてGraphical lassoを用いた相関行列を算出することで、性格特性値と施設特徴因子の関係性を明かす。なお、本モデルを用いることでスパースな相関行列を算出し、因子間の疑似相関を取り除く事を期待する。また、本モデルは多変量正規分布を仮定していることから、データの前処理としてYeo-Jonson変換を行った。この算出結果を用いて性格特性と施設特徴因子の関係性を可視化し考察を行う。なお、モデルにおける正則化パラメータは、交差検証を行い尤度が最大となるものを選択した。

表 2 各施設群にて作成したモデルにおける重要度の高い施設特徴因子

	important_feature1	importance1	important_feature2	importance2	important_feature3	importance3	important_feature4	importance4	important_feature5	importance5	Sum of importance
cluster1	Min_Charge	0.3589	Roomtype_Single	0.0964	Follower_None	0.0896	Roomtype_Triple	0.0721	Follower_Lover	0.0464	0.663
cluster2	Min_Charge	0.2108	Num_of_reviews	0.1639	Follower_Family	0.1148	Follower_Friend	0.0861	Roomtype_Etc	0.0751	0.651
cluster3	Concern_Money	0.1857	Concern_Clean	0.1406	Follower_Friend	0.1032	Roomtype_Triple	0.0966	Roomtype_Twin	0.0811	0.607
cluster4	Follower_None	0.3198	Follower_Family	0.1254	Roomtype_Single	0.0985	Num_of_reviews	0.0686	Concern_Onsen	0.0529	0.665
cluster5	Min_Charge	0.1693	Follower_Lover	0.1297	Roomtype_Triple	0.1182	Roomtype_Twin	0.1076	Follower_Friend	0.0788	0.604
cluster6	Follower_Lover	0.1188	Concern_FacilityPart	0.1012	Roomtype_Etc	0.0934	Min_Charge	0.0924	Roomtype_Japanstyle	0.0899	0.496
cluster7	Follower_Family	0.2559	Roomtype_Single	0.1310	Roomtype_Twin	0.1141	Num_of_reviews	0.0680	Follower_Friend	0.0660	0.635
cluster8	Follower_Lover	0.3192	Follower_Friend	0.1009	Roomtype_Japanstyle	0.0716	Roomtype_Etc	0.0568	Roomtype_Twin	0.0560	0.604
cluster9	Min_Charge	0.4712	Roomtype_Double	0.0789	Follower_Friend	0.0783	Follower_Lover	0.0637	Num_of_reviews	0.0422	0.734
cluster10	Num_of_reviews	0.1196	Min_Charge	0.1105	Concern_FacilityPart	0.0766	Roomtype_Twin	0.0758	Follower_Lover	0.0752	0.458
cluster11	Follower_Lover	0.1672	Min_Charge	0.1341	Roomtype_Japanstyle	0.1248	Concern_Money	0.0802	Roomtype_Single	0.0729	0.579
cluster12	Min_Charge	0.1970	Roomtype_Single	0.1498	Follower_Lover	0.1122	Follower_Friend	0.0920	Roomtype_Etc	0.0743	0.625

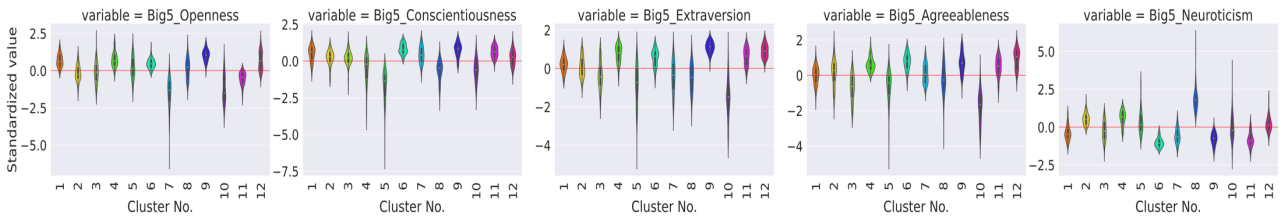


図 2 クラスタリングの結果

り、レビュー評価値自体が施設選択において大きな影響を持っており、外向性の高い人ほどそれに惹かれ施設選択を行うことが示唆される。

一方で、所在地に関しては性格特性値との間に相関が見られなかった。日本においては、多くの場合、ある地方での観光体験が他の地方でも体験可能である。例えば、温泉地といっても東北地方では銀山温泉、近畿地方では有馬温泉、九州地方では下呂温泉が存在しており大義的に温泉を楽しむという同一の体験を行うことが可能である。このような背景から、所在地における特色が薄く、性格特性値との相関が示されなかったと考えられる。一方で、本分析においては所在地ごとのデータ数の差を考慮し地方レベルに集約し分析を行なったが、より詳細に分割できれば、性格特性値と所在地に相関が見られる可能性もある。

4.2. 性格情報を用いた宿泊満足度予測モデルの提案

4.2.1. 満足度のモデリングにおける重要因子の妥当性

はじめに、k-means++法によるクラスタリングを用いて、利用者の性格特性を基に施設の層別を行なった。なお、シルエット分析により、分割数を12とした結果を最適なクラスタリングと判断し施設群を作成した。作成された各群における利用者の性格特性値の分布（図4）をバイオリンプロットにより評価する。いくつかのクラスタにおいて外れ値のような施設が存在するが、平均付近にボリュームを持つ分布が得られた。従って、性格特性値において概ねまとまりを持つ施設群を作成できた。

次に、各施設群にて作成された満足度モデルにおいて、重要度の高い因子として選択された施設特徴因子（表2）を基に、各モデルの妥当性の評価を行う。まず、全施設群を横断し考察を行う。重要度の高い因子には、所在地を除いた全てのカテゴリの因子が選択された。所在地に関して

は、4.1.の結果において性格との関係性が見られなかった因子であるが、満足度に対しても寄与が低い因子であることが明かされた。一方で、最低価格の因子は、重要度の高い因子として広く選択された。日本政策金融公庫による消費者調査において、旅行の際に約90%の人が価格を重視して施設の選択を行う¹⁰⁾という調査結果がある。大多数の人が施設選択において重視する因子であり、満足度への寄与度が高い因子として広く選択されていることは妥当と言える。他の因子においては、同伴者、部屋の形態に関する因子があり、旅行のシチュエーションや旅客施設の雰囲気に対する嗜好が含まれていると考えられる。これらの因子は性格との関係性が既に示されており、性格特性が異なる施設群においては各因子の満足度への振る舞いが異なる事が予想される。

各施設群で作成されたモデルに対する個々の評価に関しては、重要度の高い因子として選択された施設特徴因子を基に、4.1.の結果を用いて嗜好の解釈を行い実際のデータにおける傾向との乖離を検討した。この結果より、利用者の性格特性において開放性、情動性の分布が平均値付近の施設群においては、嗜好の解釈が困難であることが明らかとなった。この一例として、クラスタ5の結果を挙げる。クラスタ5は開放性、情動性ともに平均やや上部に分布する性格特性を示しており、重要度の高い因子には最低価格、同伴者（恋人、友達）、部屋の形態（トリプル、ツイン）が選択されていた。ここで、4.1.の結果を用いて嗜好の解釈を行うと、旅行における予算の設定が高く複数人での旅行を好むとされる。しかし、実際のデータにおける傾向では一人の旅行を好む傾向が見られており、反対の解釈がなされた。従って、本手法により性格による嗜好の差異を反映した満足度モデルを作成できるが、適用範囲には限界があることが示唆される。

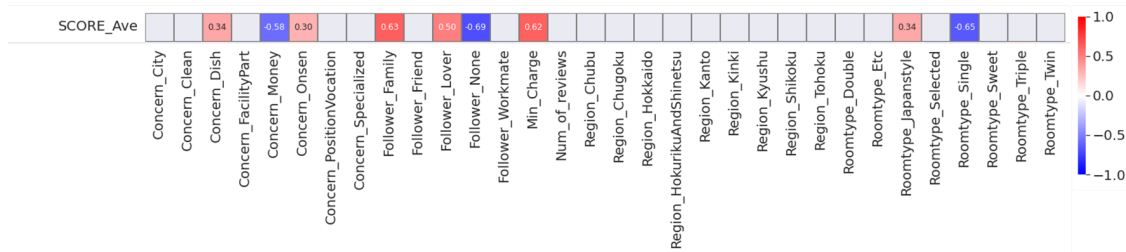


図 3 各施設における予測値の平均値と施設特徴因子の相関行列

表 3 各施設群における予測値の分布

	Cluster No.												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
mean	4.28	4.30	4.16	4.34	4.03	4.31	4.15	4.14	4.39	4.03	4.24	4.33	4.21
std	0.1142	0.0747	0.1238	0.1179	0.1707	0.1019	0.1278	0.1228	0.1283	0.1010	0.1045	0.0918	0.2532
min	3.91	3.92	3.56	3.90	3.48	3.86	3.86	3.66	3.97	3.57	3.93	4.07	3.23
25%	4.20	4.26	4.09	4.28	3.91	4.25	4.05	4.07	4.33	3.98	4.18	4.27	4.08
50%	4.29	4.30	4.17	4.36	4.02	4.32	4.17	4.17	4.41	4.04	4.25	4.34	4.24
75%	4.36	4.35	4.25	4.42	4.14	4.38	4.24	4.23	4.48	4.10	4.32	4.39	4.38
max	4.62	4.58	4.50	4.61	4.50	4.57	4.47	4.45	4.62	4.50	4.57	4.63	4.86

表 4 予測値の標準偏差が大きい施設

施設名	平均評価値	予測値標準偏差
1 熱海温泉 伊東園ホテル熱海館	3.41	0.2444
2 ホテル ドリームゲート舞浜	4.46	0.2326
3 東横イン中部国際空港1	3.31	0.2322
4 ホテルグリーンプラザ上越	3.86	0.2300
5 スパリゾートハワイアンズ モノリスタワー	4.47	0.2281

表5 予測値の最大値と最小値の差異が大きい施設

施設名	平均評価値	予測値差異
1 東横イン中部国際空港1	3.31	0.95
2 熱海温泉 伊東園ホテル熱海館	3.41	0.94
3 定山溪温泉 定山溪ホテル	3.56	0.88
4 アパホテル&リゾート<札幌>	3.44	0.86
5 アパホテル<東京板橋駅前>	3.38	0.85

4.2.2. 性格特性を基とした層別解析による満足度予測

これまで示した結果の応用手法として、性格特性を基にした満足度予測値の層別解析を考える。このために、研究対象の全 1000 施設に対して、各施設群で作成されたモデルを用いて満足度を予測する。この結果より、本手法で作成したモデルの特性を把握し有用性を検討する。はじめに、各施設群における予測値の分布（表 3）より、クラスごとに予測値の平均値に差異があることや元のレビュー評価値に比べばらつきが抑えられていることから、性格特性を基に予測値の層別が行えていることが確認できる。次に、各施設群における予測値の平均値と施設特徴の相関行列（図 3）から見られるように、4.1. にて性格特性との相関が示された施設特徴因子と関係性を持つことがわかる。従って、予測値においても性格特性による差異が十分に反映できている。最後に、本モデルの特性について検討する。各施設における予測値の標準偏差、最大値と最小値の差異を算出した。これらにおいて、高い値を取る施設（表 4, 5）を抜粋すると、大半がレビュー評価値の低い施設であることがわかる。これより、もとよりレビュー評価が低い施設においては、満足度の予測値がばらつき層別されやすいことが示唆される。従って、本手法はレビュー評価の低い施設においてより効力を持ち、利用者における好印象層の特定や普遍的に好まれる施設特徴の分析を行うことができる。

5. 結論

楽天トラベルのレビューデータを解析し、旅客施設における利用者の性格を考慮した嗜好分析を試行した。

「旅客施設における利用者の性格情報と嗜好の関係性の分析」においては、旅客施設における利用者の性格特性と嗜好の関係性を明かした。本結果より、情動性の差異が旅客施設におけるニーズを大きく左右していると示唆され

た。また、性格特性と所在地には関係性が見られないことやレビュー評価値自体が施設選択に大きな影響を与えていることも明かした。

「性格情報を用いた宿泊満足度予測モデルの提案及び有用性の検証」においては、性格特性を基に満足度モデルを層別化することを試みた。この結果より、本手法を行うことで性格特性による差異が反映された満足度モデルを作成できることが示唆された。さらに、予測値を解析することにより、普遍的に好まれる施設の分析や好印象層や低印象層の特定が可能であることを示し本手法の有用性を示した。一方で、嗜好捕捉が難しい層が存在することも明らかとなった。

今後の課題として、結果の実用性を追求することが必要だと考える。そのためには適用範囲を広げることが必要であり、施設特徴因子の細分化や低評価レビューを分析に組み込み多面的な分析を行うことが有効とされる。

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提供を受けた「楽天データセット」を利用した。

参考文献

- 1) 国土交通省. 令和 3 年度観光白書. 本文(第 I 部観光の動向). P18-21
- 2) 国土交通省. 令和 3 年度観光白書. 本文(第 II 部新型コロナウイルス感染症を踏まえた観光の新たな展開). P79-8
- 3) 楽天グループ株式会社(2014):楽天データセット. 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. <https://doi.org/10.32130/idr.2.0>
- 4) Hiroki NAKAYAMA. asari, <https://github.com/Hironsan/asari>
- 5) IBM. personality insights. <https://www.ibm.com/jp-ja/cloud/watson-natural-language-understanding>
- 6) Megagon Labs. Ginza. <https://github.com/megagonlabs/ginza>
- 7) 眞保祐樹, 高橋直己, 浜田百合, 庄司裕子. 旅行施設における決定要因の性格による差異の分析
- 8) 泊 真児, 吉田 富二雄. 性格特性 Big Five と日常生活におけるプライベート空間の 7 機能. 社会心理学研究. 第 16 巻第 3 号. p147-158. 2001
- 9) インターワイヤード株式会社ディムスドライブ事務局. DIMSDRIV. 国内のホテル宿泊に関するアンケート. 2007
- 10) 日本政策金融公庫 国民生活事業本部 生活衛生融資部. 国内宿泊施設の利用に関する消費者意識と旅館業の経営実態調査. 国内旅行で利用する際のポイント. p16. 2013