

# ホテル特化型評価表現辞書に基づくレビュー自動スコアリング方式

酒井 聖† 松波 友稀†† 中島 伸介†

† 京都産業大学コンピュータ理工学部 〒603-8047 京都府京都市 北区上賀茂本山

†† 京都産業大学大学院先端情報学研究科 〒603-8047 京都府京都市 北区上賀茂本山

E-mail: †{g1544629,nakajima}@cc.kyoto-su.ac.jp, ††i1788223@cse.kyoto-su.ac.jp

あらまし 近年、ホテルなどの宿泊施設を予約するサイトが数多く存在しており、ユーザが宿泊するホテルを予約サイトから調べ、予約するということが盛んに行われている。予約サイトでは実際に利用したことのあるユーザによるレビュー文と各項目に対するスコアが閲覧可能であり、このようなユーザの評価が書かれたレビューは、まだ宿泊したことのないユーザがホテルの利用を検討する際に大いに参考になるため、存在意義は非常に大きいと言える。しかし、同じ項目に対する評価が同じスコアでも、レビュー文の内容が異なっている場合がある。本研究では、ホテル予約サイトに投稿されたレビュー文を分析し、ホテルに関する評価項目別の評価表現辞書を構築し、構築した評価表現辞書を元に詳細化した項目別のホテルレビュー自動スコアリングを行う手法を提案する。

キーワード レビュー・評判情報・レビュー分析

## 1. はじめに

近年、楽天トラベル [1] やじゃらん net [2] といった利用者のレビューや各項目に対するスコアが掲載されたホテル予約サイトが数多く運営されている。ユーザは、宿泊するホテルを選ぶ際、各項目に対するスコアが高いホテルを選ぶ行動をとっている。ホテル予約サイトでのレビュー投稿時には、あらかじめ用意された項目に対してスコアを付与することができる。しかしこのスコアは、あらかじめ用意されている項目に対しての評価であり、同じ項目の評価でもユーザによって評価のニュアンスが異なっている場合がある。例えば、「部屋」に対する評価が同じ5点という評価でもレビューの内容が部屋からの眺めについてや、広さについてなど異なる場合がある。そのため、自分の求めている項目の評価が本当に高いのか、膨大なレビューを読んで判断しなければならない。

そこで我々は、ホテルに対するレビューを分析し、項目を詳細化した上で項目別にスコアを付与する手法の提案を行う。これを実現することで、将来的にはユーザが重視したい項目の評価が高いホテルの推薦が可能なシステムの構築を目指す。具体的には、ホテル予約サイトに投稿されているレビューを分析し、キーワード、特徴、程度などホテルの評価に関する評価表現辞書を項目別に作成し、この辞書に基づいて図1のようにホテルの評価項目ごとに詳細なスコアを付与する。これを実現することで、ユーザが着目する項目の評価が高いホテルの推薦が可能になると考えている。

以下、2章では関連研究との比較を述べ、3章では、ホテル特化型評価表現辞書に基づくレビュー自動スコアリング方式について述べる。4章では自動スコアリング手法の妥当性の確認実験、また作成した評価表現辞書の精度確認実験について述べ、最後に5章でまとめと今後の課題を述べる。

## ホテルAの「部屋」の評価



図1 ホテルの「部屋」に対する自動スコアリング結果の例

## 2. 関連研究

有名なホテル予約サイトに楽天トラベルがある。楽天トラベルでは、ホテルや旅館など多くの宿泊施設が掲載されており、宿泊する地域を元にホテルの検討が容易にできる。ユーザは、レビュー投稿時に総合評価の他、立地、部屋、食事、風呂、サービス、設備・アメニティの項目に5段階で評価する。そのため、ユーザからの主観的な評価が数字で分かるようになっている。

楽天トラベルのようなレビューを取り扱うサイトが広く普及したことで、レビューを分析する研究が多く行われている。

松波らは、コスメアイテムに対して、評価表現辞書を作成し自動スコアリングを行う方式を開発している [3]。辞書の作成については参考になるが、対象物が異なるため、本研究ではホテルの特徴について特化した辞書を作成する。

浅野らは、飲食店のレビューをメニュー単品、メニュー全体、雰囲気 の3つの分類に分け、ユーザが目的のレビューを取得しやすくする手法を提案している。レビューの文にラベル付けを行い、その伝播度によってレビューを分類している [4]。

櫛見らは、映画のレビューから映画の特徴ベクトルを抽出し、ユーザの嗜好に合わせた映画を推薦する手法を提案している。特徴ベクトルごとにコサイン類似度を求め、類似度が高い映画

をユーザに推薦している [5].

本塚らは、レビューから素性を抽出し、観点ごとの評価値を正確に推定する手法を提案している [6].

ZHANG Bo らは、楽天トラベルのレビューに対して、評価視点毎にユーザの評価点を推測する手法を提案している [7].

南らは、対象ユーザが既知ユーザに対する信頼度に基づいて、未知ユーザにに対する信頼度を予測する手法を提案している。レビューを元に信頼情報を取得し、その情報を利用して書籍推薦を行なっている [8].

上述の通り、ユーザによって投稿されたレビューを分析し、商品推薦を行う研究が数多く行われているが、本研究にて実施しているホテルに対する詳細化された項目別のスコアリング方式に関する研究開発は行われていない。

### 3. ホテル特化型評価表現辞書に基づくレビュー自動スコアリング方式

本章では、ホテル特化型評価表現辞書に基づくレビュー自動スコアリング方式に関して説明する。評価項目別の評価表現辞書の作成 (3.1 節) および自動スコアリング方式 (3.2 節) について説明する。

#### 3.1 ホテル特化型評価表現辞書の作成

本節では、ホテル特化型評価表現辞書の作成方法について述べる。

ユーザは様々な観点でホテルに対するレビューを投稿している。そこで、実際に使用されている表現を辞書に登録することで項目別の評価表現辞書を構築した。辞書の構築に利用するホテルレビューデータはホテル予約サイトである楽天トラベルのレビューデータを採用した。

具体的な辞書の構築は、レビューの情報を分析し、項目ごとにキーワードを分類し辞書に登録する。出現したキーワードを元のレビューから調べ、その前後を確認し、どのような特徴、程度で評価しているかを確認する。この特徴と程度も辞書に登録し、辞書全体の表現を把握した上でスコアを決定する。なお本稿では、ホテルの形態を絞るため施設名に「温泉」が入っているホテルのレビューを対象に辞書を作成した。

キーワード	特徴	程度	スコア
コンビニ	便利	多くて	★★★★★
コンビニ	ある	近くに	★★★★★
駅	近い		★★★★★
駅	離れている	少し	★★★

図2 「立地」の評価表現辞書の一部

表1に「立地」を詳細化した評価項目に当てはまる評価表現の一部を示す。「立地」の評価表現として、コンビニや飲食店といった店や、観光名所などの施設、駅からの距離について述べたものが見られた。

表2に「食事」を詳細化した評価項目に当てはまる評価表現

表1 「立地」の評価表現の例

従来の評価項目	提案する評価項目	見られる評価表現
立地	コンビニ・スーパーへのアクセス	コンビニが近くにある 近隣にコンビニがない 見える範囲にコンビニがなかった コンビニが多くて便利
	周辺の飲食店	近場に飲食店 飲食店が豊富 飲食店が多くて便利
	観光スポットへのアクセス	劇場が目の前 温泉の目の前 観光名所が近い 散策に向かない
	ホテルまでのアクセス	駅に近い 駅から徒歩で少々 駅からやや遠い 駅から少し離れている 駅から結構距離がある 歩きだと微妙な位置

表2 「食事」の評価表現の例

従来の評価項目	提案する評価項目	見られる評価表現
食事	豪華さ	饗されると思っていたのに期待外れ 地元関係のものは秀逸 工夫された料理に大満足 食事が驚くほど粗末 どこにでもあるものはいまいち しょぼい料理
	品数	バイキングで種類少なめ 種類も豊富 品数があまりにも少なく
	量	食べきれないほど 朝食バイキングが物足りない ボリュームがある
	味	とてもおいしかった おいしかった 味付けは可もなく不可もなく

の一部を示す。「食事」の評価表現では、料理の品数、量、味の他、料理の内容についての評価も見られたので、「豪華さ」という項目として分類する。

表3に「部屋」を詳細化した評価項目に当てはまる評価表現の一部を示す。「部屋」の評価表現では、部屋からの眺め、部屋の広さ、清潔さや部屋の周辺の音について述べたものが見られた。

表4に「風呂」を詳細化した評価項目に当てはまる評価表現の一部を示す。「風呂」の評価表現では、温泉や露天からの眺めについて述べたものが見られた。

表5に「サービス」を詳細化した評価項目に当てはまる評価表現の一部を示す。「サービス」の評価表現では、スタッフの対応や送迎について述べたものが見られた。

表6に「設備・アメニティ」を詳細化した評価項目に当てはまる評価表現の一部を示す。「設備・アメニティ」の評価表現で

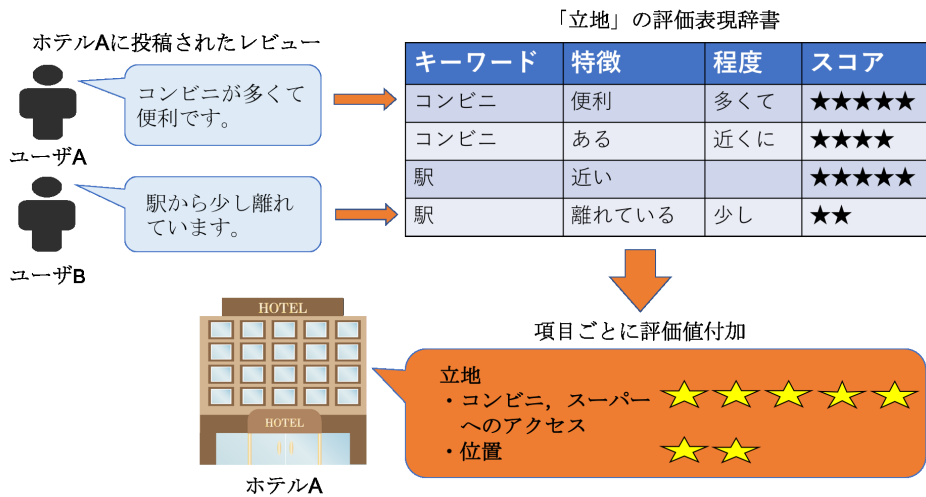


図 3 辞書を用いた項目別スコアリングの例

表 3 「部屋」の評価表現の例

従来の評価項目	提案する評価項目	見られる評価表現
部屋	眺め	眺めがいい部屋 部屋から広がる絶景 窓からは景色がばっちり
	広さ	部屋は広い 部屋は狭い 広々とした部屋 部屋が狭く身動きが取れない
	清潔さ	部屋の臭いが気になる 部屋は清潔です 部屋の掃除が行き届いていない
	静かさ	部屋はとても静か 隣の部屋の音も全く気にせず 廊下の音が若干聞こえる 電車の音が気になる

表 4 「風呂」の評価表現の例

従来の評価項目	提案する評価項目	見られる評価表現
風呂	温泉	泉質もいい お肌スベスベ とてもいい泉質 お風呂はかけ流し
	眺め	お風呂からは海が見えて眺めは最高 お風呂は海に面した絶景 露天とかはない

表 5 「サービス」の評価表現の例

従来の評価項目	提案する評価項目	見られる評価表現
サービス	スタッフ	夕食の対応が非常に悪かった スタッフの方も親切丁寧に対応 スタッフの皆さんの笑顔がよかった スタッフの態度がイマイチ
	送迎	送迎してくれる 送迎もあり大変満足 送迎バスがある

表 6 「設備・アメニティ」の評価表現の例

従来の評価項目	提案する評価項目	見られる評価表現
設備・アメニティ	アメニティ	洗顔フォームがなかった ヘアブラシなどない アメニティも歯ブラシだけ 一通り必要なものは揃っている 子供向けのアメニティが充実
	ネット環境	回線が遅い Wifi が途切れ途切れ ネットワーク環境も良好
	機器	マッサージチェアも無料で使用 紅茶の無料サーバー 洗濯機が無料

は、アメニティの有無やホテルのネット環境、備え付けられている機器について述べたものが見られた。

これらの評価表現を元にスコアを設定し、辞書に追加する。「立地」の評価表現辞書の例を図 2 に示す。

### 3.2 評価表現辞書に基づく自動スコアリング方式

本節では、未知のレビューデータに対する自動スコアリングを行う手順を述べる。

まず、あるホテルのレビュー全てを取得する。項目ごとにスコアを付与するために、レビュー文に対して形態素解析を行う。辞書に登録したキーワードが出現したなら、その前後を調べ、

特徴、程度を探す。辞書に当てはまるものがあれば、そのスコアを項目に付与する。

最後に項目ごとに平均値をとることで、そのホテルの平均スコアとして決定する。このようにスコアを統合する。

実際にスコアリングを行う過程を図 3 に示す。ユーザー A は「コンビニの数」について述べ、ユーザー B は「駅からの距離」について述べている。辞書に登録しているものがあるので、そのスコアをホテル A の評価項目「近くの店」、「位置」に付与する。

## 4. 評価実験

### 4.1 実験手順

評価表現辞書のスコアリングの妥当性を検証するために、実験を行なった。被験者である20代の男女5名に、実際に宿泊したホテルについてのレビュー文と詳細化した項目ごとにスコアを付けてもらった。被験者が付けたスコアを正解データとする。次に、被験者が記述したレビューの内容を当てはまる項目ごとに分類し、評価表現辞書からスコアを付ける。このスコアと正解データを比較する。今回用いる評価表現辞書は、228件のレビューを元に作成した。

### 4.2 実験結果

今回の実験では、全ての項目についての評価文をレビュー文から得ることができなかった。評価文が書かれた評価項目と正解データ、評価表現辞書からのスコアを表7に示す。辞書に表現を登録しておらず、スコアを付けられなかった項目には「-(ハイフン)」を付けている。

表7 実験結果

評価項目		正解データ	辞書からのスコア
立地	コンビニ・スーパーへのアクセス	3	4
	周辺の飲食店	1	-
	観光スポットへのアクセス	3	-
	ホテルまでのアクセス	3	-
		2	1
食事	味	4	2
		3	4
部屋	清潔さ	5	4
		5	4
		4	4
風呂	温泉	5	5
		3	-
サービス	送迎	4	-
設備・アメニティ	アメニティ	3	-

表7より、スコアを比較してみると、正解データと辞書からのスコアの平均絶対誤差は約0.89と抑えることができた。そのため、辞書からのスコアリングは妥当であると考えられる。さらに誤差を小さくするためにも辞書に登録するスコアの検討が必要である。

また、「ホテルまでのアクセス」に正解データが4点で辞書からのスコアが2点となったレビューがあった。これは、駅から歩いて少し遠く感じたというレビュー文を見て2点と付けたが、続いて送迎バスを利用すれば大丈夫という文があったため、送迎の評価を「ホテルまでのアクセス」に4点と付けたと思われる。

評価表現辞書に特徴や程度を登録していなかった表現を表8に示す。今回の実験で用いた評価表現辞書では、登録した表現が少なかった。そのため、さらに多くのレビューから表現を登録する必要がある。また、登録していない表現を減らすためには、「ある」という表現を登録した際に「ない」という表現も登録することで登録する表現を増やすことができると考えた。

表8 特徴や程度を登録していなかった評価表現

評価項目		見られた評価表現
立地	周辺の飲食店	周りに飲食店がなかった
	観光スポットへのアクセス	観光地からは少し離れている
	ホテルまでのアクセス	バスがなくて駅から歩いた
部屋	清潔さ	部屋はまあまあ綺麗でした ホテルの部屋はきれいで心地よかった
風呂	温泉	温泉も行ったのですが普通の印象でした
サービス	送迎	シャトルバスに間に合えば便利
設備・アメニティ	アメニティ	アメニティも悪くはなかった

## 5. おわりに

本稿では、個々のユーザが重視している項目の評価が高いホテルを推薦するために、ホテル予約サイトのレビューデータに対して、詳細化した評価項目別スコアを付与する自動スコアリング方式を提案した。提案手法では、従来の評価項目を更に詳細化し、ホテルに関する実際のレビューから抽出した評価表現を利用したキーワード、特徴、程度からなる評価表現辞書を作成した。構築したホテル特化型評価表現辞書を用いたスコアリングによって、ユーザが着目する項目の評価が高いホテルの推薦が可能になると考えている。

今後は、辞書に登録するスコアの再検討や詳細化評価項目の妥当性の検証を行い、評価表現辞書の自動構築、表現の自動追加登録の手法の検討を行う必要がある。

評価表現辞書の拡張も引き続き行い、1つだけでなく2つ以上の項目が良いホテルの推薦を行うといった、自動スコアリング方式に基づくホテル推薦システムの設計についても行う予定である。

## 謝 辞

本研究を遂行するにあたり、楽天データ公開において提供された楽天トラベルのレビューデータを分析対象データとして使用した。ここに記して謝意を表します。

## 文 献

- [1] 楽天トラベル. <https://travel.rakuten.co.jp>
- [2] じゃらん net. <https://www.jalan.net>
- [3] 松波友稀, 上田真由美, 中島伸介. コスメアイテムに対する評価項目別レビュー自動スコアリング方式の開発, DEIM Forum 2017, B5-3, 2017.
- [4] 浅野祥汰, 北山大輔. レビューサイトの差異に基づくラベル伝播を用いたユーザレビュー分類手法, DEIM Forum 2017, D4-3, 2017.
- [5] 梶見圭司, 北山大輔. ユーザレビューを用いた全体的・部分的観点の類似に基づく映画推薦, DEIM Forum 2017, B5-4, 2017.
- [6] 大塚達也, 青野雅樹. 単語分散表現と文法的な表現に着目したカスタマーレビューの観点ごとの評価値推定, DEIM Forum 2017, F8-2, 2017.
- [7] 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回年次大会発表論文集, 2017年3月.
- [8] 南大智, 牛尾剛聡. このユーザは信頼に値するか?- SNS における協調的な信頼推定モデル-, DEIM Forum 2017, B1-4, 2017.