

レビュー自動スコアリング方式に基づくホテル推薦システム

安部 克[†] 中島 伸介^{††}

[†] 京都産業大学コンピュータ理工学部 〒603-8555 京都府京都市 北区上賀茂本山

^{††} 京都産業大学情報理工学部 〒603-8555 京都府京都市 北区上賀茂本山

E-mail: [†]1644042@cc.kyoto-su.ac.jp, ^{††}nakajima@cc.kyoto-su.ac.jp

あらまし ホテル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と星の数で表示された各項目に対するスコアが閲覧可能であるが、多くのホテル予約サイトでは、レビューへのスコア付けが可能な評価項目数は5～6個程度であり、レビューにて表現された評価を適切に表現することはできていない。そこで我々は先行研究として、ホテルに関する評価表現辞書を構築することで評価項目別のホテルレビュー自動スコアリング方式を提案している。本稿では、このレビュー自動スコアリング方式に基づいたホテル推薦手法を提案する。

キーワード レビュー・評判情報・レビュー自動スコアリング, 情報推薦, ユーザ間類似度, ホテル推薦
検証する。

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択している。また、多くの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は部屋からの眺めについてや、広さについてなど評価している内容が異なる場合がある。そのため、自分の予約する予定のホテルが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み理解し判断する必要がある。

そこで我々は、先行研究 [3] でホテルに対するレビューを分析、評価項目を詳細化した上で評価項目別にスコアを付与し、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する提案、およびそれを用いてレビューのスコアリングを行なった。本研究では、評価表現辞書に基づき図1のようなスコアリングされた結果を用いることで、既存のホテル推薦より精度の高い推薦を行うことが可能ではないかと考えている。

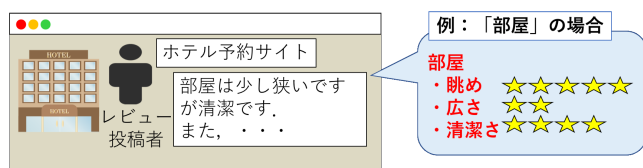


図1 「部屋」を詳細化した評価項目別の自動スコアリング結果の例

そこで本稿では、評価項目別自動スコアリングを用いた推薦手法の提案を行い、検証実験を行うことで提案手法の妥当性を

以下、2章では関連研究との比較を述べ、3章では、先行研究の内容である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章ではレビュー自動スコアリング方式に基づくホテル推薦手法について述べ、5章ではホテル推薦手法の妥当性の確認実験について述べる。最後に6章でまとめと今後の課題を述べる。

2. 関連研究

有名なホテル予約サイトの一つとして、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。評価はレビュー投稿者が主観的に感じた評価を数値化しているため、他のユーザも参考としやすい。

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 [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].

林部らは、宿レビューから「肯定的事実」と「推薦対象」という言語知識を、宿のレビューから抽出する手法を提案している。 [6]

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 [7].

田熊らは、ユーザが順位付けしたホテルの評価項目を元にホテルを推薦する手法を提案している。投稿されたレビューから、評価項目ごとの名詞の数をカウントし、数が多い順に評価項目

の順位付けを行い、ユーザ類似度が高いレビューを推薦している [8].

和田らは、宿泊予約サイトのレビューに記述されているギャップ文を抽出し、評価項目ごとに提示する手法を提案している [9].

野津らは、Twitter に投稿された TV 番組についての tweets の感情を 10 種類の感情軸に分類し、作成した 10 次元の感情ベクトルを作成。その後 TV 番組間の感情ベクトルのコサイン類似度を求めることで、TV 番組の推薦を行っている [10].

村瀬らは、マンガのストーリーに応じた漫画を、利用者の指定した作品に基にベクトル空間モデルを用いて類似度を判定し、推薦している。[11]

大山らは、ゲームレビューサイトから抽出したレビュー文を元にゲーム辞書を作成、Word2vec を用いた類似度計算によって入力データとゲーム間の類似度が高いゲームの抽出・提示をするシステムの提案を行っている [12].

Aditi A.Mavalankar らは、100 個のホテルのクラスタを用いて、その中でユーザが予約する可能性が高いホテルのクラスタ 5 個を予測し、推薦することを目指している。[13]

Nurul Akhmal binti Mohd Zulkefli らは、ブログに存在するホテルに関する有用な情報を元に、ユーザにブログランキングに基づいた手法で推薦している。[14]

上述の通り、ユーザによって投稿されたレビューや特徴ベクトル、ランキングなどを分析し、情報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評価項目別自動スコアリング方式によって得られた詳細化された評価項目別スコアを特徴ベクトルとして用い、類似ユーザを検索することでホテル推薦を行う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、先行研究である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。3.1 節では評価項目別の評価表現辞書の作成について述べ、3.2 節では自動スコアリング方式について説明する。

3.1 ホテル特化型評価表現辞書の作成

本節では、先行研究で行われた、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、過去に投稿されたレビュー文に含まれている評価表現をキーワード、特徴、程度に分割し評価項目ごとに分類して辞書に登録する。

表 1 に従来の評価項目と先行研究で提案した詳細化された評価項目の一覧を示す。

これらの詳細化された評価項目に対する評価表現に対してスコアを設定し、評価表現辞書に登録する。評価表現に対してのスコア設定は、20 代の男女 5 名にアンケートを行い 0 から 5 の

キーワード	特徴	程度	スコア	提案する評価項目
部屋	眺め	良い	★★★★★	部屋/眺め
部屋		広い	★★★★★	部屋/広さ
部屋		狭い	★★	部屋/広さ
部屋		綺麗	★★★★★	部屋/清潔さ

図 2 「部屋」の評価表現辞書の一部

表 1 従来の評価項目と提案する詳細化した評価項目

従来の評価項目	詳細化した評価項目
立地	コンビニ・スーパーへのアクセス 周辺の飲食店 観光スポットへのアクセス ホテルまでのアクセス
食事	豪華さ 品数 量 味
部屋	眺め 広さ 清潔さ 静かさ
風呂	温泉 眺め
サービス	スタッフ 送迎
設備・アメニティ	アメニティ ネット環境 機器

範囲でスコアをつけてもらったものの平均と、楽天トラベルのデータを元に設定している。「部屋」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、先行研究で行われた、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。詳細化した評価項目ごとの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エンジン mecab を用いて形態素解析を行う。分割されたキーワード、特徴、程度の組み合わせを評価表現辞書から探索を行い、辞書に当てはまる評価表現があれば、設定されているスコアをその評価項目に付与する。

実際にスコアリングを行う過程を図 3 に示す。このレビューには、部屋の眺めに対する評価と部屋の広さに対する評価が書かれている。レビューを解析し、評価表現辞書に当てはまる表現があるので、設定されているスコアを部屋の「眺め」と部屋の「広さ」に付与している。

4. 詳細化された評価項目別スコアを用いたホテル推薦手法

本章では、自動スコアリングを用いて算出された詳細化された評価項目別スコアを用いた推薦手法に関して説明する。図 4 に被推薦ユーザにホテルを推薦するまでの流れを示す。4.1 節では、ホテルレビューデータベースについて述べ、4.2 節では、

1. レビューに含まれている評価表現を辞書から探索

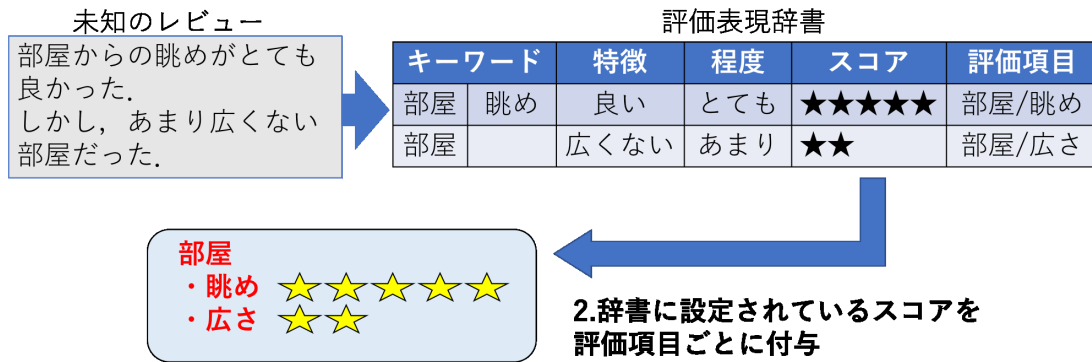


図 3 評価表現辞書を用いた項目別スコアリングの例

ユーザモデルの作成について述べる。4.3 節では類似評価ユーザの検出について述べ、4.4 節では、類似評価ユーザを用いた推薦に関して説明する。

4.1 ホテルレビューデータベース

本節では、過去のレビューデータを保存する、ホテルレビューデータベースについて述べる。

ホテルレビューデータベースには、過去にレビューを行ったユーザのデータが、図 5 のように格納されている。このデータベース内にあるデータを用いて、推薦を行う。このデータベースの管理にはデータベース管理システム MySQL を使用しており、各カラムの内容はユーザ判別を行うための”username”，ホテル名の”hotelname”，レビュー文が保存されている”review-text”，それ以外の”location”，”room”，”food”，”bath”，”service”，”facility”，”total”は各項目に対する評価という形になっている。このデータベースを作成するサンプルデータとして、楽天トラベルのレビューデータ約 250 万件を採用した。

4.2 ユーザモデルの作成

本節では、自動スコアリング方式によって算出された、詳細化された評価項目別スコアを用いてユーザモデルを作成する手法について述べる。

まず、被推薦ユーザがあるホテルに対してレビューを投稿する。投稿したレビューを 3.2 節にて述べた自動スコアリング方式を用いて詳細化された評価項目別スコアを算出する。その後、算出されたスコアデータを元に、図 7 のようなユーザプロフィールデータを作成する。その後、被推薦ユーザが過去にレビューを行っている場合は、過去に行ったユーザのレビューデータと作成したユーザプロフィールデータを元に、そのユーザの評価などのレビューデータをまとめたユーザモデルの作成を行う。

4.3 類似評価ユーザの検出

本節では、4.2 節にて述べたユーザモデルを用いてホテルレビューデータベースに問い合わせ、類似評価ユーザを検出する手法について述べる。

作成したユーザモデルには、過去に被推薦者であるユーザがレビューを行ったホテルの評価、ホテル名、レビュー文などが格納されている。それらを用いて類似評価ユーザの検出を行う。

類似評価ユーザを検出する際、4.1 節にて述べたホテルレビューデータベースに問い合わせを行う。問い合わせの内容は、被推薦ユーザが今までにレビューを行ったホテルに対してレビューを付けた他のユーザの中で、最も評価が類似しているユーザを検出することである。類似ユーザを検出する際には、図 6 に示したコサイン類似度を用いる。これにより、被験者と思いや趣味が似たユーザを検出することができ、推薦の精度を上げることができると考える。

4.4 類似評価ユーザを用いたホテル推薦

本節では、4.3 節にて検出した類似ユーザを用いて被推薦者にホテルを推薦する手法について述べる。

まず、検出した最類似ユーザの全てのデータをホテルレビューデータベースからホテル推薦システムに提供を行う。提供された最類似ユーザのデータのうち、最も高評価をつけたホテルを検出、その後検出したホテルを被推薦者に提示することによりホテル推薦を行う。

5. ホテル推薦手法の妥当性の確認実験

5.1 実験手順

4 章章で示した手法の妥当性を検証するために、実験を行った。まず、被験者である 20 代の男女 8 名に対して過去に宿泊したことのあるホテルについてのレビューを書いてもらった。書いてもらったレビューから自動スコアリング方式を用い詳細化された評価項目別スコアを付ける。このスコアを用いて、約 88 万人のユーザのレビューデータが存在するデータベースから最類似ユーザをコサイン類似度を用いて検索し、ホテル推薦を行う。今回は被験者が過去に他のホテルに対してのレビューを書いたデータが存在しないため、今回レビューを行ったホテル一種に対してレビューを付けたユーザの中から最類似ユーザを検索する。その後、推薦結果として出されたホテルの公式サイトを被験者に提示し、食事、温泉等の項目確認を行ってもらい、推薦したホテルに宿泊してみたいかを調査した。

5.2 実験結果

今回の実験では、被験者 8 名中 7 名が推薦したホテルに宿泊したいと回答した。コサイン類似度のグラフを図 8 に示す。

図 8 に示した A~H の被験者のうち、被験者 E 以外が推薦

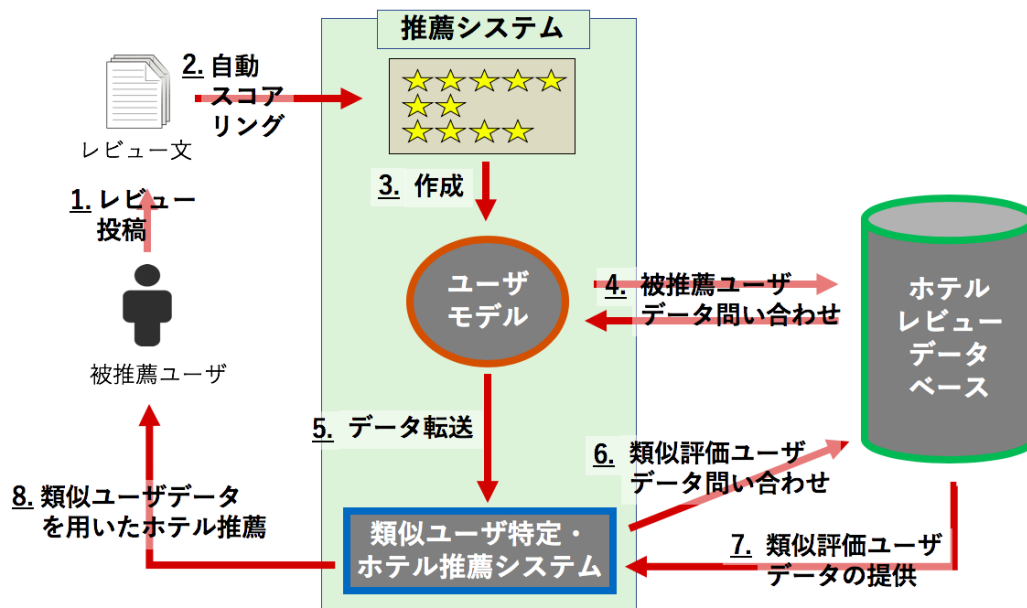


図 4 推薦の流れ

username	hotelname	location	room	food	bath	service	facility	total	reviewtext
...	...	4	3	0	3	3	3	4	とても清潔な部屋でした。周辺は箱崎~千代にかけ...
...	...	5	5	4	5	5	5	4	貸し切り風呂がとても良かったです！色んな風呂...
...	...	5	5	0	3	5	5	5	ダブルでの利用は初めてでした。相変わらず綺麗...
...	...	4	4	0	3	4	4	4	フロントの方の対応も感じよく、部屋も清潔でした。
...	...	2	5	4	4	3	4	4	部屋は広く、ペランダがあったのはとてもよかつた...
...	...	5	5	2	5	5	3	5	ホテルの外観も内装もとても素敵でした。部屋は広...
...	...	3	4	2	4	3	3	4	とにかく安いプランがあって部屋は広くて安くて良...
...	...	5	3	3	4	3	2	3	ビジネスホテルのユニットバスではくつるがないの...
...	...	5	5	5	4	5	4	4	この度は両親子供と4人でお世話になりました。部...
...	...	5	5	4	3	4	4	5	駅から近くとても便利で良かったです。部屋もきれ...
...	...	4	4	3	3	4	4	4	いつも利用させて頂いています。今回はいつもと違...
...	...	5	5	5	4	4	5	5	深夜に予約・チェックインしましたが、対応が素早...
...	...	4	4	4	3	5	2	4	夜遅くのチェックインだったのにも関わらず、嫌な...
...	...	4	4	5	4	4	4	4	久しぶりの利用でした。朝食が以前と比べて、凄...
...	...	5	4	0	3	4	3	4	二度目の宿泊です。施設は隅々まで綺麗にされて...
...	...	4	4	5	4	5	4	5	多少、古さを感じましたが、清潔感もあり、芦ノ湖...
...	...	5	4	4	5	4	3	5	熊本出張でいつも利用しているの、
...	...	1	1	0	1	1	1	1	二度と泊りません。【ご利用の宿泊プラン】【室数...
...	...	5	5	5	5	5	5	5	社員が宿泊させて頂きましたが、感想として、と...
...	...	2	3	0	2	3	3	3	値段相応の施設でした。単に寝て泊まる方には申し...
...	...	4	4	0	3	5	4	5	服务人员真的非常親切。交流让人愉快。旅館就在...
...	...	3	3	0	4	4	4	4	普通の都市圏でこの値段は十分過ぎるかと思いま...
...	...	4	4	3	3	5	5	4	同僚と泊まりましたが、満足してました。カウンタ...
...	...	4	5	5	4	4	4	5	今回初めて利用させて頂きました。他宿泊施設の一...

図 5 データベース内データの一例

$$\frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}}$$

n: レビューしたユーザと比較ユーザの両者が評価をつけた項目の数
a: レビューしたユーザの評価値
b: 比較ユーザの評価値

図 6 コサイン類似度

ユーザ名	ホテル名	評価項目	スコア	
ユーザ1	ホテルX	立地	コンビニ・スーパーへのアクセス	4
			ホテルまでのアクセス	1
		食事	味	3
			部屋	広さ
			清潔さ	5
		サービス	送迎	4
		設備・アメニティ	アメニティ	3
機器	5			

図 7 ユーザプロフィールの一例

したホテルに宿泊したいと回答した。被験者 E はこの検証の中で最もコサイン類似度が低い最類似ユーザを用いた推薦であったため、満足のできるホテル推薦ができなかったと考えられる。この結果から、コサイン類似度の値が高いユーザを用いた推薦であるほど、被験者へのホテル推薦の精度が向上し、この二つには相関性があると考えられる。しかし、本検証は 8 人という少人数で行われたものであったため、被験者の人数が増加して

も同様に相関性が存在するかどうか、確認する必要があると考えるまた、結果全体より、自動スコアリング方式を用いたホテル推薦が有効な推薦手法であることを示していると考えられる。

6. まとめと今後の課題

本稿では、ユーザが重視している観点の評価が高いホテルを推薦するために、詳細化された評価項目にスコア付けをするこ

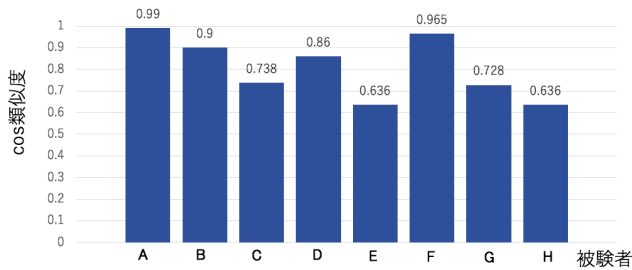


図 8 被験者と最類似ユーザとのコサイン類似度

とができる自動スコアリングシステムを用いて推薦する手法の提案をした。提案手法では、詳細化された評価項目別スコアを用いることにより従来よりも精度の高いホテル推薦システムの開発が可能ではないかと述べた。その後提案手法の検証を行い、手法の有効性が示され、コサイン類似度と推薦精度に相関性があることを確認した。今後の課題としては、被験者が 8 人という少人数であったため人数が増加しても同様に相関性が確認できるのか、考える必要がある。

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。また、本研究の一部は、JSPS 科研費 17H01822 による。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] 安部克, 酒井聖, 松波友稀, 中島伸介. レビュー評価項目別スコアを用いたホテル推薦手法の提案, DEIM Forum 2019, P1-134, 2019.
- [4] 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.
- [5] 榎見圭司, 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] 林部祐太. 宿レビューからの肯定的事実と推薦の対象の抽出, 言語処理学会 第 25 回年次大会, 2019. .
- [7] 南大智, 牛尼剛聡. 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.
- [8] 田熊浩二, 福原楓, 亀井清華, 藤田聡. 口コミを用いた嗜好抽出に基づくホテル推薦手法, 情報処理学会第 78 回全国大会, 2016.
- [9] 和田知亜紀, 豊田哲也, 大原剛三. 期待と現実のギャップに着目した宿泊施設レビューの解析とその検索への応用, 情報処理学会, 2018.
- [10] 野津貴之, 若原徹. 放送中に投稿された Tweet を用いた TV 番組の推薦, 情報処理学会第 80 回全国大会, 7K-04, 2018,3.
- [11] 村瀬尊好, 柊和佑, 安藤友晴. マンガの概要に基づく作品推薦システム, FIT2012 第 11 回情報科学技術フォーラム N-013, 2012.
- [12] 大山光暉, 竹山佳成, 平田圭二. レビュー分を考慮したゲーム推薦システムの実現に向けた単語類似度調整の取り組み. エンタテインメントコンピューティングシンポジウム (EC2017), 2017,9.
- [13] Aditi A.Mavalankar, Ajitesh Gupta, Chetan Gandotra, Rishabh Misra. Hotel Recommend System, Cornell University cs.LG, 2019.
- [14] Nurul Akhmal binti Mohd Zulkefli, Baharum bin Baharudin. Hotel Travel Recommendation based on Blog In-