

レビュー分析に基づく観光地周辺の宿泊施設推薦手法の提案

佐藤 海龍[†] 王 元元[†]

[†] 山口大学大学院創成科学研究科 〒755-8611 山口県宇部市常盤台 2-16-1

E-mail: †{i036vg, y.wang}@yamaguchi-u.ac.jp

あらまし 現在、旅行先の観光情報を集める手段として Web 上の観光サイトを利用することで、ユーザのレビューや評価などの情報を総合的に判断して計画することが多くなっている。しかし、これらは観光スポットもしくは宿泊施設のどちらかの情報に寄っているなどで双方の関係性までを考慮されていない。また、人気順や評価順に基づいた施設ランキングでは、施設に対する不満や観光ニーズなどを考慮されていない。そこで本研究では、効率的な旅行計画を支援するために、観光スポットと宿泊施設の関係性を分析することで、ユーザのニーズに合わせた宿泊施設推薦システムの構築を目指す。具体的には、観光スポットと宿泊施設の関係性に着目し、場所およびレビュー情報を用いて観光スポットと宿泊施設間の物理的な距離やレビュー内容の分析に基づき関係性を抽出することで、観光スポットに最適となる宿泊施設のランキングを推薦する手法を提案する。これにより、利用者は観光スポットを検索しながらその周辺の宿泊施設を検索することができる。本論文では、提案した宿泊施設推薦手法について述べ、ユーザの観光ニーズに合わせた宿泊施設推薦の実行例を示す。最後に、アンケート調査による宿泊施設推薦手法の有用性を図る評価実験を行った。

キーワード 観光支援, 観光スポット, 宿泊施設, 推薦システム

1 はじめに

2020年に東京オリンピック・パラリンピックが開催されることに続き、2025年には日本の誘致成功で20年ぶりに大阪で日本万国博覧会が開催することが決定した。これに伴いインバウンド観光客である訪日外国人数は2013年度の1,036万人に対して2015年度は1,973万人、2017年では2,869万人のような増加傾向である[1]。さらなる目標として訪日観光客4000万人が掲げられ、大きな経済効果が期待される。だが、課題も多く存在する。たとえば、観光の地域による偏りや観光の経済効果を充分地域に取り込めていない、観光安全に懸念が残るといったものがあげられる¹。これに対して、王と岳[2]は地方観光協会のTwitterアカウントを用いた地方都市の魅力を効率的に情報発信する地方都市の観光客誘致戦略を提案している。平久江ら[3]はマイクロブログのTwitterからジオタグ付きツイートを取得して座標からメジャー・マイナー判定を行った。観光ガイドブックや観光サイトに乗っていないようなマイナーなスポットを見つけることで観光客の地方の分散化を試みた。また、開地と檜垣[4]はユーザが今まで意識することがなかった観光地の推薦を行った。Word2Vecを用いて観光地の分散表現を作成し、興味となる単語の足し引きを行うことで潜在的な興味の発見方法を提案した。これにより、ユーザの嗜好を捉えつつユーザが思いつかないような観光地を推薦した。

同時に、近年では爆発的なスマートフォンなどの普及により、観光地の情報を調べることや宿泊施設の予約は多くがインターネットによるものとなっている。旅行を計画する際の情報収集

の手段でも、TripAdvisor²やYahoo!トラベル³などの観光情報サイトが時代の流れに伴い、大きな役割を担っているようになった[5]。これらの旅行サイトでは観光スポットや宿泊施設についての基本的な情報として写真やアクセスに加えて、実際に訪れた旅行者の感想を、ユーザレビューとして多くが投稿されて、共有することができる。さらには、そのままオンライン予約することやツアーの参加も可能となるため、旅行の一連を計画することができる。しかし、これらのサイトは観光スポットと宿泊施設の両方に着目した推薦ができない。多くは検索結果に対して、おすすめ順や評価が高い順といったようになっているため、施設に対する不満や観光ニーズなどを考慮されていない。また、たとえば宿泊が必要な旅行を計画しようとするとき、まず訪れたい先を決定し、観光したい場所を決めた後に宿泊の施設を決めるような流れで決めていくと考えられる。このとき、観光スポットに対しての最適な宿泊施設を決めるには、多くの宿泊施設の中からそれを、地図で見比べながらレビューで見比べながらと選ぶことになってしまい、多くの時間・労力が必要となる問題点が考えられる。

そこで、本研究ではこの問題に対して、観光スポットと宿泊施設の関係性が相互に与える影響を考慮して、観光スポットに対する適切な宿泊施設のランキングを推薦する手法を提案する。具体的には、効率的な旅行計画を支援するため、ユーザが検索した観光スポットに対して、投稿されたレビュー内容の分析や評価、観光スポットと宿泊施設間の物理的な距離などの複数の要素から最適となる宿泊施設をランキングしユーザに推薦する。これにより、利用者は観光スポットを検索しながらその周辺の

1 : <http://www.ryoko-net.co.jp/?p=36478>

2 : <https://www.tripadvisor.jp/>

3 : <https://travel.yahoo.co.jp/>

宿泊施設を検索することができる。

本論文の構成は以下のとおりである。2章では情報推薦，レビューを用いた推薦および検索に関連する研究の紹介を行う。3章では，提案する宿泊施設推薦手法について説明する。4章では，実データを用いた提案手法の評価実験について述べる。最後に，5章でまとめと今後の課題について述べる。

2 関連研究

近年，Web 情報を基にした観光地推薦や情報提供，およびその分析に関するさまざまな研究やシステムの開発が盛んである [6], [7]。中に，Web 上のデータから観光イメージを分析するもの [8] や，旅行者の感想を分析するもの [9]，Flickr や Twitter などの SNS から観光ルートの推薦を行っているもの [10], [11] などがあげられる。中嶋ら [12] は，観光地でつぶやかれたジオタグ付きツイートを分析し，旅行者の好みにあわせた観光ルートを推薦する方法を提案している。石野ら [13] はガイドブック等では知ることが難しい観光スポットの詳細な情報を効率良くユーザに提示するための情報支援システムを提案した。土田ら [14] はニューラルネットワークを用いた言語モデル Word2Vec を利用した。SNS から観光情報を抽出し，解析を行うことで生成したコーパスから単語の加算による他の地域における類似スポットの発見手法を提案した。また，三笠ら [15] はユーザが効率的に観光情報を集めるために，観光トピックごとに文章分類を行い，動的に概要文章を生成する手法を提案している。興味に応じた要約文章を見ることができ，不必要だと事前の判断が可能になるためサイトの閲覧時間の削減を行った。

同時に，ユーザレビューを用いた情報推薦も多く存在する。Sakai ら [16] は情報推薦を「耳より」となるキーワードを観光レビュー抽出し，それらを話題ごとに分類した。ユーザが観光地名を入力することで，システムで抽出された耳より情報を話題毎に分けて提示され，ユーザが選択することで知りたい情報をわかりやすく表示することで情報収集の効率化を図った手法を提案した。潘ら [17] の研究では観光情報サイトから投稿されたレビューのうち，スポット・季節・タイプに着目し，それぞれの特徴を解析してユーザに示した。これを選択させ，レビューの類似度を比較することで各ユーザの嗜好にあった観光スポットをランキング形式で推薦することができた。

3 提案システムの概要

本章では，選択された観光スポットと宿泊施設が相互に与える影響を考慮した推薦手法について述べる。まず，図 1 に本研究の提案システムの処理の流れを示す。本研究で使用するデータとして，「楽天データ公開」⁴より公開されている，楽天トラベルに関するデータ [18] を用いた。この宿泊施設とユーザレビューのデータおよび TripAdvisor より収集したデータから，必要な情報を選出する。また，本研究の提案システムの宿泊施設の推薦は大まかに 3 点の要素から構成される。それぞれの宿

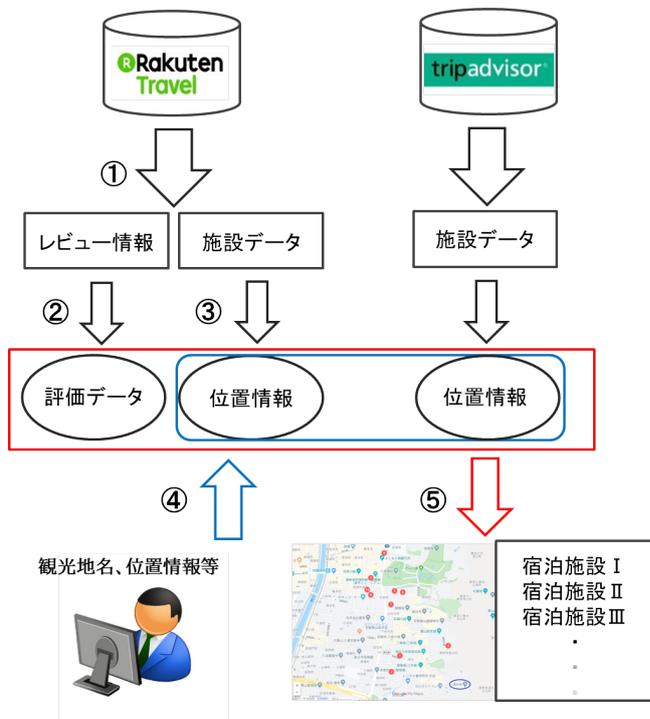


図 1 宿泊施設推薦手法の処理の流れ

泊施設が持つユーザのレビュー評価値・口コミレビュー数をもとに計算されたものを宿泊施設の評価スコア。ユーザが検索したい観光スポットに対して周辺に点在する宿泊施設までの距離から計算したものを距離スコア。宿泊施設のレビューの観光スポットに関連したレビュー分析し数値化したものを関連性スコアとして，この 3 つの観点から推薦ランキングを決定する。提案システムの大まかな手順は以下のようになる。

- (1) 公開されるデータから必要となるデータのレビュー情報と施設データに関するものを抽出する。
- (2) 欠損データ・不足データがある場合は補完作業を行う。
- (3) 公開で取得したデータは位置情報が付加されていないのでジオコーディングにより位置情報の取得を行う。
- (4) 取得したデータから後に計算式で使いやすくするために，前処理を行う。
- (5) 利用者が観光スポット名または住所や緯度経度などの位置情報を入力する。
- (6) 入力に対応した宿泊施設を 3 点の要素から数値化したスコアを総合判断し，ランキング形式でユーザに提示する。

3.1 データ取得および前処理

本研究で必要となる宿泊施設に関するデータは楽天データ公開の楽天トラベルのものを使用した。具体的には，1996 年から 2015 年までに登録されたもので，施設データが約 13 万施設，レビューデータが約 558 万レビューが対象となっている。公開データはユーザ評価データ，レビューデータ，ホテルマスターデータの 3 つにわけられ構成されている。

ここで，本研究の提案システムでは宿泊施設を主に扱うため，関連するデータ属性では不十分だと考えた。具体的には，宿泊施設の五段階評価となる「総合評価」，ユーザによって投稿され

4 : https://rit.rakuten.co.jp/data_release_ja/

表 1 データ属性追加後のホテル・旅館データ

データ属性名	例
施設 ID	121212
施設名	品川楽天タワーホテル
総合評価	4.32
アンケート件数	5678
クチコミ数	4321
住所	東京都世田谷区玉川 1 丁目 2 3 - 4 5
郵便番号	123-4567
緯度	35.6104826
経度	139.6301228

たクチコミの「クチコミ数」、宿泊施設の位置情報となる「住所」や「経緯度」などがあげられる。これらは楽天トラベルがホテル・旅館として Web ページに掲載されている情報をもとに、ホテルマスターデータの新しい属性として追加し、使用した。追加後のホテルマスターのデータ構造を表 1 に示した。同時に、公開データの中には空データや欠損したデータとなっていたデータもあり、情報推薦の精度高める観点や効率的にデータ処理を行いたい理由からデータの補完する処理を行った。

観光スポットに関連したデータは Tripadvisor から収集した。今回の実験では京都府内の観光スポットを想定しているため、Tripadvisor 上のエリアを京都内と指定し、評価順としてランキングされた上位約 1,000 件の観光名所を使用した。提案システムでは入力した観光スポットに関連したものを想定しているため、その観光スポットの詳細なデータやレビューデータが必須ではなく、観光名だけを用いる。将来的な課題として、より観光スポットと宿泊施設の関連性を重視するために、双方の推薦が必要となるとなってくる。よって、観光スポットに関するデータの収集は観光情報サイトの公開している API や SNS などを利用してのデータの分析が考えられる。

3.2 宿泊施設に関する評価スコアの算出

ユーザが宿泊施設を選ぶ際には、その宿泊施設の評判を気にすることがよくあることだと考えられる。総務省の調査⁵によると約 7 割の利用者が気にすると答えている。これを考慮して、本研究では楽天トラベル上のホテル・施設に対して、評価対象となる部分を数値化することで推薦ランキングに反映し、これを宿泊施設の評価スコアとして定義した。複数の要素からなる、評価の数値化は式 (1) の計算方法で表した。

$$E_S = \frac{CE_S}{CE_{max}} \times \log\left(1 + \frac{NA_S}{Mean(a \in A)}\right) \quad (1)$$

式 (1) において、 CE_S は宿泊施設 S における総合評価の数値、 CE_{max} は宿泊施設の総合評価の最大値であり、ここでは 5 である。対数関数の真 $E_{S,score}$ 数部分における NA_S は宿泊施設のアンケート件数の総数であり、 $Mean(a \in A)$ は候補となる全てのホテルのアンケート総数である集合 A における平均値を表す。また、対数の底は 2 とした。対数関数の真数および

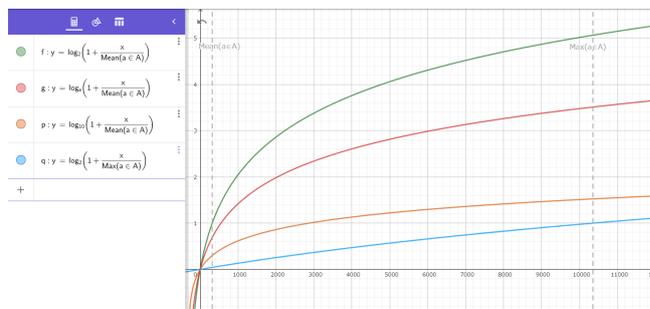


図 2 底および真数の変化に対する対数関数の推移

び底に関しては複数の候補となる関数の中から、適切に推薦ランキングに影響を与えられると考えられるものを選んだ。候補となった関数を図 2 に示す。各関数においての各宿泊施設のクチコミ総数の集合 A における平均値と最大値の関係性を考慮し、式 (1) に含まれる、緑色となる対数関数を使用した。

最後に式 (1) によって求められた値に対して、データの正規化を行い、これを評価スコアとして扱う。推薦ランキングでは 3 つ要素のスコアからなるため、同じ比率にする必要がある。宿泊施設 S における評価スコアを $E_{S,score}$ としたとき、式 (2) のように正規化される。ここで、 $E_{S,score}$ は元の数値となる E_S の値によっては正規化する際にスコアが 0 になるものが生じてしまうことがあるために、+1 することで全ての宿泊施設の評価スコアが最大値 2、最小値 1 になるように正規化を行う。

$$E_{S,score} = \frac{E_S - E_{min}}{E_{max} - E_{min}} + 1 \quad (2)$$

3.3 距離スコアの算出

距離スコアとはユーザが指定された場所に対して、候補となるそれぞれの宿泊施設がどれほど離れているかを示すための数値となる。まず、ジオコーディングにより得られた緯度経度を使用し、2 点間の距離を求める必要がある。一般的に、緯度経度から距離を求める際にさまざまな手法が存在するが、よく使われるものとして、国土交通省国土地理院⁵が使用している算出方法やヒュベニの公式 [19], [20] があげられる。本研究の実験ではこれらの方法以外に、提案手法も加えて、いくつかの手法の測定結果を比較して距離スコアに取り入れる。

まず、観光スポットと宿泊施設 S の緯度経度をそれぞれ P_{lat} , P_{lng} , C_{lat} , C_{lat} とする。この時、提案手法 1 として、2 点間からなる直線距離を D_{S1} とした場合、距離 D_{S1} は式 (3) によって求められる。ここで、 γ とは緯度経度から計算した 2 点間の距離に対して、km に変換する際に使用する係数となる。

$$D_{S1} = \gamma \times \sqrt{(P_{lat} - C_{lat})^2 + (P_{lng} - C_{lng})^2} \quad (3)$$

次に提案手法 2 として、緯度と経度の 1° の差を考慮した距離算出法を D_{S2} としたとき、 α と β を用いて式 (4) で算出する。式 (4) の特徴として、2 点間の距離を計算する平方根の前に距離の単位を度数から km に直してから計算している点があげられる。

5 : <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/html/nc114230.html>

5 : <https://vldb.gsi.go.jp/sokuchi/surveycalc/surveycalc/bl2stf.html>

$$D_{S2} = \sqrt{\alpha^2(P_{lat} - C_{lat})^2 + \beta^2(P_{lng} - C_{lng})^2} \quad (4)$$

比較手法として地球表面の曲率を距離計算に反映できるヒュペニの公式を使用した。これらの手法を実際の地名の緯度経度から距離を測定して精度の比較を行う。そして、適切などと考えられる手法から算出された距離 D_S を用いて、距離スコアとして推薦ランキング反映される。最後に、距離スコアも 3.2 節における評価スコア同様に正規化を行う。しかし、今回の距離スコアは距離が長くなるに連れ数値が大きくなってしまふので、正規化に加え、距離が近いほどスコアが大きくなるようにした。宿泊施設 S と観光スポットに対する距離スコアを $D_{S, score}$ とすると式 (5) で表すことができる。式 (5) により、距離が近いほどスコアが大きくなるようになり、最大値 2, 最小値 1 となるように正規化することができる。

$$D_{S, score} = 2 - \frac{D_S - D_{min}}{D_{max} - D_{min}} \quad (5)$$

3.4 関連性スコアの算出

最後に、観光スポットと宿泊施設がどれほど関わりがあるのかを表す関連性について算出し数値化する。具体的には、宿泊施設に投稿されているクチコミのテキストデータから関連単語となるものがどれほど含まれるかで関連スコアが決められる。本研究では関連単語として 3.1 節で収集された。京都府内にある観光スポット名の集合を $w \in K$ とし、キーワード単語として扱う。宿泊施設 S に対して、関連性を示す R_S は式 (6) として以下のように算出される。 DF_S はキーワードを含んだクチコミ数、 NR_S は宿泊施設 S が有するクチコミの総数、 TF_S は宿泊施設 S におけるキーワードの出現頻度である。ここで、 DF_S に対して NR_S で除算を行うことは全体のクチコミのうち、キーワードを含んだクチコミの割合を意味し、大きいことはそれだけ観光スポットとの関連性が高くなる。しかし、宿泊施設のクチコミ数が少ない場合には極端な数値になる可能性を考慮して、 TF_S を用いた。式 (2) と同様に R_S をもとに正規化を行い、関連性スコア $R_{S, score}$ を算出する。

$$R_S = \frac{DF_S}{NR_S} \times \log(1 + TF_S) \quad (6)$$

3.5 候補宿泊施設ランキングの推薦

最後にそれぞれ算出し正規化を行った、宿泊施設の評判に関する要因を考慮した評価スコア E_{score} 、観光スポットと宿泊施設の距離スコア D_{score} 、宿泊施設が持っている観光スポットとの関連性を表す関連性スコア R_{score} を統合することで、複数要因を総合的に考慮した宿泊施設のスコアを算出する。これを総合スコア T_{score} とした場合、 T_{score} は式 (7) で求められる。

$$T_{score} = E_{score} \times D_{score} \times R_{score} \quad (7)$$

これにより、宿泊施設の評判を考慮した評価スコア、観光スポットとの距離を考慮した距離スコア、観光スポットとの関連性を表す関連性スコアのそれぞれをもとにしたランキングと複数要因を総合的に考慮した宿泊施設の総合スコアをもとにしたランキングをユーザに推薦する。

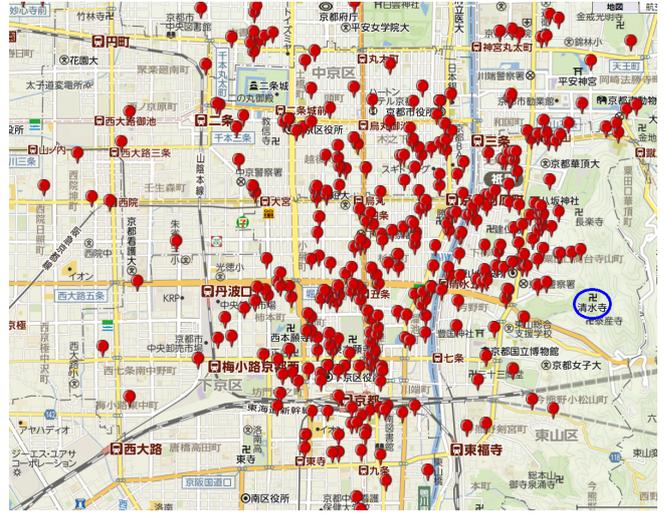


図3 対象の候補となった周辺ホテルの一覧

4 評価実験

4.1 宿泊施設推薦の実行例

本節では、推薦ランキングを提示する際の数値化についての実行例を試した。また、候補となった宿泊施設は多く存在しているため周辺一部の宿泊施設を図3に図示する。「清水寺」を観光スポットとしたときのホテルの評判、観光スポットの距離差、関連性を考慮した宿泊施設の総合ランキングの数値化および観光スポットと候補となった宿泊施設を地図上に表したのを図4に示す。ただし、ここでの図示は数値化した各スコアを表すものであり、ユーザが使用する際には違う表示となることを想定している。また、推薦ランキングでは上位の10件だけを表示するようにした。

推薦ランキングでは宿泊施設の評判、観光スポットの距離差、関連性がすべて考慮されているため、どれかの属性に偏ることはなく、総合的に高くなるホテルが高い順位となっている。たとえば、1位の「京都東山荘」では、5段階評価の点数は4.8点と高かったがその評価回数が168回であり、他のランキング上位と比べると少ないため評価スコアが低くなっている。これに対して、目的地との距離が近くてその関連性も高いことであるため総合スコアが最大となっている。逆に、4位の「アパホテル<京都祇園> EXCELLENT」では評価回数が2947回であり全宿泊施設で最も高く、その分が評価スコアに反映されているのに対して、目的地との距離がかなりあったためこの順位となった。

4.2 距離算出手法の比較実験

本節では、予備実験として 3.3 で示した複数の距離算出手法の比較を行った。まず、奈良市にある奈良駅、東大寺、奈良国立博物館、京都駅の4地点の緯度経度を表2に示す。

次に、距離測定の2地点の区間として A. 奈良駅-東大寺, B. 奈良駅-奈良国立博物館, C. 奈良駅-京都駅の3区間をそれぞれの手法から算出した結果を表3に示す。また、比較参考の対



順位	ホテル名	評価スコア	距離スコア	関連性スコア	総合スコア
1	京都東山荘	1.711	1.858	2.000	6.358
2	京都東山温泉 ホテル りょうぜん	1.756	2.000	1.748	6.139
3	雫 庵【かむろあん】	1.836	1.844	1.616	5.471
4	Aパホテル<京都祇園>EXCELLENT	1.996	1.474	1.730	5.090
5	祇園の宿 舞風館	1.664	1.605	1.855	4.954
6	料理旅館 花染	1.744	1.637	1.652	4.716
7	旅館 富久家	1.653	1.650	1.672	4.560
8	ホテル ささりんどう	1.826	1.541	1.588	4.468
9	知恩院 和順会館	1.679	1.522	1.700	4.344
10	ギオン福住	1.685	1.407	1.773	4.203

図4 宿泊施設推薦ランキングの例（観光スポット：清水寺）

表2 4地点の緯度および経度

地点名	(緯度, 経度)
奈良駅	(34.680365, 135.81898)
東大寺	(34.688985, 135.83981)
奈良国立博物館	(34.683609, 135.83665)
京都駅	(34.985849, 135.75876)

表3 提案手法による距離の算出結果および比較 [km]

区間	提案手法1	提案手法2	ヒュベニの公式	国土地理院
A.	2.28	2.13	2.14	2.13
B.	1.82	1.67	1.66	1.66
C.	31.52	34.34	34.33	34.33

象として国土地理院の Web サイトから緯度経度を入力して算出された距離を追加した。

表3の結果から、ヒュベニの公式と国土地理院の手法ではどの区間においてもほぼ同じ結果となったことがわかる。これに対して、提案手法1においてはおよそ8%の誤差が確認できた。一方、提案手法2の結果ではどの区間においても10m以下の誤差となり、とても精度が高いものとなったことがわかる。

これを踏まえて、推薦システムの評価実験では提案手法2を使用することとした。理由としては、本研究では大量の観光スポットや宿泊施設を扱う。しかしながら、ヒュベニの公式と国土地理院の手法では計算式が複雑である。これを推薦ランキングごとに位置情報の処理が必要となるため、式が簡潔で軽量化を図るとともに精度の良い提案手法2を3.3節における D_S として用いることにした。

4.3 アンケート調査による推薦ランキングの有用性検証

評価実験では4.1節で用いた推薦ランキングの有用性を確かめるために、被験者には提案した推薦システムを用いて観光スポットを検索を行ってもらい、推薦された宿泊施設ランキングに対するアンケート調査を実施した。具体的な実験内容として「京都にある観光名所『清水寺』を徒歩で観光する時の宿泊施設を探したい」と想定して、用意した4つの宿泊推薦ランキングについて、それぞれの設問項目についてアンケートの回答を行ってもらった。用意した4つの宿泊推薦ランキングはそれぞれ違うランキング算出方法であり、以下のように定めて実験を行った。本研究からの提案手法による推薦ランキングは、ランキング①、ランキング②、ランキング③である。ランキング④は、既存サービスとして楽天トラベルによる「清水寺」に対する推薦ランキングを使用して比較した。各被験者に対して4つのランキングを提示した。また、以下の宿泊施設ランキングを提示する際は、それぞれの比較を行うために何を考慮したランキングかは示さず、提示する順序も変えて行った。

- (1) ランキング①：観光スポットと宿泊施設の距離を考慮した宿泊施設ランキング (D_{score})
- (2) ランキング②：観光スポットとの距離および宿泊施設の評価を考慮した宿泊施設ランキング ($D_{score} \times E_{score}$)
- (3) ランキング③：観光スポットとの距離、宿泊施設の評価および観光スポットとの関連性を考慮した宿泊施設ランキング ($D_{score} \times E_{score} \times R_{score}$)
- (4) ランキング④：楽天トラベルによる「清水寺」に対する宿泊施設の推薦ランキング

4.3.1 推薦ランキングの評価項目

まず、提示したランキングに対して項目1-1~1-3, 2-1~2-4, 3-1~3-4, 4-1~4-4を5段階評価(1.全くそう思わない, 2.そう思わない, 3.どちらともいえない, 4.そう思う, 5.とてもそう思う)を行ってもらった。

ランキング① (D_{score}) に対して、以下の設問項目についてアンケートを実施した。

- 項目1-1 宿泊施設の評価が適切に反映された
- 項目1-2 ランキングの順位が適切であった
- 項目1-3 宿泊したいホテルが上位にきていた

ランキング② ($D_{score} \times E_{score}$) に対して、以下の設問項目についてアンケートを実施した。

- 項目2-1 宿泊施設の評価が適切に反映された
- 項目2-2 目的地の情報が宿泊施設のクチコミ(レビュー)に含まれていた
- 項目2-3 ランキングの順位が適切であった
- 項目2-4 宿泊したいホテルが上位にきていた

ランキング③ ($D_{score} \times E_{score} \times R_{score}$) に対して、以下の設問項目についてアンケートを実施した。

- 項目3-1 宿泊施設の評価が適切に反映された
- 項目3-2 目的地の情報が宿泊施設のクチコミ(レビュー)に含まれていた
- 項目3-3 ランキングの順位が適切であった

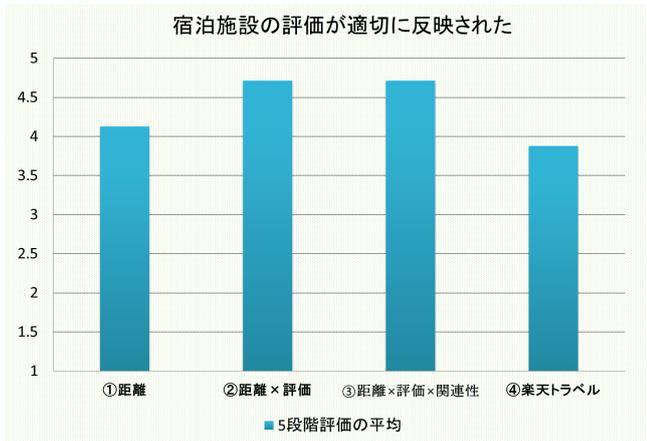


図5 「宿泊施設の評価が適切に反映された」の結果

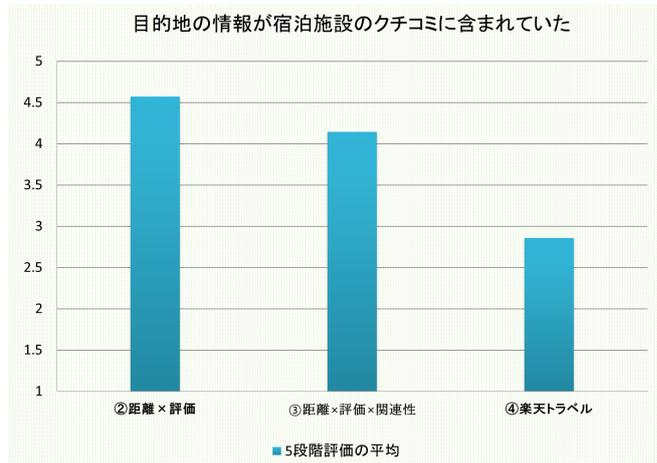


図6 「目的地の情報が宿泊施設のクチコミに含まれていた」の結果

項目 3-4 宿泊したいホテルが上位にきていた

ランキング④（楽天トラベル）に対して、以下の設問項目についてアンケートを実施した。

項目 4-1 宿泊施設の評価が適切に反映された

項目 4-2 目的地の情報が宿泊施設のクチコミ（レビュー）に含まれていた

項目 4-3 ランキングの順位が適切であった

項目 4-4 宿泊したいホテルが上位にきていた

これらのランキングを比較するため、以下の設問項目についてアンケートを実施した。

項目 5-1 推薦ランキング4つを結果が良かったと思う順に書いてください

項目 5-2 そう思った理由を空欄に書いてください

最後に、提示した推薦ランキングと提案手法に対して以下の記述項目についてアンケートを回答してもらった。なお、設問項目7では提案手法によるランキングを指定していることから、被験者にランキング内容を知られないためにそこまでの用紙を回収したのちに設問項目7に答えてもらった。

項目 6 推薦ランキングに追加してほしい機能または改善してほしい点を書いてください

項目 7 ランキング②およびランキング③について良かった点と悪かった点を書いてください

4.3.2 アンケート結果に対する考察

本節では、4.3.1節の設問項目に対する回答を比較しながら考察について述べる。最初に、推薦ランキングの宿泊施設に対する評価が適切に反映されたかどうかを確認した。それぞれのランキングに対する設問項目1-1、2-1、3-1、4-1の5段階評価を比較するため、図5に5段階評価の平均値を示す。図5の結果から、既存サービスであるランキング④と評価スコアを考慮していないランキング①より、ランキング②およびランキング③がいずれも高い評価となった。

推薦ランキングが観光スポットとの関連性を考慮しているか確認するために、目的地の情報が宿泊施設のクチコミに含まれていたを問う項目2-2、3-2、4-2の5段階評価の平均値を図6に示し比較する。既存サービスのランキング④の5段階評価の平均値は3.0以下と低い結果となった。これに対して提案手法

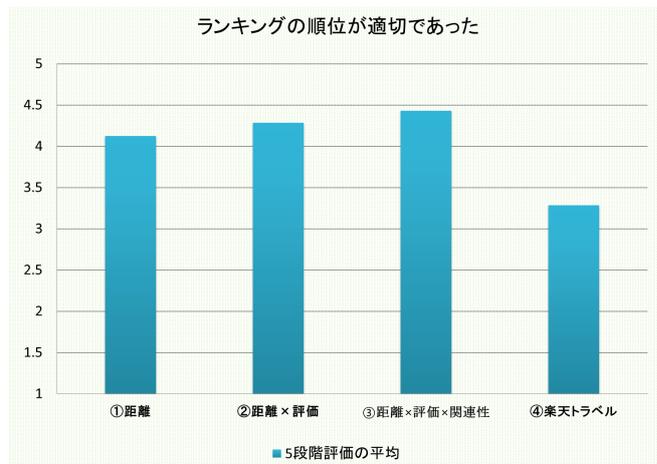


図7 「ランキングの順位が適切であった」の結果

のランキング②とランキング③は両方4.0以上の結果ではあったが、関連性スコアを考慮した③が最大ではなかった。

次に、それぞれのランキングで提示された宿泊施設の順位が適切かどうかを項目1-2、2-3、3-3、4-3の5段階評価の平均値を図7に示す。結果として提案手法のランキングはいずれも高い平均値となった。特に距離スコアと評価スコアと関連性スコアを同時に考慮したランキング③が最も良い評価となった。

続いて、ランキングで提示された宿泊施設の中で実際に泊まっていた宿泊施設が上位にあったかに関して、項目1-3、2-4、3-4、4-4の5段階評価の平均値を図8に示す。こちらも同様に、結果として提案手法のランキングはいずれも高い平均値となったが、距離スコアと評価スコアを考慮したランキング②が最も高い平均値となった。

上記の結果をまとめてから、いずれの項目の比較においても提案手法による推薦ランキングが既存サービスとなる楽天トラベルのランキング④より高い平均値となった。ランキング④が比較的低い水準となった原因は、楽天トラベルによる推薦基準が宿泊施設の5段階評価とその評価回数に大きく依存したためと考えられる。そのため、ランキング④では「清水寺」との距離差が提案手法によるランキングと比べると離れていることがわかる。加えて、図6から宿泊施設のクチコミにて目的地の情

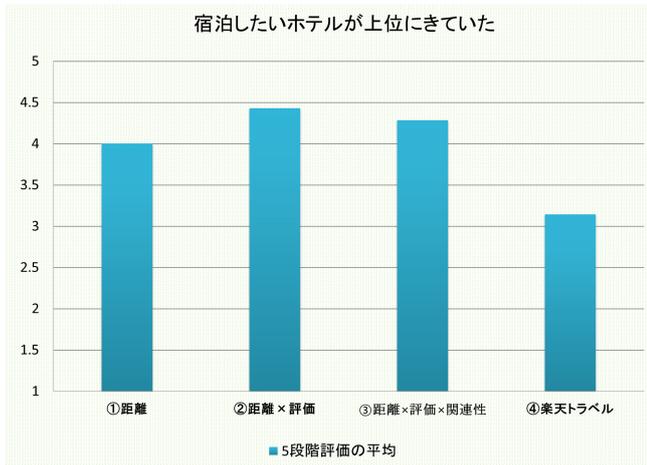


図8 「宿泊したいホテルが上位にきていた」の結果

報に関する数値が低いため、関連性が高い宿泊施設が提示できないという問題点があげられる。これに対し、ランキング①では距離スコアのみ考慮したランキングだが、高めの数値となっていることからユーザが如何に距離を重視しているか読み取れる。一方、提案手法によるランキングでランキング②とランキング③が全体的に高い評価となった。複数の要素を総合的に考慮したことでランキング①より精度が上がった結果といえる。

設問項目 5-1 においては、被験者に答えてもらったランキングの良かったと思う順番を点数化した。1番良かった思った場合は4点、2番の場合は3点、3番の場合は2点、4番の場合は1点として採点した。それぞれのランキングの点数と全体点数の占める割合を図9に図示する。これにより、ランキング②が最も順位が高い結果となり、今回の目的として適した宿泊施設の推薦となった。また、項目 5-2 では順位を選ぶ際、以下の理由などをもとに判断したと述べられた。

- レビューの数が多いほうが参考になるかなと思ったから
- 5段階評価と距離を重視したから
- 泊まりたいと思える宿が上位に来ているため
- 観光なら距離は少し遠くてもいいと思っているから
- レビューの内容が納得できるものが多かった
- 距離の近さを優先的に見て次に評価回数の多さを見たから
- 距離と評価を総合的に反映していると判断したから

以上より、宿泊施設の5段階評価と評価数、レビュー数を基に判断した意見が多くあげられた。また、距離に関しては「観光目的の旅行なら距離は少し遠くてもいい」と「距離の近さを優先的に」と分かれた意見があったが、「観光目的の旅行なら距離は少し遠くてもいい」という意見の被験者はレビュー数の次に距離を重視していると述べていたため、概ね距離を参考しているものとなった。しかし一方では、アンケートの全体を見ても「関連性」についての記述は少なく、これは被験者にとっては見えない潜在的な要素であることが大きな要因だと考えられる。加えて、図9でも示されたように、距離スコアと評価スコアからなるランキング②が最も点数が高いこともこれを示唆している。これに対して、今後の課題として宿泊施設の関連性を

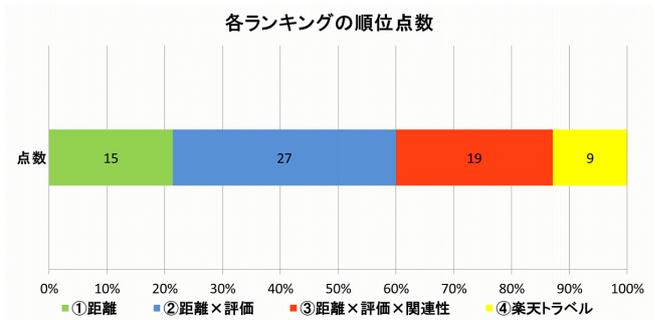


図9 各ランキングの順位点数と割合

ユーザに見える形で表すことがテーマになると考えられる。

設問項目 6 では推薦ランキングに追加してほしい機能と改善すべき点について下記の意見があげられた。

- 料金についての記載の追加
- 価格を範囲指定で反映する機能
- 検索期間に対応した価格の表示
- 周辺の駅または交通機関の情報
- 宿泊利用者の年代分布の表示
- ホテルの料金水準と清潔感を気にするのでその表示機能
- 宿泊施設のサービスおよび清潔感を考慮すべきだ
- 残念なレビュー・無駄なレビューを弾くような方が良い
- 値段を指定できるようになるとシステムとしてアプリ化ができる

回答の中に、多く見られたのが宿泊の料金に関するもので、単に料金の表示する以外にも検索期間に対応した価格表示する機能、価格の範囲を絞り込む機能があげられた。本研究の提案手法では料金を考慮せずに推薦ランキングを提示していた。これは料金を考慮したいが難しいためであり、その理由としてデータの公開がされていない点と料金の時期によって変化する点があげられる。これ問題に対して、料金をスコアとして推薦の計算式に取り込むのではなく分別すべく、サービス化の際やアプリケーション化の際に表示することや絞り込む機能を実装することで解決すべきと考えた。また、交通・アクセス情報に関することも同様にいえると考えられる。一方、「残念なレビュー・無駄なレビューを弾く」については、レビュー文字数の制限やレビュー分析などの処理を行うことで除外することができる。最後にサービス、清潔感に対してあげられた意見に対して、サービスならば「評価5(サービス)」が参考になると考えられる。清潔感であるなら「評価2(部屋)」「評価6(設備)」が同義ではないが判断材料となり得ると予想できる。

最後に、設問項目 7 では提案手法による推薦ランキングであるランキング②とランキング③について良かった点と悪かった点の回答は下記の通りである。良かった点として、

- 高評価と近い距離のホテルが上位にきていた
- レビュー数が多くてバランスが良いと思った
- 泊まりたいと思うのが上位にランクインしていた
- 両方とも上位は高得点のホテルだったが、レビューの内容を見て納得できるものが多かった
- 距離が近くて評価が高い宿が上位にある

・ ランキング③の順位は精度が良く、評価回数と距離を総合的に考慮されている

- ・ レビューに「清水寺」に関する情報多く含まれていた
- ・ 宿泊施設の順位が妥当で、目的地との関連性について知ることができる

などの意見があった。多くの回答が評価と距離、または総合的に適切に反映されていたものであり、実際に宿泊してみたいなどの意見があげられた。中でも「レビューに『清水寺』に関する情報多く含まれていた」、「宿泊施設の順位が妥当で、目的地との関連性について知ることができる点が良かった」の記述もあり、今回の研究の目的に関する意見が確認できた。

悪かった点としては、

- ・ 単純に距離のみで比較しており交通手段を比較していない
- ・ ランキング②ではレビューが少なめでキーワードも少ないホテルがあった
- ・ ランキング②は宿泊施設の評価に偏っている
- ・ 価格について知りたい
- ・ 最寄りの駅が考慮されていない

という意見が得られた。ランキング②だけに関する記述があり悪い点としてあげられていたが、これはランキング②の推薦手法として距離スコアと評価スコアをもとに算出されているため、適切な意見だと考えることができる。「単純に距離のみで比較しており、交通の便を比較していない」について、本研究の距離測定方法が2点間距離によるものであったためと考えられる。これに対し、推薦の計算式としては引き続き2点間距離を用いて算出し地図上に宿泊施設を表示するが、サービス化を想定した際はランキングから選択したホテルについてのルート検索結果を提示することによる解決を考えている。

5 おわりに

本論文では観光スポットと宿泊施設の関係性分析に着目したインタラクティブな推薦システムの提案を行った。観光Webサイトと楽天トラベルの公開情報をもとに、必要となるレビューと施設を選出し、投稿されたレビュー数および施設の評価から算出した施設の評価スコアと観光スポットと宿泊施設との位置情報から計算した距離スコアから利用者から見つけたい施設をランキングしユーザに提示する手法を提案した。アンケート調査による提案手法の有用性が確認できた。

今後の課題として、提案手法である投稿されたレビュー数と施設の評価から算出する施設の評価スコアをの定義などの課題があげられる。また、評価、距離と関連性以外の要素（たとえば、料金、季節、交通手段など）も検討する必要がある。

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。また、本研究の一部は、JSPS 科研費JP17K12686の助成を受けたものである。ここに記して謝意を表す。

- [1] インバウンド訪日外国人動向。
https://www.jnto.go.jp/jpn/statistics/visitor_trends/. 日本政府観光局, 2019年01月.
- [2] 王元元, 岳五一. Twitterを活用した地方都市の観光客誘致戦略. パーソナルコンピュータ利用技術学会論文誌, Vol. 13, No. 1, pp. 16-24, 2019年3月.
- [3] 平久江知樹, 早川智一, 疋田輝雄. マイクロブログにおけるジータグのクラスタリングを用いたマイナー観光地抽出手法の改良. DEIM Forum 2018, H1-5, 2018年3月.
- [4] 開地亮太, 檜垣泰彦. 観光地推薦システムへの単語分散表現の適用. 電子情報通信学会技術研究報告, Vol. 116, No. 488, pp. 129-134, 2017年3月.
- [5] デジタル化による旅行スタイルの変化.
<https://www.kankokeizai.com/>【データ】デジタル化による旅行スタイルの変化/. 観光経済新聞社, 2018年7月.
- [6] 川村秀憲, 鈴木恵二, 山本雅人, 松原仁. 観光情報学. 情報処理学会デジタルプラクティス, Vol. 3, No. 4, 2012.
- [7] 斎藤一. Webにおける観光情報提供と分析(〈特集〉観光と知能情報). 人工知能学会誌, Vol. 26, No. 3, pp. 234-239, 2011年.
- [8] William C. Hunter. The Social Construction of Tourism Online Destination Image: A Comparative Semiotic Analysis of the Visual Representation of Seoul. *Tourism Management*, Vol. 54, pp. 221-229, 2016.
- [9] G. McCartney, R. Pao Cheng Pek. An Examination of Sina Weibo Travel Blogs' Influence on Sentiment Towards Hotel Accommodation in Macao. In *Journal of China Tourism Research*, Vol. 14, No. 2, pp. 146-157, 2018.
- [10] Kwan Hui Lim. Recommending Tours and Places-of-Interest based on User Interests from Geo-tagged Photos. In *Proceedings of the 2015 ACM SIGMOD on PhD Symposium (SIGMOD '15 PhD Symposium)*, pp. 33-38, May 2015.
- [11] 新妻弘崇, 新井晃平, 太田学. 観光ルート推薦のための効率的な制約条件. 情報処理学会論文誌データベース(TOD), Vol. 9, No. 2, pp. 34-45, 2016年6月.
- [12] 中嶋勇人, 新妻弘崇, 太田学. 位置情報付きツイートを利用した観光ルート推薦. 研究報告データベースシステム(DBS), Vol. 2013, No. 28, pp. 1-6, 2013年.
- [13] 石野亜耶, 藤井一輝, 藤原泰士, 前田剛, 難波英嗣, 竹澤寿幸. 旅行ブログエントリと質問応答コンテンツを利用した旅行ガイドブックの情報拡張. 人工知能学会論文誌, Vol. 29, No. 3, pp. 328-342, 2014.
- [14] 土田崇仁, 遠藤雅樹, 加藤大受, 江原遥, 廣田雅春, 横山昌平, 石川博. Word2Vecを用いた地域やランドマークの意味演算. DEIM Forum 2016, H5-1, 2016.
- [15] 三笠弘貴, 奥野拓. 観光サイトにおける閲覧目的に基づいた旅行記概要の動的生成. 情報処理学会研究報告, DD, Vol. 2014, No. 4, pp. 1-8, 2014年3月.
- [16] Keigo Sakai, Akiyo Nadamoto. Extracting Welcome News from Travel Reviews. In *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2015)*, pp. 141-148, 2016.
- [17] 潘健太, 北山大輔. 観光スポット検索のためのユーザのレビュー選択と特徴抽出に関する考察. 信学技報, Vol. 118, No. 107, DE2018-5, pp. 21-24, 2018年6月.
- [18] 楽天株式会社. 楽天トラベルデータ. 国立情報学研究所情報学研究所データリポジトリ. (データセット). <https://doi.org/10.32130/idr.2.2>, 2016年.
- [19] Karl Hubeny. Zur Entwicklung der Gauss'schen Mittelbreitenformeln. *sterreichische Zeitschrift fr Vermessung und Geoinformation*, 42, pp. 8-17, 1954.
- [20] Karl Hubeny. Weiterentwicklung der Gauss'schen Mittelbreiten Formeln. *sterreichische Zeitschrift fr Vermessung und Geoinformation*, 84, pp. 159-163, 1959.