

宿泊施設のレビューに基づいた穴場ホテル探索

松井 悠太¹ 伊藤 正彦^{1,a)}

概要: 現在のホテルレビューサイトでマイナーだが評価の高い宿を探そうとすると、探したい宿以外にも候補として多数上がってしまうことがあり、効率よく宿が探しにくい場合がある。本論文では、レビュー数が少なくマイナーではあるが、食事などの評価の高い宿を穴場ホテルとし、可視化しながら穴場ホテルを探索できるツールの構築を目指す。検索条件として、レビュー件数、レビュー評価項目、レビュー評価数値を用いたホテル探索ツールを作成した。結果ある程度目的とするマイナーで評価の高い宿泊施設を探し出せるツールの作成に成功した。

1. はじめに

旅行をするときに心の置きになりえるのが宿探しである。実際、2018年行われた全国意識調査 [1] では旅行のストレスを感じる要因として「飛行機やホテルの予約の手続きが面倒なため」という回答が多くあった。多くの場合大手レビューサイトを使用すると思うが、自分が探す宿をすぐに見つけることは時として非常に難しい。特に、地方観光において食事がおいしかったり景色がよかったりといった宿泊施設でペンションや個人経営の民宿などを探したい場合は非常に難航する。

本研究ではこういったレビュー数が低くマイナーで、評価スコアが高い宿泊施設を穴場ホテルと定義しているが、ホテルのレビューデータのレビュー件数やレビュー評価をグラフによって可視化する。それにより効率的に穴場ホテルを探索することができるツールの開発を目指した。

2. 関連研究

吉田らは宿泊レビューデータを用いて不満レビューのみを抽出する新たな手法を提案した [2]。宿泊業の不満傾向把握や不満解消策を検討し易くすることを目的として、従来の感情極性辞書による分析よりも正確にレビューから不満点を抽出することを目指したものである。自然言語処理モデルのBERTなどを使い、不満レビューを抽出し、従来の感情極性辞書を用いた方法と比較した。結果従来の方法よりも正確に不満点が抽出できた。[2]の研究では宿泊施設の不満傾向の把握を目的としているが、本研究では穴場ホテルの探索を目的としている点で大きく異なる。

野本らは観光レビュー文を用いて穴場スポットを発見する新たな手法を提案した [3]。観光において地方には、人気な観光地と同等の魅力を持ちながらも十分認知されていない観光地（穴場スポット）があるとして、そういった観光地を地方観光の際に旅行者が発見しやすくすることで地方の観光業を活性化させることを目的としている。大手レビューサイトを使用し、認知が十分されていない観光地と魅力度が人気な観光地と同等にある個所を分析によって分類することで穴場スポットを抽出した。結果穴場スポットとして妥当な観光地を抽出できた。[3]と本研究とは穴場スポットを発見する手法が異なっている。[3]では穴場スポットを発見する際に高認知度スポットとその他のスポットに分類する方法を使用しているが本研究ではレビュー数によって穴場を発見している。[3]では観光地の穴場スポットを発見することを目指しているが本研究ではあくまでもホテルに注目している。

3. レビュー数と評価スコアに基づいたホテル探索

3.1 手法の概要

本研究では、レビュー数が低く、評価の高いホテルを穴場ホテルと定義している。主に穴場ホテルの探索を目的として、探索項目に都道府県、レビュー数、評価項目、評価値を設定している。各項目を選択すると表示される結果にはレビュー評価、ホテル名、地域名、レビュー本文などが含まれる (図 1)。

3.2 ホテルレビューデータ

本研究では、楽天トラベルのレビューデータセットを使用する。

¹ 北海道情報大学

^{a)} imash@do-johodai.ac.jp

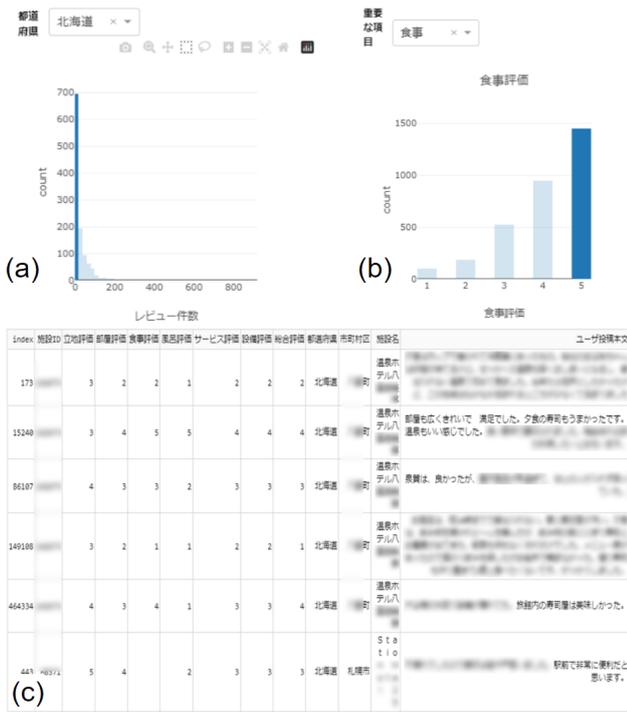


図 1 レビュー数と評価スコアで穴場ホテルを探索するシステムの全体図

- レビューデータ：2019 年の 526262 件のデータを使用する。本研究では施設 ID、立地評価、部屋評価、食事評価、風呂評価、サービス評価、設備評価、総合評価、レビュー本文を使用する。各評価には 5 段階の数字評価が含まれる。
- ホテル施設名データ：レビューデータにある施設 ID に対応した施設名が入っているデータで、このデータを上記レビューデータと組み合わせることで各評価と施設名を紐づける。また、施設名には都道府県名が含まれないため、geocoding を使用し施設ごとの都道府県を求めた。これにより、レビューデータレビューデータに都道府県を追加する。

3.3 レビュー件数による絞り込み

具体的な選択方法としてはまず、ユーザには探すホテルの地域とレビュー件数を選択してもらう (図 1(a))。都道府県を選択するのは、ホテル探しとしての一つの項目といった意味もあるが、レビュー数を選択するときに地域で絞っていないと、目的とする穴場ホテルの条件であるレビュー件数が低くマイナーなホテルという条件にぶれが生じるためという理由が大きい。なぜなら地域によってマイナーなホテルが存在するレビュー件数のボーダーラインが違うためである。都道府県という大まかなくりでよいのかという点には課題が残されている。

レビュー件数というのは選択した都道府県のホテルごとのレビュー数が棒グラフで表示されたもので、横軸にレビュー件数が表示され、縦軸にホテル数が表示される。こ

のヒストグラムの x 軸の最大数値は、地域ごとのレビュー件数によって決まるため選択した都道府県によって規模感は異なる。このグラフでレビュー数を選択することでユーザはホテルのある程度の規模やホテルがどのくらい人気かどうかを選択することができる。

レビュー件数が少ないものには多くの場合、小規模なホテルや個人経営の部屋数の少ないペンション、または新設間もない宿泊施設が含まれる。逆にレビュー件数の圧倒的に多いものには、全国的に人気な大規模ホテルなどが多い。このレビュー件数を選択できることがこのツールの大きな特徴で魅力である。一般的にレビューサイトなどではレビュー件数で検索はできないからだ。また宿泊施設は、都道府県ごとにレビュー件数の分布や件数の傾向が全く違ったりするが、情報を可視化することによってどの都道府県ごとのホテルのレビュー分布を瞬時に把握することができる。効率的な探索に有効なツールであると言える。

3.4 評価項目のスコアによる絞り込み

次にユーザには重要視する項目と評価の値を選択してもらう (図 1(b))。重要視する項目とはレビューデータの評価項目として存在する、「立地評価」「部屋評価」「食事評価」「風呂評価」「サービス評価」「設備評価」「総合評価」の 7 項目のことである。この項目はレビュー投稿時に集計されるデータだが、レビューサイト上の宿泊施設一覧では確認できない数値である。ホテル一つ一つページを開いて初めて確認できる。デフォルトでは総合評価が設定されているため、この項目はユーザが設定しなくてもよい。

都道府県とレビュー件数を選択した時点でレビュー評価のグラフが表示される。ここでいうグラフとはレビュー件数のグラフとは別のグラフである。x 軸に評価の 1 から 5 の数字、縦軸にレビュー件数が表示されたデータで、このグラフを選択することで評価値を選択できる。都道府県選択で北海道を選択し、ヒストグラムでレビュー数 20 から 40 のバーを選択し、重要視する項目で立地評価を選択したなら北海道でレビュー件数が 20 から 40 件のホテルの立地評価の分布が表示される。多くの場合ここでは最高評価の 5 を選択すると思うが、4 や 1 も選択できるようにすることで効率よくホテルを探すというだけでなくツールを使うおもしろみが出ると思う。これによって探索に自由度が生まれ、インタラクティブ性がまずだろう。

3.5 選択結果の表示

すべての項目を選択するとホテル名が表示される (図 1(c))。最終的に表示されるデータには「施設 ID」「立地評価」「部屋評価」「食事評価」「風呂評価」「サービス評価」「設備評価」「総合評価」「都道府県」「市区町村」「施設名」「レビュー本文」が含まれる。レビュー本文も含めているためここで表示されるのはホテル名の一覧ではなく、選択

された宿泊施設のレビュー一覧である。

3.6 実装方法

開発言語は Python で、Web アプリケーションの作成には Python ライブラリの Dash*1 を使用している。可視化作成時に plotly.express*2 も使用した。

4. 探索結果の傾向と課題

図 2 は探索結果の一例として、穴場ホテルの探索を行ったものである。都道府県を島根県に指定し、レビュー件数を 0 から 9 件に指定、評価項目を食事に指定したあと評価の数値を 5 に指定した。結果には島根県でレビュー件数の少ない食事の評価が高い宿が表示されている。

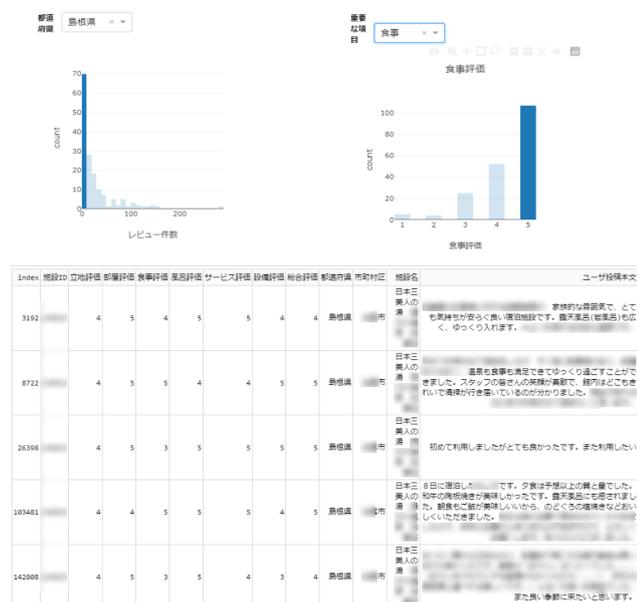


図 2 島根県で探索した例

このホテル探索ツールでレビューやホテルを探索しているといくつか特徴が見えてきた。たとえば、北海道の食事評価 5 と東京都の食事評価 5 はまったく意味が違うと感じたこと。北海道の場合の食事評価 5 というのは土地の傾向からか宿泊施設も観光に寄ったものが多いのか観光客が本当に食事がおいしかった場合につけられていることが多い印象だが、東京都の場合は同じ検索方法でもそうではない傾向にある。図 3 は東京都と北海道のレビューの傾向の例である。どちらも同じ条件で比較しているが傾向が違うことが分かる。

東京都の宿泊施設が食事におもきを置いていない傾向にあるのか、朝御飯が無料で出てそこそこ満足したというレビュー本文の内容で食事評価 5 になっている事例も見受けられた。見方を変えると、東京都に食事が特においしい宿

Index	施設ID	立地評価	設備評価	食事評価	風呂評価	サービス評価	設備評価	総合評価	都道府県	市町村	施設名	ユーザ投稿本文
126948		5	5	5	3	5	5	4	東京都	港区	ホテル	朝ごはんが無料でした。朝食・夕食も食べたいですが、朝のチェックインも無料です。
363456		4	4	5	5	5	5	5	北海道	札幌市	旅館	食事が本当においしくて、また食べたいくらいです。浴衣と朝飯も準備があります。浴衣は無料です。朝食は無料です。朝食は無料です。朝食は無料です。

図 3 上が東京の例、下が北海道の例

泊施設が存在しないわけないにもかかわらずそういった印象を受けたということは、食事評価の数値だけでは望む宿を瞬時に見つけることはできないとも言える。東京都に関しては宿泊施設があまりにも多いため、データも膨大になるのでレビュー件数、評価値だけではまったく宿泊施設を絞り切れていない。完全に絞り切れることを目的にしているわけではないが、5 分かかっても見切れないほど結果が出てしまうのは本意ではない。その点を改善できるのが宿一覧を出す機能の追加だと考える。

現状では結果に絞り込んだ宿泊施設で投稿されている一件一件のレビューの情報が大量に羅列してしまっているため、レビューの本文を見ることは出来るが宿泊施設を比較することは簡単ではない。しかし単に宿泊施設を羅列したところで見ごたえはない。そこで求められるのが、宿泊施設ごとの URL の追加や画像の追加である。実装を検討したい。

そして結果に選択した評価以外のレビューも表示されるという課題もある。例えば食事評価 5 の絞り込みをした際にも食事評価 3 のレビューは表示される。それはホテルをレビューで絞り込むのではなくあくまでホテルの評価で絞り込んでいるためだが、今後絞り込んだ条件に当てはまるレビューのみ表示する機能の追加も検討したい。

また全体に言えることだが、条件の複数選択ができないというのも課題である。たとえば、栃木県と群馬県のレビュー件数 10 から 100 の間の食事と立地で評価 5 を取っている宿泊施設、などという検索は現状不可能である。条件が一つの場合探索方法によっては結果のデータ数が見切れないほど多くなることもあるが複数追加することができればより細かな指定ができ、結果も数が絞られることによって見やすいものになるだろう。

5. おわりに

本研究では大手レビューサイト「楽天トラベル」のレビューデータを使用し、探索項目としてレビュー件数の追加や評価項目、評価項目の数値を選択できるようにした宿泊施設の探索ツールを作成した。結果レビュー件数やレビューの評価使った新しいホテル探索ツールが完成した。

課題として、探索地域によって同じ食事評価でも結果として出力されるホテルの傾向が異なる問題や、探索条件の複数選択ができないことで結果のデータが細かく絞り切れないなど技術的な問題も見つかった。

*1 <https://dash.plotly.com/>

*2 <https://plotly.com/python/plotly-express/>

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」*3を利用した。

参考文献

- [1] tabikobo: 最新！旅行意識調査 (2018). <https://about.tabikobo.com/news/press/2018/06/180605/>.
- [2] 吉田伊武貴, 倉田陽平: 投稿者による評価スコアや感情極性辞書を用いない, 文意に基づく宿泊施設不満 レビュー自動抽出手法の構築及びその分析による不満予防策の検討, *DEIM2022* (2022).
- [3] 野本 輝, 上野 史, 太田 学: 観光レビュー文を用いた穴場スポットの発見, *DEIM2022* (2022).

*3 https://rit.rakuten.com/data_release/