

観光スポットと宿泊施設の関係性分析に基づく インタラクティブな推薦システムの提案

佐藤 海龍[†] 王 元元[†] 河合由起子^{††,†††}

[†] 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

^{††} 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

^{†††} 大阪大学サイバーメディアセンター 〒567-0047 大阪府茨木市美穂ヶ丘 5 番 1 号

E-mail: [†]{i036vg, y.wang}@yamaguchi-u.ac.jp ^{††}kawai@cc.kyoto-su.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できていない。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明らかにすることで、ユーザのニーズに合わせたインタラクティブな推薦システムの構築を目指す。本論文では、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とレビューの類似度に基づき関係性を抽出することで、推薦ランキングを相互的に提示できるインタラクティブなシステムを提案する。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できる。また、利用者は宿泊施設を探す際に、おすすめの観光スポットも検索できる。

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挙げられる^(注1)。平久江ら [2] はマイクロブログの Twitter からジオタグ付きツイートを収め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、開地ら [3] 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 を使っ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。旅行を計画する際に使われる

TripAdvisor^(注2) や Yahoo!トラベル^(注3) などの観光情報サイトでは観光スポットや宿泊施設についての情報が掲示され、実際に訪れたユーザレビューが多く投稿されている。潘ら [4] 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・タイプごと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。

本研究では、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、観光スポットと宿泊施設のインタラクティブな推薦システムを提案する。一般的な検索サイトでは、どちらかを検索した後に改めて残りのほうを検索する必要がある。この手間を減らすため、ユーザが検索した観光スポットか宿泊施設に対して、投稿されたレビューの解析や評価、物理的な距離などの要素から宿泊施設もしくは観光スポットをランキングしユーザに推薦する。これにより旅行の計画の際に観光スポットの検索と同時に宿泊施設も検索できるようになる。同じように宿泊施設を探す際に、おすすめの観光スポットも調べることができる。

本論文の構成は以下のとおりである。次章では、提案する推薦システムの概要について述べ、3 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、4 章でまとめと今後の課題について述べる。

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

(注2) : <https://www.tripadvisor.jp/>

(注3) : <https://travel.yahoo.co.jp/>

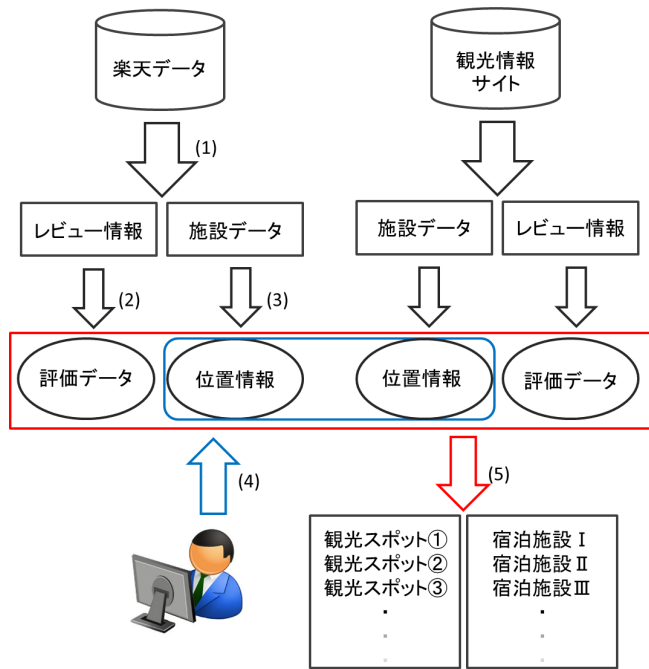


図1 提案システムの処理流れ

2. 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インタラクティブな推薦手法について述べる。楽天トラベルが公開している宿泊施設とユーザーレビューのデータおよび TripAdvisor のデータから、必要な情報を選出する。また、本研究ではお互いが与え得る影響として、ユーザのレビュー評価値・レビューの文章の形態素解析結果を評価値として、緯度経度物理的な距離との複数の要素から算出する。提案システムの全体の流れを図1に示す。大まかな処理としては以下になる。

- (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) 取得したレビューデータのうち評価の数値・テキストデータなどから評価値を算出する。
- (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を行う。
- (4) 利用者が観光スポットか宿泊施設を入力する。
- (5) 入力に対応した宿泊施設もしくは観光スポットをランキング形式で表示する。

2.1 データ取得

本研究では宿泊施設では楽天トラベル、観光情報は TripAdvisor を使用した。楽天トラベルでは2016年までに公開した施設データのうち2015年度と2016年度のデータを利用し分析する。施設データでは海外の施設も含んでいるので、日本国内だけに限定した。ユーザレビューでは、投稿した施設の「旅行の目的」の項目において、「レジャー」となるレビューだけを選定した。なお、将来的にはレビュー文章中に「観光」や「旅行」といった観光目的となるような単語を含むものだけを選定し、2つの方法を比較する予定である。また、今回は観光スポット

表1 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)

と宿泊施設の関係性分析を行うにあたり、緯度経度の情報が必要だが、住所の情報しかないため、郵便番号・住所などからジオコーディングを行った。

2.2 距離算出

ジオコーディングにより得られた緯度経度から、距離スコアとして2点間の距離を求める。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^(注4)が使用している算出方法やヒュベニの公式があげられる。本研究の実験ではいくつかの手法から結果を比較して精度が良いものを使用する。

まず、提案手法1として2点間の距離を D_1 とする時、観光スポットと宿泊施設の緯度経度をそれぞれ P_{lat} , P_{lng} , C_{lat} , C_{lng} とすると距離 D_1 は式(1)で算出する。

$$D_1 = \gamma \times \sqrt{(P_{lat} - C_{lat})^2 + (P_{lng} - C_{lng})^2} \quad (1)$$

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。具体的には緯度1度あたりのkm距離 α , 経度1度あたりのkm距離 β から、平均を使用したものである。

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_2 としたとき、 α と β を用いて式(2)で算出する。式(2)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。

$$D_2 = \sqrt{\alpha^2(P_{lat} - C_{lat})^2 + \beta^2(P_{lng} - C_{lng})^2} \quad (2)$$

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ベ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。

2.3 候補地ランキングの提示

最後に利用者に対応した施設・スポットのランキング方式について述べる。前節で求めた距離スコアに加え施設の総合評価値、投稿レビュー数から計算した施設の評価スコアの2つの数値から適切な施設・スポットを数値の大きさ順に利用者に推薦する。

3. 評価実験

3.1 距離算出手法の比較実験

実験では、はじめに2.2で示した複数の距離算出手法の比較を行った。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1に示す。

(注4) : <https://vldb.gsi.go.jp/sokuchi/surveycalc/surveycalc/bl2stf.html>

表 2 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

次に、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 2 に示す。また、比較参考の対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 2 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分かる。これに対して、提案手法 1 においてはおよそ 8% の誤差があったが、提案手法 2 がどれにおいても 10m 以下の誤差となり、とても精度が高い結果となった。

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。理由としたは、本研究では大量の観光スポットや宿泊施設を扱う。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共に精度の良い提案手法 2 を使用する。

3.2 推薦システムの評価

評価実験では提案した推薦システム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および宿泊施設の検索を行ってもらい、結果に対するアンケート調査を実施する予定である。具体的には、被験者が奈良市内にある観光スポットを検索するならば、検索結果の観光スポット一覧を地図上に表示して行きたいスポットを選択する。観光スポットを決定したのち、この観光スポットに対応したおすすめの宿泊施設を提示して選択してもらう。逆も同じように、宿泊施設を選択したのちにおすすめの観光スポットを推薦する。テストユーザに推薦の結果が満足いくものかのアンケート調査を行う予定である。

4. 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。

今後の課題として、提案した推薦システムの有用性を検証するため、被験者によるアンケート調査を行う予定である。また、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。

謝 辞

本研究では、楽天株式会社が国立情報学研究所の協力により

研究目的で提供している「楽天公開データ」を利用した。また、本研究の一部は、JSPS 科研費 17K12686 の助成を受けたものである。ここに記して謝意を表す。

文 献

- [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] 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オ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.
- [3] 開地亮太, 檜垣泰彦: 潜在的興味に基づく観光地推薦システムの試作, 信学技報, LOIS2015-14, 2015 年 7 月.
- [4] 潘健太, 北山大輔: ユーザのレビュー選択に基づく観光スポット検索手法, DEIM Forum 2018, H1-2, 2017.