

嗜好の合致度と非接触度に基づく穴場スポットの個人化

松本 華奈[†] 北山 大輔[†]

[†] 工学院大学情報学部システム数理学科 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: †j318256@ns.kogakuin.ac.jp, †kitayama@cc.kogakuin.ac.jp

あらまし インターネットなどを利用して観光スポットの情報を取得しようとする、過去に訪れた観光スポットや訪問予定の観光スポットと類似しているにも関わらず、見つけることが困難な観光スポットがある。このような個人の嗜好に合致するが、発見が困難な観光スポットを「個人の穴場スポット」と定義した。本研究では、観光スポットの説明文とレビュー文に出現する単語の tf-idf 値を計算し、過去に訪れた観光スポットと推薦候補の観光スポットの類似度を嗜好との合致度として算出する。さらに、推薦候補の観光スポットごとに算出した、ユーザとその観光スポットの非接触度と合わせ、個人の穴場スポットを推薦する手法を提案する。

キーワード 観光地推薦, 穴場スポット, 非接触度, Web 閲覧履歴, 訪問履歴

1 はじめに

旅行の計画を立てるときや観光スポットを探すとき、一般的に観光情報サイトやガイドブックが利用され、検索する単語には観光スポット名や観光地名が用いられる。このとき、検索結果には検索した単語と同じ単語が含まれる観光スポットが推薦されやすい。また、観光地の中でも一般的に知名度の高い観光スポットがランキングの上位に表示されやすい。しかし、検索結果に表示されない観光スポットでも、ユーザの嗜好に合い、そのユーザにとっては知名度の高い観光地と同等の満足度が得られる観光スポットがあると考えられる。本研究では、このようにユーザの嗜好に合うにも関わらず発見に至りにくい観光スポットを「個人の穴場スポット」と定義する。しかし、従来の研究によって推薦される穴場スポットは、一般的な評価が高く、知名度が低い場所と定義されていることが多い。したがって、個人の穴場スポットを推薦するには異なる指標が必要となる。

そこで本研究では、一般的な評価に対する個人の評価として、嗜好との合致度を考え、知名度を個人化したものとして、非接触度を考える。個人の穴場スポットを発見するため、ユーザが閲覧している Web ページから、観光スポットを探しているエリアを推定し、エリア内の観光スポットに対し嗜好との合致度と非接触度を算出する。そして、それらを掛け合わせた値が高いスポットを、そのユーザの穴場スポットの候補として推薦する。そのため、本手法は、ユーザが観光地検索を行なっている最中に推薦結果を逐次更新していく。本研究では、じゃらん net¹などの観光レビューサイトの観光スポットを対象とする。

本研究における嗜好との合致度とは、推薦候補の観光スポットとユーザが過去に訪れた観光スポットとの類似度であり、観光スポットの説明文とレビュー文に出現する単語の tf-idf 値を計算し、その値を要素としたベクトルで計算したコサイン類似度から得られる。非接触度は、ユーザが推薦候補の観光スポットと接触する難度を表す値であり、ユーザごとの接触の難しさ

と一般的な接触の難しさを掛け合わせて算出する。ユーザごとのその観光スポットに対する接触の難しさとして、ユーザが閲覧したページの中で、その観光スポット名が出現した回数の逆数を用いる。また、一般的な接触の難しさとして、その観光スポットのレビュー数の逆数を用いる。これらの手法により、以下の2つの研究課題の解決を目指す。

RQ1 非接触度を用いることで、ユーザが個人的に知らないスポットを抽出可能か。

RQ2 嗜好のみを用いるよりも非接触度を用いることは、推薦の性能に寄与するか。

本稿の構成を以下に示す。2章では関連研究、3章では提案手法、4章では予備検討、5章では評価実験、6章ではまとめを述べる。

2 関連研究

2.1 個人の嗜好に合う観光スポットの発見に関する研究

個人の好む観光スポットを発見する研究は数多く行われている。このような研究の多くは、利用者の嗜好を分析し、その嗜好と類似した観光スポットを推薦するものが多い。市村ら [1] は、観光スポットの口コミの解析と利用者の好み診断により、手軽に旅行先を決定できるシステムを提案した。旅行サイトの口コミデータセットを機械学習したシステムが、ユーザの好みを診断する質問を自動生成する。ユーザの回答に基づき、ユーザの嗜好に合った旅行先を提案する。小南ら [2] は、ユーザが選択した観光スポットの画像から、画像に付与されているタグをベクトルの要素とし、コサイン類似度によって観光スポットを推薦するシステムを構築した。ユーザが言語化できない、潜在的な嗜好と類似する観光スポットの推薦を可能にする。奥蘭ら [3] は、検索作業を行わず、ユーザが選択した観光イメージの画像により個人の嗜好を分析し、その嗜好を反映した観光地を推薦するシステムを構築した。

嗜好との合致度については、これらの研究による手法をそ

1: <https://www.jalan.net/kankou/>

のまま用いることが可能である。本研究では、単純な、観光スポットの説明文とレビュー文中の名詞の類似度を利用し、非接触度を用いる効果を明らかにすることに着目する。

2.2 穴場スポットの発見に関する研究

穴場の観光地推薦に関する研究も、これまでに多く行われている。Kitayama [4] は、知名度は低いが他の景観地点と比べて同等以上の満足が得られる景観地点を穴場スポットと定義した。景観地点とは、景観を眺めることに適した観光スポットであり、Flicker の投稿写真に付与されるジオタグをクラスタリングし決定した。クラスタ内の写真のお気に入り数の総和を撮影者数で割った値を穴場スポット度と定義し、これが高いクラスタを穴場スポットの候補とした。平久江ら [5] は、観光客がマイクロブログを頻繁に投稿している場所を、観光ガイドブックに掲載されない穴場スポットと定義した。主要観光地周辺で投稿されたジオタグ付きの Twitter の投稿のうち、観光客と判定したユーザのツイートのみをクラスタリングし、観光地点候補を発見する。観光客判定は投稿数、キーワード、投稿期間により判定する。

これらの研究に対し、本研究では一般的な穴場スポットではなく、よりユーザの嗜好に合った穴場スポットを推薦することを目指している。ユーザごとにその観光スポットの情報を目にするの難しさに着目することで、ユーザからみた穴場スポットの推薦が可能になると考える。

2.3 観光情報の取得・可視化に関する研究

観光情報の取得や、可視化の研究も数多く行われている。渡邊ら [6] は、有益な観光情報可視化のため、LDA (Latent Dirichlet Allocation) による観光特徴抽出に関する研究を行った。抽出した観光情報を基に、観光地名と調べたい要素を検索すると、その観光地の情報とともに、類似した観光地の情報を可視化するシステムを開発し、LDA が有用であることを明らかにした [7]。

廣嶋ら [8] は、地理情報検索支援のため、ポアソン確率を用いて場所に関する特徴的な語を抽出する手法を提案した。ブログ記事から特徴語を抽出し地点の特定を行い、有用性を明らかにした。松本ら [9] は、廣嶋ら [8] の手法を用い、クチコミから観光地を表す言葉を抽出した。抽出した言葉をその観光地の特徴語とし、ユーザに提示することで観光地検索の支援を行う手法を提案した。

本研究では、ユーザが閲覧した Web ページの情報に着目した。観光情報を取得するために閲覧した Web ページに出現する観光情報に着目することで、ユーザが言語化できていない好みの観光スポットや、詳しく調べていないが見逃している情報まで網羅することが可能になると考える。

3 提案手法

3.1 提案手法の処理の流れ

提案手法の処理の流れを図 1 に示す。提案システムでは観光スポットとその説明文とレビュー文、緯度・経度を必要とする。

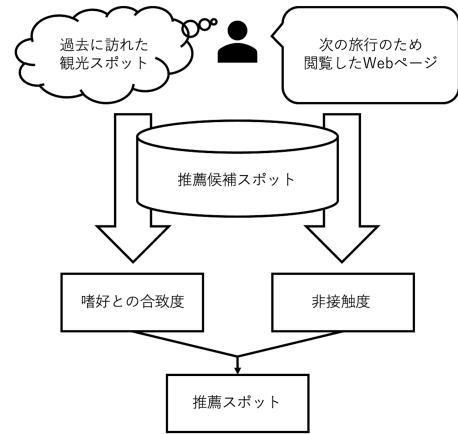


図 1 提案手法の処理の流れ

また、提案システムは、ユーザが次に訪れる観光スポットを探すとき、まだユーザが見つけれられていないが、ユーザの嗜好に合いそうなものを提示するために使うことを想定している。そこで最初に、ユーザが過去に訪れた観光スポットと、次の旅行のために閲覧した Web ページを取得する。過去に訪れた観光スポットは、ジオタグ付き写真や移動ログから自動で取得することが考えられるが、本稿では最初にアンケートでユーザが入力した観光スポットとする。

次に、取得した Web ページによって推定したエリアから、エリア内のすべての観光スポットを取得し、推薦候補の観光スポットとする。過去に訪れた観光スポット H と推薦候補の観光スポット c に対し、嗜好との合致度 $sim(c, H)$ を算出する。嗜好との合致度とは、過去に訪れた観光スポットと推薦候補の観光スポットの類似度であり、それぞれの説明文とレビュー文に出現する単語の tf-idf 値を計算し、その値を要素としたベクトルで計算したコサイン類似度から得られる。

取得した観光スポット c に対し、非接触度 $con(c)$ を算出する。非接触度はユーザごとの接触の難しさと一般的な接触の難しさによって算出する。嗜好との合致度と非接触度を掛け合わせ、値が大きいスポットを穴場スポットの候補として推薦する。穴場スポットの候補は、次の式 (1) より求められる。

$$rec(c) = sim(c, H) \times con(c)^\alpha \quad (1)$$

ここで sim は、嗜好との合致度であり 3.3 節で定義する。 con は非接触度であり 3.4 節で定義する。 α は重みである。

3.2 エリアの推定

本研究では、次に旅行に行く計画を立てているとき見逃してしまっているスポットを推薦する。そのため、個人の穴場スポットとしては、全ての観光スポットを対象とするのではなく、ユーザが旅行を計画しているエリアに限定して推薦を行う。ここで我々は、ユーザの閲覧行動を用いることで、旅行を計画しているエリアを推定できると考えた。エリアはユーザが閲覧した Web ページに出現する観光スポットの MBR を拡大した範囲とする。MBR (Minimum Bounding Rectangle) とは、緯度・経度の最小値・最大値で囲んだ四角形のことであり、閲覧

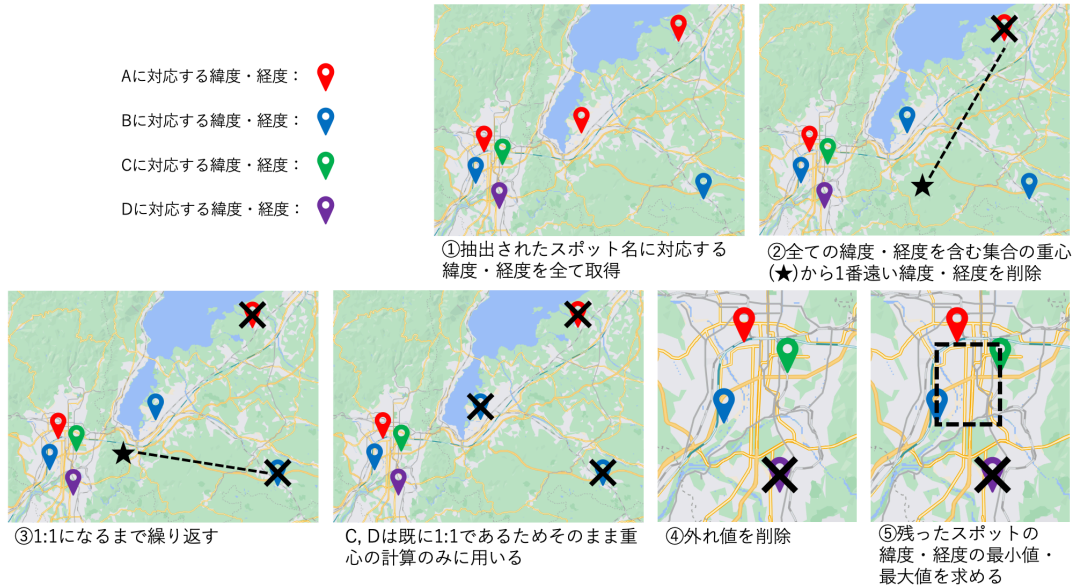


図 2 緯度・経度の決定手順

表 1 観光スポット名と対応する複数の緯度・経度の例

観光スポット名	県 (緯度, 経度)
法住寺	愛知県 (34.8193, 137.301), 京都府 (34.9875, 135.773)
南禅寺	静岡県 (34.7454, 138.980), 京都府 (35.0112, 135.795)
清水寺	青森県 (40.4718, 141.491), 岩手県 (39.3701, 141.030)

した Web ページのテキストには明示的な位置情報は含まれていないため、MBR を推定することで、ユーザが観光情報取得しようとしているエリアを推定できると考えた。手順を図 2 に示す。まず、ユーザが閲覧した Web ページのテキストから、観光スポット名を抽出する。抽出には、観光スポット名で作成した辞書を追加した Janome²を用いる。抽出された観光スポット名のうち、英字の観光スポット名以外の、3 文字以上の漢字のみの観光スポット名と、4 文字以上の観光スポット名を使用した。そして、抽出された観光スポット名に部分一致する緯度・経度を取得する。これは、より正確にユーザが観光情報を得ようとしているエリアを推定するために、可能性のある候補を得るためである。また、部分一致のスポットに加え、そもそも異なる地域にも同じ名前の観光スポットが存在するため、あるスポット名に対し、複数のスポットが候補として抽出される。表 1 に、観光スポット名と対応する複数の候補スポットの緯度・経度の一部の例を示す。ここで、北山ら [10] の判定方法にならない、観光スポット名とそれに対応する緯度・経度を 1 対 1 にする処理を行う。抽出された観光スポット名に対応する複数の緯度・経度の組からなる集合の重心を求める。求めた重心から最も遠い緯度・経度の組を除外する。この操作を各観光スポット名と対応する緯度・経度が 1 対 1 になるまで繰り返す。

その後、スミルノフ・グラブス検定を行い、外れ値を除いた集合から緯度・経度の最小値・最大値を決定する。スミルノフ・グラブス検定とは、データが正規分布に従うとき、測定されたデータの外れ値を検出する手法である。この処理により得られ

たスポット集合により MBR を決定する。ユーザが観光情報取得しようとして実際に閲覧しているエリアはこの MBR より少し広いと予想されるため、1.3 倍に拡張したエリアをユーザが観光情報取得しようとしているエリアと推定する。

3.3 嗜好との合致度

本研究での嗜好との合致度とは、推薦候補の観光スポットとユーザの嗜好との一致度を測るものである。ユーザが過去に訪れた観光スポットの説明文とレビュー文から抽出された名詞は、ユーザの観光スポットの嗜好の傾向を表すものであると推定する。これを推薦候補の観光スポットの説明文とレビュー文中の名詞と比較することで、ユーザの嗜好と類似する観光スポットを推薦できると考える。まず、観光スポットの説明文とレビュー文の tf-idf ベクトルを生成する。全ての観光スポットの説明文とレビュー文を総文書とし、説明文とレビュー文中の名詞の tf-idf 値を計算する。tf-idf とは、文書に含まれる単語の重要度を評価する手法である。tf-idf 値は以下の式 (2)(3)(4) より得られる。

$$tf_{i,j} = \frac{n_{i,j}}{\sum_{s \in j} n_{s,j}} \quad (2)$$

$$idf_i = \log \frac{N}{df(i)} \quad (3)$$

$$tfidf_{i,j} = tf_{i,j} \cdot idf_i \quad (4)$$

式 (2) について、 $tf_{i,j}$ は文書 j 内のある単語 i の tf 値を表す。 $n_{i,j}$ は文書 j 内のある単語 i の出現回数、 $\sum_{s \in j} n_{s,j}$ は文書 j 内のすべての単語の出現回数の和である。式 (3) について、 idf_i はある単語 i の idf 値を表す。 N は総文書数、 $df(i)$ は単語 i が出現する文書の数である。式 (4) について、式 (2)(3) より、文書 j 内のある単語 i の tf-idf 値は $tf_{i,j}$ と idf_i を掛け合わせた値である。

観光スポットごとに計算した tf-idf 値をその観光スポットのベクトルの要素とし、過去に訪れた観光スポットのベクトルと

2 : <https://mocabeta.github.io/janome/>

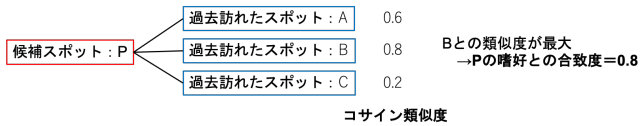


図3 嗜好との合致度の決定方法

推薦候補の観光スポットのベクトルのコサイン類似度を計算する。コサイン類似度とは、ベクトル同士の成す角度の大きさを表すものであり、1に近いほど類似しており、0に近いほど類似していないことになる。嗜好との合致度は以下の式(5)より得られる。

$$sim(c, H) = \max \left(\frac{c \cdot h}{\|c\| \|h\|}, h \in H \right) \quad (5)$$

H はユーザが過去に訪れた観光スポット集合である。ある推薦候補の観光スポット c に関して過去に訪れた観光スポット h のコサイン類似度をすべて計算し、コサイン類似度の最大値をその推薦候補の観光スポットの嗜好との合致度とする。嗜好との合致度の決定方法を図3に示す。推薦候補の観光スポット P と過去に訪れた観光スポット A, B, C があつたとき、 P に対し A, B, C それぞれのコサイン類似度を計算する。そのうち、 P と B のコサイン類似度が最大であつたとき、この値を推薦候補の観光スポット P の嗜好との合致度とする。

3.4 非接触度

本研究での非接触度とは、ユーザが推薦候補の観光スポットと接触する難度を表す値であり、ユーザごとの接触の難しさと一般的な接触の難しさを掛け合わせて算出する。

ユーザごとの接触の難しさとは、ユーザがその観光スポットを目にする機会の少なさと同義である。そこで、ユーザが閲覧したページ中にその観光スポット名が出現した回数を用いる。このとき、Web ページ中にスポット名が完全な形で出現するとは限らない。例えば、「東京国立近代美術館」は、「近代美術館」とだけ書かれることがある。そのため、Web ページ中の単語がスポット名に部分一致するとき、エリア内の部分一致スポット数の逆数回出現したとみなしている。出現数の数え方を図4に示す。接触の難しさを表す値として出現数の逆数で表現する。

また、一般的な接触の難しさを用いる理由は、ある情報と接触できる要因にはユーザが意識して発見しようとする行動の他に、その情報の知名度もあると考えたためである。そこで、服部ら [11] の研究にならい、観光スポットの知名度が高いほどレビューの数も多いとして、一般的な接触の難しさを表す値にレビュー数の対数の逆数を用いることにした。レビューの数が少ないほど逆数は大きくなるため、接触の難しさを表す値も大きくなる。また、対数を用いる理由は、レビュー数が極端に少ない場合や多い場合に、一般的な接触の難しさを表す値の影響が大きくなることを防ぐためである。

例として、「東京タワー大展望台」の非接触度を説明する。「東京タワー大展望台」は、東京の観光情報を探す人であればよく接触する観光スポットであり、ユーザごとの接触の難しさを表す値は小さくなる。さらに、レビュー数は1875件あり、全て

対象スポット：東京国立近代美術館

Webページ中の完全一致する単語と部分一致する単語		
単語	出現数	部分一致スポット数
東京国立近代美術館	2	0
東京	204	29
近代	1	1
美術館	4	6
近代美術館	1	1

東京国立近代美術館の出現数
 = 完全一致単語出現数 + 部分一致単語出現数 / 部分一致スポット数
 = 2 + 204/29 + 1/1 + 4/6 + 1/1 = 11.7

図4 非接触度の決定方法

の観光スポットのレビュー数の平均19.1件と比べても多く、一般的な接触の難しさを表す値も小さくなる。したがって、「東京タワー大展望台」は東京の観光情報を探す人にとって、非接触度の低いスポットといえる。

非接触度は、以下の式(6)(7)(8)より得られる。

$$U_c = \frac{1}{f_c + 1} \quad (6)$$

$$G_c = \frac{1}{\log_e(r_c + e)} \quad (7)$$

$$con(c) = U_c \times G_c \quad (8)$$

式(6)について、 f_c はユーザが閲覧したWeb ページ中の観光スポット c の出現数であり、 U_c は観光スポット c のユーザごとの接触の難しさを表す。式(7)について、 r_c は観光スポット c のレビュー数であり、 G_c は観光スポット c の一般的な接触の難しさを表す。対数の計算には、自然対数の底 e を用いる。式(8)について、 $con(c)$ は推薦候補の観光スポット c の非接触度を表す。

4 予備検討

4.1 嗜好との合致度に関する検討

嗜好との合致度の計算の精度を高めるため、評価実験の前に名詞抽出に関する検討を行った。

本研究では、ユーザの嗜好や観光スポットを特徴付けるものとして、観光スポットの説明文とレビュー文から抽出した名詞を用いる。そこで本実験の前に、観光スポットの説明文とレビュー文から名詞を抽出する精度を高めるため、不必要な品詞の単語を除外するフィルタを作成した。フィルタを作成するため、名詞のみを指定して抽出した結果から、不必要な単語の品詞を確認した。品詞の確認にはPythonの形態素解析器であるJanomeを用いた。

名詞のみの指定で観光地の説明文から名詞を抽出した結果を、表2に示す。表2より数字、助動詞、観光地とは全く関係のなさそうなものを人手で選び、品詞を調べた。選んだ単語の例とその品詞を表3に示す。表3の結果を「指定品詞」とし、「指定品詞」の単語を除外して名詞を抽出した結果、表2の下線の引かれた名詞が抽出されなくなった。

表2の「春日大社」の結果より、「1863」、「7」、「棟」、「3」、

表 2 「名詞」と指定して抽出した結果

観光スポット	抽出された名詞
春日大社	春日大社 1863 文久 春日 典型 形式 7 造 建造 本殿 棟 重要 文化財 3 的 指定 年
若宮神社	燈籠 おん 春日大社 榎社 若宮 本社 庶民 遺構 2 両側 参道 奈良 祭り 祭 平安 石 有名 時代
夫婦大國社	春日大社 末社 大國 夫婦 唯一 様 日本 1
東大寺	那仏始 盧 若草山 巨刹 古都 資産 大仏 域 舎 め 重文 国宝遺産 うち 奈良 麓 級 広大 群 ひとつ 世界 8 寺 文化財
奈良国立博物館	院展 仏教 倉 本館 実施 美術 博物館 毎年 重要 秋 中心 文化財 指定 展示
猿沢池	うねめ 管絃 猿沢 ミス 采女 七草 雅楽 中秋 名月 興福寺 使 扇 隻 演奏 華やか あまり 最後 坂 行事 南側 なか 奈良 祭 景色 塔 船 池 秋 はじめ 花 m 日 2 中
元興寺	智光 西小 飛鳥寺 南都七大寺 僧房 元興寺 古都 資産 東大 曼荼羅 大寺 極楽 五重塔 小 改築 寺宝 坊 重 遺産 うち 奈良 院 五 群 塔 ひとつ 創建 史跡 一つ 世界 8 跡 境内 国 現在 文化財 もの 指定
興福寺	古都 資産 出入り 五重塔 天平 寺宝 数々 遺産 うち 奈良 自由 彫刻 群 ひとつ 創建 見事 建物 世界 8 はじめ 時 境内 文化財
興福寺国宝館	1250 興福寺 34 天平 仏像 絵画 重文 国宝 様式 堂 奈良 伝統 建築 江戸 昭和 歴史 もの 指定 時代 年

表 3 不要な単語とその品詞の例

品詞	単語
名詞, 数	1863 五
名詞, 副詞可能	ひとつ はじめ
名詞, 接尾	年 時め (始め) あまり (2m あまりの)
名詞, 非自立	日 うち なか もの

“年” が抽出されなくなった。新しく設定したフィルタで観光スポットを特徴付けるような名詞のみを抽出できるようになった。この結果を用い、ユーザの観光履歴中の観光スポットとユーザの指定したエリア内の観光スポットの説明文とレビュー文から名詞を抽出する。

4.2 非接触度の重み α に関する検討

式 (1) の α の値により、推薦結果は「発見が困難な観光スポット」から「個人の嗜好に合うスポット」まで変化する。そこで、非接触度の重み α に関する検討を行った。

式 (1) について、 α の値を 0 (嗜好との合致度のみによる推薦) から 1 まで 0.1 ずつ変更しながら結果を比較し、結果の上位に 10 件に含まれる嗜好との合致度による推薦結果の割合が 50% になる値を確認した。50% とした理由は嗜好との合致度のみによる推薦結果にも、「個人の穴場スポット」を満たすスポットがあると考えられるためである。

表 4 に、式 (1) の α の値を 0.1 ずつ変更しながら出力した結果に、嗜好との合致度のみによる推薦結果が占める割合を示す。入力に用いた過去に訪れた観光スポットと、次の旅行のため閲覧した Web ページの組み合わせは 3 組用意した。

結果より、1 組目と 2 組目が 50% になる 0.3 を α として採用した。

表 4 嗜好との合致度のみによる推薦結果が占める割合

α	1 組目	2 組目	3 組目
0.0	100%	100%	100%
0.1	70%	90%	70%
0.2	50%	60%	40%
0.3	50%	50%	30%
0.4	20%	40%	0%
0.5	10%	30%	0%
0.6	0%	30%	0%
0.7	0%	30%	0%
0.8	0%	20%	0%
0.9	0%	10%	0%
1.0	0%	10%	0%

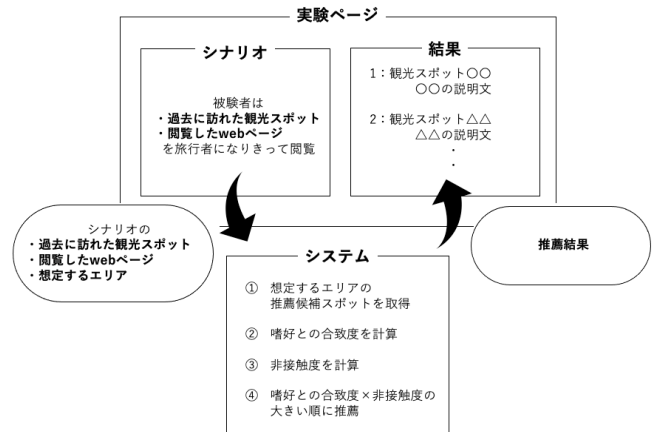


図 5 提案手法を用いた実験の概要

5 評価実験

5.1 実験方法

提案手法の嗜好との合致度に対応する比較尺度として、一般的な評価 (星の数)、非接触度に対応する比較尺度として一般的な接触の難しさ (式 (7)) を用意する。これらの組み合わせによる 6 手法を比較する。

提案手法を用いた実験の概要を、図 5 に示す。本実験では、あらかじめ用意したシナリオを被験者が行ったことのあるスポットと、これから旅行に行くとして閲覧した Web ページだと想定してもらい、被験者は出力結果のみを評価する形で実験を行った。シナリオは、本実験の被験者とは異なる被験者より取得した、過去に訪れた観光スポットと閲覧した Web ページを用いて作成した。なお、この実験の主眼は、推薦されたスポットのスコアの評価を行うことであるため、エリア推定は人手で行っている。

一般的な評価を用いた観光スポット推薦 (以下比較手法 1) では、被験者が閲覧した Web ページから推定したエリア内のレビュー数が 5 以上の観光スポットに関して、じゃらん net より取得したレビューの投稿者がそのスポットを評価して付けた星の数の多い順に推薦する。一般的な評価と一般的な接触の難しさをういた観光スポット推薦 (以下比較手法 2) では、星の数と式 (7) を掛け合わせた結果を用いる。一般的な評価と非接触度をういた観光スポット推薦 (以下比較手法 3) では、星の

表 5 推薦手法の表すスポットと用いる尺度

手法	表すスポット	嗜好との合致度	星の数	非接触度	一般的な接触の難しさ
比較手法 1	一般的な評価が高い		○		
比較手法 2	一般的な穴場		○		○
比較手法 3	一般的な評価が高いが自分では見つけにくい		○	○	
比較手法 4	嗜好に合う	○			
比較手法 5	嗜好に合うが一般的に見つけにくい	○			○
提案手法	個人の穴場	○		○	

表 6 6 手法による推薦の被験者全員の評価の平均値

	観点 a	観点 b	観点 c	観点 d
比較手法 1	2.86	4.31	2.51	3.73
比較手法 2	2.77	4.36	2.41	3.78
比較手法 3	3.06	4.31	2.76	3.72
比較手法 4	3.95	2.74	3.56	2.42
比較手法 5	3.61	3.61	3.27	3.02
提案手法	3.58	3.38	3.20	2.86

数と式 (1) の第 2 項を掛け合わせた結果を用いる。嗜好との合致度を用いた観光スポット推薦（以下比較手法 4）では、式 (1) の第 1 項のみを用いて推薦を行う。嗜好との合致度と一般的な接触の難しさを用いた観光スポット推薦（以下比較手法 5）では、式 (1) の第 1 項と式 (7) を掛け合わせた結果を用いる。全ての手法でレビュー数が 5 以上のスポットを用いている。表 5 に、それぞれの推薦手法の表すスポットと用いる尺度を示す。非接触度の α は 0.3 であるが、一般的な接触の難しさの α は 1.0 を用いている。

提案手法、比較手法 1、比較手法 2、比較手法 3、比較手法 4、比較手法 5 からそれぞれ上位 5 件の観光スポットを推薦する。なお、シナリオ 2 の比較手法 2 のみ、実験上のミスで推薦件数が 4 件となっている。推薦された観光スポットについて、「個人の穴場スポット」を満たすものとして以下の観点を 5 段階のリッカート尺度（1:「そう思わない」、2:「あまりそう思わない」、3:「どちらとも言えない」、4:「ややそう思う」、5:「そう思う」）で評価してもらう。

- 観点 a この観光スポットを訪問してみたい
- 観点 b この観光スポットを知らなかった
- 観点 c この観光スポットは探していたものに合致する
- 観点 d この観光スポットは自分で見つけることは困難である

シナリオは 3 種類用意し、各推薦スポットにつき 50 名が評価し、その平均を用いる。被験者はクラウドソーシングで集めたのべ 3600 人を対象とした。

5.2 実験結果と考察

まず、3 種類のシナリオから推薦されたスポットの全ての評価の平均値を手法ごとに算出した結果を表 6 に示す。比較手法 5 はどの観点でも評価の平均値が 3 を超え、6 つの手法の中で推薦結果が最も「個人の穴場スポット」に近いものであった。

比較手法 5 と提案手法の結果を比較する。比較手法 5 の一般的な接触の難しさの重みと提案手法の非接触度の重みは違うため、提案手法の非接触度の重みを 1 にした結果を比較手法 5 の結果と比較した。どのシナリオでも結果が大きく異なることがわかった。表 7 にシナリオ 1 の比較手法 5 の推薦結果、提

表 7 重みごとの推薦結果と U_c の値

順位	比較手法 5		提案手法 (重み: 0.3)		提案手法 (重み: 1)	
	スポット	U_c	スポット	U_c	スポット	U_c
1	安井金比羅宮	0.09	知恩院	0.30	六波羅蜜寺	0.91
2	京都きものレンタル麗	0.01	六波羅蜜寺	0.91	酢屋	0.88
3	市比賣神社	0.05	京都水族館	0.08	妙蓮寺	0.63
4	京都水族館	0.08	元離宮二条城	0.06	智積院	0.43
5	下鴨神社	0.02	壬生寺	0.39	拾得	0.5

案手法の推薦結果、提案手法の重みを 1 にした場合の推薦結果を示す。結果から、比較手法 5 には用いず、提案手法では用いるユーザごとの接触の難しさ（式 (6) の U_c ）の影響が考えられる。 U_c の値から、ユーザごとの接触の難しさを用いると U_c の値の大きいスポット、つまりユーザが閲覧した回数が少ないスポットが上位に含まれやすくなることがわかった。したがって、ユーザごとの接触の難しさを用いると「ユーザが知らないスポット」を推薦しやすくなるが、提案手法では重みがかかることにより、「ユーザが知っているスポット」が上位に推薦されやすくなり、提案手法の評価が低くなったと考えられる。

今回、0.3 という重みを 0 のときから推薦結果が 50% 変化する点で決定したが、より適切な値にすることで、ユーザの知らない興味に合致するスポットを推薦する必要がある。

また、 U_c の計算でスポット名に部分一致する単語を考慮しているが、「東京」や「京都」などの多くのスポット名に含まれる単語が結果に効きすぎていることも考えられる。そのため、不必要に U_c の値が下がっているケースも考えられ、この数え方も改善が必要である。

6 まとめ

本稿では、個人の嗜好に合うにも関わらず、発見が困難な観光スポットを「個人の穴場スポット」と定義し、各推薦候補の観光スポットの嗜好との合致度と非接触度を計算し、個人の穴場スポットを推薦する手法を提案した。また、本稿では以下の 2 つの研究課題の解決を目指した。

- RQ1 非接触度を用いることで、ユーザが個人的に知らないスポットを抽出可能か。
- RQ2 嗜好のみを用いるよりも非接触度を用いることは、推薦の性能に寄与するか。

過去に訪れた観光スポットと推薦候補の観光スポットの類似度と、ユーザが推薦候補の観光スポットを目にするの難しさに着目することで、ユーザがまだ発見できていないが、嗜好に合う観光スポットを推薦できると考えた。

実験の結果、RQ1 に関して、非接触度によって、ユーザが知らないスポットを推薦できる可能性があることを確認した。しかし、重み α の設定によっては、ユーザが検索中に接触している可能性が高いスポットを推薦しており、 α の調整が必要である。

RQ2 に関して、非接触度の一部である一般的な接触の難しさを用いることで、推薦の質が良くなることが示唆された。ユーザにとっての接触の難しさを加えた場合の性能について、適切

な重み α を決定した後に実験により評価する必要がある。

謝 辞

本研究の一部は、2021 年度科研費基盤研究 (C)(課題番号：21K12147) によるものです。ここに記して謝意を表すものとします。

文 献

- [1] 市村哲, 上石萌恵, 堀口莉里花. 口コミ解析と好み診断に基づいた旅行先推薦. 研究報告グループウェアとネットワークサービス (GN), Vol. 2020-GN-110, No. 11, pp. 1–8, mar 2020.
- [2] 小南佑介, 赤石美奈. 感性的嗜好に基づく旅行先推薦システムの構築. 第 2020 巻, pp. 381–382, feb 2020.
- [3] 奥園基, 牟田将史, 平野廣美, 益子宗, 星野准一. 複数人での旅行における嗜好分析による観光地推薦システムの提案. 研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2015-HCI-162, No. 19, pp. 1–8, mar 2015.
- [4] Daisuke Kitayama. Extraction method for anaba spots based on name recognition and user's evaluation. *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services, ii-WAS 2016, Singapore, November 28-30, 2016*, pp. 12–15, Nov 2016.
- [5] 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観光地発見. 第 79 回全国大会講演論文集, Vol. 2017, No. 1, pp. 523–524, mar 2017.
- [6] 渡邊小百合, 吉野孝. 観光情報可視化システムのための web 上における観光特徴抽出手法の提案. 第 78 回全国大会講演論文集, Vol. 2016, No. 1, pp. 43–44, mar 2016.
- [7] 渡邊小百合, 吉野孝. 観光地間の類似性を基にした向上点発見のための観光情報可視化システム. マルチメディア, 分散協調とモバイルシンポジウム 2016 論文集, Vol. 2016, pp. 1357–1362, jul 2016.
- [8] 廣嶋伸章, 安田宜仁, 藤田尚樹, 片岡良治. 地理情報検索におけるクエリ入力支援のための特徴語の提示. 人工知能学会全国大会論文集, Vol. JSAI2012, No. 26, jun 2012.
- [9] 松本敦志, 杉本徹. クチコミから抽出した特徴語を利用する観光地検索支援. 第 75 回全国大会講演論文集, Vol. 2013, No. 1, pp. 307–308, mar 2013.
- [10] 北山大輔, 李龍, 角谷和俊. 地理的正確性と空間的コンテキストに基づくデフォルメ地図分析. 情報処理学会論文誌データベース (TOD), Vol. 4, No. 2, pp. 172–184, jul 2011.
- [11] 服部裕基, 灘本明代. Wikipedia を用いたマイナースポーツ検索手法の提案. 研究報告 データベースシステム (DBS), Vol. 2011-DBS-152, No. 11, pp. 1–7, jul 2011.