

ユーザの嗜好に基づく観光スポット説明文の個人化手法

山田 祥輝[†] 北山 大輔[†]

[†] 工学院大学情報学部 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: tj115105@ns.kogakuin.ac.jp, tkitayama@cc.kogakuin.ac.jp

あらまし 近年、観光スポットを選択する際にじゃらんや TripAdvisor といった Web 上の観光情報サイトを利用することが多くなっている。これらのサイトには、実際に観光スポットを訪れたユーザの体験に基づくレビューが数多く掲載されており、観光スポット選択を行うユーザにとって有用である。しかし、膨大な情報から観光スポット選択に有用な情報を見つけることは容易ではない。したがって、観光スポット選択を行うユーザは興味のある情報を見落とす可能性がある。そこで本研究では、観光スポットの訪問履歴からユーザの嗜好を推定し、その嗜好に合致する観光スポットのレビューを提示する手法を提案する。観光スポットには複数の要素があることに着目し、ユーザの訪問済み観光スポットのレビューをクラスタリングすることによって、嗜好を推定する。また、プロトタイプシステムを構築し、被験者を用いた提案手法によるレビュー提示の評価実験を行う。

キーワード 観光スポット, レビュー, 個人化, 階層的クラスタリング, 分散表現

1 はじめに

近年、じゃらん¹や TripAdvisor²といった観光情報サイトの普及に伴い、Web 上の観光情報サイトを利用する機会が多くなっている。その背景には、ライフスタイルの多様化や個人の嗜好の細分化により、観光の主体が団体から個人に変化し、旅行者自らが観光スポットの選定をする機会が増えていることがあげられる [1]。観光情報サイトは、そうした個人観光客の観光計画を立てる判断材料として利用されている。

観光情報サイトには、実際に観光スポットを訪れたユーザの体験に基づくレビューが数多く掲載されている。その中には、観光スポットの公式 Web サイトにはない貴重な情報や、ユーザの興味を引くような情報が含まれており、ユーザが観光スポットを選択する際に有用である。しかし、膨大なレビューの中からそのような情報を見つけることは容易ではない。観光情報サイトに掲載されているレビューは、評価の高い順やレビューの投稿日順などで整理されているが、これらはユーザの嗜好に必ずしも合致しているとはいえず、ユーザの嗜好に合った情報や興味のある情報を見落としてしまう可能性がある。

本研究では、ユーザの観光スポット訪問履歴から観光スポットに対する嗜好を推定し、その嗜好に合致するレビューを観光スポットの説明文として提示する手法を提案する。図 1 は提案手法の概要を示したものである。まず、検索ユーザの観光スポット訪問履歴から、各訪問スポットに付随する個々のレビューを全て取得する。次に、取得したレビュー群に対し階層的クラスタリングを行い、各クラスタに対しユーザの嗜好を表現してスコア付けを行う。その後、ユーザの検索スポットのレビューベクトルと、スコアの高いクラスタの重心ベクトルとの類似度を計算する。最後に、類似度の最も高いレビューを

ユーザの嗜好に合致した文として、検索ユーザに提示する。

本稿の構成を以下に述べる。2 節ではユーザの嗜好に基づく推薦や検索に関する研究の紹介を行う。3 節では Doc2Vec によるレビュー文の分散表現の適応とその有効性について述べる。4 節では提案手法の概要について述べる。5 節では提案手法を用いたユーザの嗜好抽出に関する予備実験について述べる。6 節では提案手法によるユーザの嗜好に基づくレビュー提示の評価について述べる。最後に、7 節でまとめと今後の課題について述べる。

2 関連研究

2.1 ユーザの嗜好に基づく観光スポット検索・推薦

ユーザの嗜好を利用した地理情報の検索および推薦に関する研究は広く行われている。開地ら [2] は、単語の分散表現を用いてユーザの観光に対する潜在的興味を推定し、観光スポットを推薦する研究を行った。潜在的興味はユーザの好む要素と嫌う要素の加減算によって表現できると仮定し、Word2Vec を用いて作成した単語ベクトルの加減算によりユーザの潜在的興味を推定した。本研究とはユーザの観光に対する興味をベクトル空間モデルで表現している点で類似しているが、ユーザによる明示的な単語の加減算では推定することが難しいより潜在的な嗜好を推定するため、レビュー文そのものに Doc2Vec を用いて特徴ベクトルを作成している点で異なる。

阪井ら [3] は、観光スポットの耳より情報を提示する研究を行った。この研究では、観光スポットに対する有益かつ意外な情報を耳より情報と定義し、検索スポットのレビューをクラスタリングすることによりレビューを分類することで耳より情報を抽出した。また、作成したクラスタの中心から閾値以上離れたレビューを意外な情報と定義し、ユーザにとって意外かつ耳よりな情報を提示する手法を提案した。本研究とはクラスタリングによって観光スポットのレビューを分類し、ユーザに有益

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

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

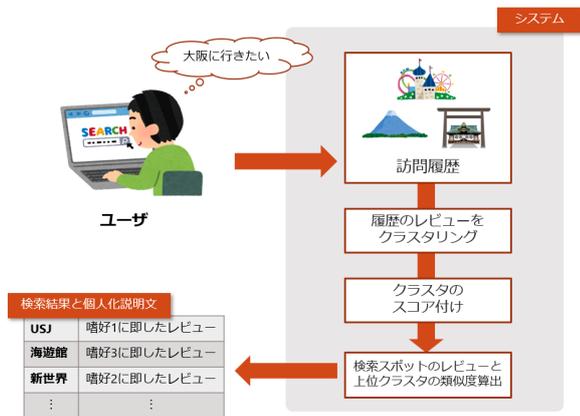


図1 ユーザの嗜好に基づく観光スポット説明文の個人化手法

な情報を提示する点で類似するが、本研究では検索ユーザの嗜好と観光スポットの特徴との一致度が高い情報が検索ユーザの観光スポット選択に有用であると考えたため、意外性を明示的に考慮していない点で異なる。

2.2 観光スポットに対するユーザの嗜好推定手法

観光スポットに対するユーザの嗜好の推定に関しても、様々な研究が行われている。鈴木ら[4]は、個人の嗜好と訪問地域の魅力に基づく観光地推薦を行った。この研究では、観光嗜好の構造を観光動機論に基づき構築した。観光嗜好の構造は、「個人嗜好を考慮した観光資源の選定」と「観光動機」の二段階となっており、「観光動機」は「観光動機」と「観光目的」に分かれていると定義した。その後、AHP（階層分析法）を用いてユーザの個人嗜好を推定している。日高[5]は観光に対する個人の嗜好をAHPとコンジョイント分析を組み合わせることで、AHPによるユーザの負担を軽減しつつ、個人の嗜好を高精度に取得する手法を提案した。しかしながら、一般的にAHPを行う場合、一対比較表の作成や回答に時間や手間がかかるという問題がある。本研究では、訪問済み観光スポットを入力するという簡単な作業のみで、検索ユーザの観光スポットに対する嗜好を推定する。

奥蘭ら[6]は複数人の嗜好を反映させた観光スポット推薦システムを提案した。画像選択により個人の嗜好モデルを作成し、見解間距離均等法によって集団の嗜好を統合した。結果として、5人以下の嗜好モデルであれば1個人の嗜好に基づく観光スポット推薦と同程度の推薦ができることを示した。

川崎ら[7]は、ユーザの価値観を意味の数学モデルによってモデル化し、知的価値を有する観光地を推薦する手法を提案した。

上原ら[8]は、Web上にある複数の要素から特徴ベクトルを作成し、観光地間の類似度を図ることで観光地を推薦するシステムを提案した。観光地の特徴ベクトル生成には、Yahoo知恵袋³、ブログ上での共起キーワードと時系列分布、知恵袋上でのカテゴリ構造、観光地周辺施設、地図画像を利用している。

2.3 ユーザの嗜好に基づく観光スポット以外の検索・推薦

地理情報以外でのユーザの嗜好を利用した情報検索、推薦に関する研究も広く行われている。中岡ら[9]はジオタグツイート、言語、評判の3つの情報から国民の嗜好性を抽出し、店舗を推薦する手法を提案した。

Wangら[10]は、Twitterユーザがフォローしているユーザの情報を元に、LDAトピックモデルを利用してユーザの嗜好を推定する手法を提案した。本研究とは、ユーザの履歴情報がユーザの嗜好を表すと仮定し重み付けを行っている点で類似しているが、ユーザに提示する情報が異なる。

情報推薦における推薦理由の提示が、ユーザの行動に与える影響に関する研究も行われている。関根ら[11]は、コンテンツの推薦理由の明示化によるユーザの満足度評価を行った。その結果、推薦理由の存在が、ユーザのコンテンツ選択に影響を与える可能性が大きいことを示した。

このように、ユーザの嗜好を用いた研究ではユーザの嗜好を測る入力から、コンテンツ自体を提示することが多い。しかし、観光スポットのような1つのコンテンツに複数の要素がある場合、どの要素がユーザの嗜好と合致しているのかを推定している研究は少ない。本研究ではその点に着目し、ユーザの嗜好に即したレビューを提示することで複数の要素からユーザの好む適切な要素を推定することを目的とする。

3 Doc2Vecによるレビュー文の分散表現の適用

3.1 Doc2Vecによるレビュー文のベクトル化

本研究では、ユーザの嗜好を表すベクトルと観光スポットのレビューベクトルを比較し、ユーザの嗜好に類似するレビュー文を出力する必要がある。そのため、文章の類似度を比較するためにDoc2Vecを用いる。Doc2Vecとは任意の長さの文章をベクトル化する技術であり、文やテキストに対して分散表現を獲得することができる。これにより、観光スポットのレビューを300次元のベクトルで表現する。学習データの生成とモデルの学習には、国内でも代表的な観光情報サイトであるじゃらん⁴に投稿された1,481,838レビューを使用し、Paragraph Vectorモデル[12]を用いてベクトル生成を行う。実際のレビューベクトルの作成にはgensim⁴を利用する。また、3.2節で行った検討により、L2ノルムが1.95未満のレビューベクトルはユーザに提示しないこととする。

3.2 レビューベクトルによる類似文の例

3.1節で作成したベクトルが適切に文章を表現しているか確かめるために、例を用いて議論する。表1は上記で作成したレビューベクトルを用いた、上野動物園に関するレビューとコサイン類似度が高いレビュー上位10件である。比較元となる上野動物園のレビュー文を見ると、「動物園」、「パンダ」、「動物を見る」といった内容で構成されていることがわかる。上野動物園のレビューと類似度が高いレビューに関しても、そのような

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

4 : <https://radimrehurek.com/gensim/index.html>

表 1 上野動物園のレビューと類似度が高いレビューとノルムの関係

レビュー文	類似度	L2 ノルム
東京上野にある、上野動物園に行きました。パンダだけでなく、動物たちを楽しく見られる工夫がされています。	-	2.14
動物たちが、生き生きしていた。	0.800	1.47
全国的にも有名な動物園です。行動展示の先駆けで工夫した展示がされていて動物園たちの生き生きした姿が見れました	0.787	2.40
園内はとても広くて見応えがありました。動物だけでなく植物も見れて楽しいです。ゴリラを見て子供がすごく喜んでいました。	0.774	2.66
パンダが見れなかった。残念です。	0.767	1.64
カワウソがたくさん飼育されているのですが、とても活発で見ている楽しかったです。珍しいヤマネコもいて楽しめました。…	0.757	2.48
象のショーがとても楽しく、立ち上がったときは面白かったです。象以外にもライオンなどの動物が飼育されています。	0.751	2.48
パンダがいなくて残念。	0.750	1.45
動物が自然体で飼育されている場所です。バスで動物を見て回れるので動いている姿を間近で見れて面白かったです。	0.747	2.65
パンダが見たくて行きました。パンダは寝ていて動かなかったけどかわいかったです。他の動物もたくさん見れて楽しかったです。	0.745	2.61
パンダかわいかった。	0.743	1.31

内容を含むレビューの類似度が高くなるという結果になった。これにより、文章の類似度に基づくレビューベクトルを生成できていると考えられる。

また、観光スポットのレビューには「写真のみ」や「大変美しかった」といった内容が乏しいレビューが存在する。表 1 を見ても、「パンダかわいかった」というレビューや「動物たちが、生き生きしていた。」という内容に乏しいレビューが存在する。本研究では、クラスタの重心ベクトルと検索スポットのレビューベクトルとの類似度のみを利用して、ユーザにレビューを提示する関係上、内容が乏しいレビューもユーザに提示する可能性がある。しかし、このような内容が乏しいレビューはユーザの嗜好には合致しているかもしれないが、ユーザにとって有用であるとは考えにくい。

そこで、検索スポットのレビューを提示する際に、レビューベクトルの L2 ノルムが 1.95 未満⁵のレビューを取り除くことにより、内容が乏しいレビューをユーザに提示されないようにすることが有効であると考えられる。他のレビュー文に対しても同様の実験を行い、同様な傾向であることを確認している。

4 ユーザの嗜好に合致するレビューの選定方法

本研究では、観光スポットを検索する際に、観光スポットの訪問履歴からユーザの観光スポットに対する嗜好を推定し、嗜好に合致する検索スポットのレビューを提示する手法を提案する。まず、ユーザの観光スポット訪問履歴から各訪問スポットに付随する個々のレビューを全て取得し、階層的クラスタリングによってレビューをクラスタに分類する。次に、各クラスタに対しユーザの嗜好を表現しているかどうかのスコア付けを行う。その後、スコア上位のクラスタの重心ベクトルと検索対象の観光スポットのレビューベクトルを比較し類似度を求める。最後に、クラスタの重心ベクトルと検索スポットのレビューベクトルとの類似度、クラスタのスコア、検索スポットのレビューベクトルの L2 ノルムによりスコア付けを行い、最もスコアの高いレビューをユーザの嗜好に合致する文としてユーザに提示する。

5：全レビューのノルムの平均が 3.03、標準偏差が 0.54 のため、平均と標準偏差の二倍の差である 1.95 を採用した。

4.1 観光スポットに対するユーザの嗜好の分類

本研究では、ユーザが訪れた観光スポットにはユーザの観光スポットに対する嗜好が含まれていると仮定する。しかし、観光スポットには多くの場合、単一の要素ではなく複数の要素が存在する。たとえば、東京都にある浅草寺には雷門や本堂における歴史的な要素や、仲見世通りにおけるグルメに関する要素などの、様々な要素が存在すると考えられる。そのため、ユーザの観光スポット訪問履歴には単一の嗜好ではなく複数の嗜好が入り混じっていると考えられる。同様に、観光情報サイトにおいても観光客の観光スポットに対する感じ方によってレビューの内容は多岐にわたる。

そこで、ユーザが訪れた観光スポットに付随する全てのレビューをクラスタリングすることにより、レビューを観光スポットの持つ特徴ごとに分類する。レビューの分類には階層的クラスタリングを利用する。クラスタ間の距離の測定には、レビューベクトルのコサイン類似度による群平均法を用いる。群平均法とは、2つのクラスタに属する要素間の全ての組み合わせの距離を求め、その平均値をクラスタ間の距離とする手法である。クラスタ分けの閾値は、事前の実験により 0.65 とする。

4.2 ユーザの嗜好の表現度によるクラスタのスコア付け

階層的クラスタリングによって分類したクラスタに対し、どのクラスタがユーザの嗜好を表しているかをスコア付けする。本研究では、以下の条件を満たすクラスタをユーザの嗜好を表すクラスタと定義する。

- ユーザの各訪問済みスポットのレビューを満遍なく含んでいる

- より多くのスポットのレビューで構成されている

たとえば、クラスタを構成するレビューの大半がある単一のスポットのレビューで構成されているクラスタがあると仮定する。そのクラスタはユーザの嗜好というより、ある単一のスポットの特徴を表していると考えられる。一方で、複数のスポットのレビューを満遍に含んでいるクラスタは、単一のスポットの特徴の範囲を超え、ユーザの観光に対する嗜好を表現していると考えられる。

本研究では、クラスタを構成する訪問スポットごとのレビュー数の割合を算出するため、ジニ分散指標を用いる。ジニ分散指標とは各クラスの分散の総和を表す指標であり、クラスのばら

つきが小さいと値は大きくなり、ばらつきが大きいほど値は小さくなる。履歴のスポット数を H 、クラスタ t に所属するレビューのうち、スポット i のレビューの割合を $P(i | t)$ とすると、ジニ分散指標は式 1 で表せる。また、 $P(i | t)$ はスポット i のレビュー数を r_i 、クラスタ t に所属するレビュー数を r_t とすると、式 2 で表せる。

$$gini(t) = 1 - \sum_{i=1}^H P(i | t)^2 \quad (1)$$

$$P(i | t) = \frac{r_i}{r_t} \quad (2)$$

式 2 では、各スポットのレビューの割合を単にクラスタ構成レビュー数の割合で算出しているが、そもそも観光スポットによってレビュー数は異なる。そこで、各スポットのレビュー数とそのスポットの全レビュー数との割合を求めることにより、レビュー数の正規化を行なったものを正規化ジニ分散指標と呼称し、これを利用する。スポット i の総レビュー数を R_i とすると、正規化ジニ分散指標における $P(i | t)$ は式 3 で表せる。

$$P(i | t) = \frac{\frac{r_i}{R_i}}{\sum_{j=1}^H \left(\frac{r_j}{R_j} \right)} \quad (3)$$

ジニ分散指標は各クラスの分散の総和のみで値が決まる関係上、クラス数までは考慮されない。我々はより多くの訪問済みスポットで構成されたクラスタは、ユーザの観光に対する嗜好を表現していると考えたため、クラスタの構成スポット数による重み付けを行う。そのとき、あるスポットに関するレビューが 1 つしかない場合でも、そのクラスタを構成するスポットとしてカウントしてしまう問題がある。そこで本研究では、クラスタを構成するあるスポットのレビュー数と、そのスポットの全レビュー数との割合が閾値以上のものを、クラスタを構成するスポットとする。事前の実験により、上記に示した閾値を 1% 以上と決定した。

よって、クラスタ t の嗜好の興味度 $pref(t)$ は、あるスポットに関するレビュー数の割合が閾値以上のスポット数を s_t とすると、式 4 で表せる。

$$pref(t) = gini(t) \times \log s_t \quad (4)$$

4.3 ユーザの嗜好に基づく提示レビューの個人化

本研究では、4.2 節によるスコア付けによりスコア上位のクラスタの重心ベクトルを、ユーザの観光に対する嗜好として用いる。まず、クラスタの重心ベクトルと検索対象となる観光スポットのレビューベクトルとの類似度を算出する。この類似度は、ユーザの観光に対する嗜好とレビューが表す観光スポットの特徴との近さを表す。そのため、スコア上位の各クラスタの重心ベクトルと、観光スポットのレビューベクトルとの類似度を算出し、類似度が高いレビューはユーザの嗜好に合致しているレビュー文となる。

クラスタの重心ベクトルとレビューベクトルの類似度計算には、式 5 に示すコサイン類似度を用いる。

$$\cos(\mathbf{v}_i, \mathbf{v}_j) = \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{\|\mathbf{v}_i\| \|\mathbf{v}_j\|} \quad (5)$$

また、3.2 節で行った検討により内容に乏しいレビューをユーザに提示することを避けるため、レビューベクトルの L2 ノルムによるスコアの補正を行う。

以上から、ユーザの嗜好の興味度、クラスタの重心ベクトルとの類似度、レビューの L2 ノルムに基づく提示レビューのスコア $score(r, t)$ を式 6 に示す。

$$score(r, t) = \{\alpha \times pref(t) + (1 - \alpha) \times \cos(\mathbf{v}_i, \mathbf{v}_j)\} \times norm(r) \quad (6)$$

v_i は階層的クラスタリングによって作成した重心ベクトルとし、 v_j は検索スポットのレビューベクトルとする。 $norm(r)$ はレビュー r の L2 ノルムの値である。式 6 の値が最も高いレビューを、ユーザの嗜好に合致しているレビューとしてユーザに提示する。

5 ユーザの嗜好抽出に関する実験

5.1 実験内容

観光スポット訪問履歴から、ユーザの観光に対する嗜好の抽出が可能かどうかについて実験を行う。また、クラスタのスコア付けによりユーザの嗜好を表現しているクラスタのスコアが高くなるかについて実験を行う。まず、観光スポットに対する嗜好を 3 つ想定したデータセットを著者らが仮想的に用意する。このデータセットを用いて観光スポットのレビューを取得し、階層的クラスタリングによってクラスタを作成し、正規化ジニ分散指標とスポット数によるスコア付けを行う。その後、スコア上位のクラスタを構成するレビューを分析し、想定した嗜好を含むクラスタが作成されているかどうかを確認する。

5.2 データセット

本実験で利用するデータセットの内容を以下に示す。ユーザのデータセットを表 2 のように用意した。表 2 は「初詣」、「タワー」、「生物」を観光地に対する嗜好と仮定したユーザを想定している。また、ユーザの観光スポット訪問履歴には様々な嗜好が入り混じっていることから、「その他」としてジャンルの異なる観光スポットをそれぞれ 1 つずつ含めている。

5.3 結果と考察

ジニ分散指標とスポット数によるクラスタのスコア付けの結果を表 3 に示す。レビュー内の単語出現頻度上位 5 件は、そのクラスタの重心に対し、スポットごとにそのクラスタ内で類似度の高いレビューを 3 件ずつ取得し、それらのレビューに含まれる一般名詞、固有名詞、サ変接続名詞、形容詞の出現回数上位 5 件の単語である。

表 3 のスコア上位のクラスタの単語出現頻度を見ると、184 番のクラスタでは「イルカ」といった「生物」に関するレビュー、223 番のクラスタでは「景色」といった「タワー」に関するレビュー、273 番のクラスタでは「初詣」に関するレビューで構成されており、表 2 で想定した嗜好と一致する。以上の結果から、正規化ジニ分散指標とスポット数によるクラスタのスコア付けは、ユーザの観光に対する嗜好の抽出に有用であることが

表 2 ユーザの観光スポット訪問履歴

観光スポット	想定した嗜好	レビュー数
寒川神社	初詣	255
成田山新勝寺	初詣	788
川崎大師	初詣	241
東京スカイツリー	タワー	238
横浜マリンタワー	タワー	173
ランドマークタワー	タワー	1766
ズーラシア	生物	736
八景島シーパラダイス	生物	1894
美ら海水族館	生物	3035
花やしき	その他	879

表 4 各手法による提示レビューの平均評価点

レビュー提示手法	平均評価点
手法 1 (じゃらん高評価)	2.76
手法 2 (類似度のみ)	2.31
手法 3 ($\alpha = 0.0$)	2.91
手法 4 ($\alpha = 0.1$)	2.93
手法 5 ($\alpha = 0.2$)	2.98
手法 6 ($\alpha = 0.3$)	3.04

表 3 ユーザの正規化ジニ分散指標とスポット数による重み付けを用いたスコア付け結果

クラスタ	スコア	レビュー内の単語出現頻度上位 5 件
184	0.646	子供, イルカ, 楽しい, 圧巻, 広い
223	0.479	展望, 景色, 一望, 夜景, 店
273	0.270	初詣, 参拝, 混雑, お正月, お寺
196	0.097	感じ, 水族館, 古い, 良い, 子供
215	0.077	施設, 展望, 海, 入場, 水族館
173	0.070	誕生, 子供, イベント, 月, ジンベイザメ
261	0.061	赤札, 豆, 成田, 番号, バッグ
218	0.060	横浜, ホテル, ツリー, スカイ, ランドマークタワー
186	0.047	駐車, チケット, オススメ, 良い, 周辺
209	0.043	展望, 写真, 景色, スポット, 大勢

わかった。

6 ユーザの嗜好に基づくレビューの評価実験

6.1 実験内容

従来のレビューの評価順に基づくレビュー提示方法と、提案手法によるユーザの嗜好に基づくレビュー提示方法によるユーザのレビュー評価を比較する。被験者は大学生 11 名を対象に、実験を行った。本実験で比較するレビュー提示手法を以下に示す。

- 手法 1: じゃらんにおける高評価かつ最新レビュー
- 手法 2: クラスタの重心ベクトルとレビューベクトルとの類似度による提示レビュー
 - 手法 3: 式 6 における $\alpha = 0.0$ での提示レビュー
 - 手法 4: 式 6 における $\alpha = 0.1$ での提示レビュー
 - 手法 5: 式 6 における $\alpha = 0.2$ での提示レビュー
 - 手法 6: 式 6 における $\alpha = 0.3$ での提示レビュー

手法 4, 5, 6 における α を大きな値に設定すると、レビューのスコア算出で類似度よりもクラスタのスコアが重視され、同一のレビューしか出力されない。そのため、本実験では変化が見られた 0.0 から 0.3 の範囲を用いる。

まず、被験者から観光スポットの訪問履歴を 10 件と、旅行に行きたい都道府県を 1 つ取得した。観光スポットの訪問履歴はユーザの嗜好抽出に利用し、旅行に行きたい都道府県からじゃらんにおける観光スポットランキング上位 5 件を取得し、レビューの検索対象の観光スポットとして利用する。

次に、被験者に上記の 6 手法によるレビューを提示し、レ

ビューを読んでその観光スポットに興味を湧いたかどうかを、1 から 4 の 4 段階で評価させる。

以上の実験に基づき、既存手法と提案手法の評価の比較、L2 ノルムを用いたスコア付けによる提示レビューの評価の比較、式 6 における α の値による提示レビューの評価の比較をし、考察する。

6.2 実験結果と考察

各手法による提示レビューの評価結果の平均点を表 4 に示す。まず、提示レビューのスコア付けの際にレビューベクトルの L2 ノルムを用いたことによるユーザ評価の差について述べる。スコア付けにノルムを用いなかった手法 2 と、ノルムを用いた手法 3 を比較すると、手法 3 によって提示したレビューの方が、ユーザの評価が高いことがわかる。よって、レビューのスコア付けに関して、レビューベクトルの L2 ノルムを利用することが有用であることが判明した。

次に、既存手法によるレビュー提示と提案手法によるレビュー提示のユーザ評価の差について述べる。表 4 より、既存手法としてあげたじゃらんによる最新かつ高評価レビューの評価（手法 1）と、上記で述べたレビューベクトルの L2 ノルムに基づくスコア付けを用いた手法 3, 4, 5, 6 を比較すると、いずれも提案手法の方がレビューの評価が高いことがわかる。よって、提案手法はじゃらんの高評価かつ最新レビューによるレビュー提示より、ユーザの興味を引くレビューを提示することができた。

最後に、式 6 における重心クラスタとレビューベクトルとの類似度と、クラスタのスコアとの関係性について述べる。手法 3, 4, 5, 6 をそれぞれ比較すると顕著な差はないが、 α の値が大きくなるほどユーザの評価が高くなる傾向にあることが判明した。理由としては、 α の値が大きくなるとクラスタのスコアの影響力が大きくなるため、「クラスタが表現する嗜好の興味度は低い、重心ベクトルとの類似度が高いレビュー」よりも、「クラスタが表現する嗜好の興味度が高く、重心ベクトルとの類似度があまり高くない」レビューを出力することになり、そのようなレビューがユーザに好まれる傾向にあると考えられる。しかしながら、単純に α の値を大きくしてしまうと、類似度に関係なくクラスタのスコアに依存したレビューが出力されてしまうため、 α の値を再度検討する必要がある。

7 まとめと今後の課題

本研究では、観光スポットの訪問履歴からユーザの嗜好を推定し、ユーザの嗜好に合致した検索スポットのレビューを提示

する手法を提案した。具体的には、観光スポット訪問履歴から各訪問スポットに付随する個々のレビューを全て取得し、そのレビューを用いて階層的クラスタリングを行った。次に、得られたクラスタに対しユーザの嗜好を表しているかどうかのスコア付けを行った。その後、スコアの高いクラスタの重心ベクトルと、検索対象となるスポットのレビューとの類似度を算出する。その後、クラスタのスコアである嗜好の興味度、重心ベクトルとレビューベクトルとの類似度、レビューベクトルのL2ノルムによってレビューをスコア付けし、最もスコアが高いレビューをユーザの嗜好に基づくレビューとして提示する。予備実験により、観光スポットの訪問履歴からユーザの観光に対する嗜好の推定が可能であることがわかった。提示レビューの評価実験では、じゃんにおける高評価かつ最新レビューと提案手法による提示レビューの評価を比較した。その結果、提案手法の有効性を確認し、提示レビューのスコア付けに際しレビューベクトルのL2ノルムを利用することの有用性を確認した。また、スコア付けにおける重心ベクトルとレビューベクトルの類似度と、クラスタの嗜好の興味度について、クラスタの嗜好の興味度の割合を高くすることでユーザのレビューの評価が上昇する傾向にあることが判明した。

今後の課題として、 α の値に関する検証、実利用環境における提案システムの妥当性に関する実験を行うことを計画している。これらの実験により、提案手法がユーザの嗜好に合致するレビューを提示できるか、また、ユーザの観光スポット検索の補助に有用であるか明らかにする予定である。また、本稿では大学生11名を対象に評価実験を行なったが、年齢層により観光に対する嗜好が異なることも考えられるため、様々な年齢の被験者を対象に評価実験を行う予定である。また、本研究ではユーザの嗜好を過去の訪問履歴から推定しているが、時間とともにユーザの観光に対する嗜好が変わることも考えられるため、ユーザの観光スポット訪問順などを考慮すべきと考えられる。

謝 辞

本研究の一部は、平成30年度科研費基盤研究(C)(課題番号:18K11551)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] “観光白書 - 国土交通省”, <<http://www.mlit.go.jp/statistics/file000008.html>>, (参照2018-11-12)
- [2] 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”, ライフインテリジェンスとオフィス情報システム研究会, 信学技報, Vol.115(138), pp.29-34, 2015
- [3] 阪井 奎吾, 灘本 明代, “観光レビューを対象とした耳寄り情報提示手法の提案”, 第8回 データ工学と情報マネジメントに関するフォーラム, H3-4, 2016
- [4] 鈴木 綾子, 伊藤 史子, “個人嗜好を考慮した訪問エリア選択支援システムの基礎研究”, 日本建築学会計画系論文集, Vol.75, No.651, 1169-1174, 2010
- [5] 日高 真人, “オンサイト観光プランニングのための観光スポット推薦システムの実現”, 奈良先端科学技術大学院 修士論文, NAIST-IS-MT1651090, 2018

- [6] 奥藪 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.162, No.19, pp.1-8, 2015
- [7] 川崎 彩子, 清木 康, “知的価値を有する観光対象を訪問する個人観光客のための価値観計量による観光地推薦方式”, 第10回 データ工学と情報マネジメントに関するフォーラム, I1-6, 2018
- [8] 上原 尚, 嶋田 和孝, 遠藤 勉, “Web上に混在する観光情報を活用した観光地推薦システム”, 一般社団法人電子情報通信学会, 信学技報, Vol.112, No.367, pp.13-18, 2012
- [9] 中岡 佑輔, パノット シリアーラヤ, 王 元元, 河合 由起子, 秋山 豊和, “ジオタグツイートの多言語性と評判に基づく Venue 推薦”, 研究報告情報基礎とアクセス技術 (IFAT), Vol.132, No.5, pp.1-6, 2018
- [10] WANG Yu, 前田 亮, “トピックモデルを用いた Twitter フォロワー情報からのユーザ嗜好の推測手法の提案”, 第10回 データ工学と情報マネジメントに関するフォーラム, B4-2, 2018
- [11] 関根 裕太郎, 北山 大輔, “ユーザ満足度向上のための推薦理由透明性の影響評価”, 第8回 データ工学と情報マネジメントに関するフォーラム, B3-4, 2016
- [12] Quoc V. Le and Tomas Mikolov. Distributed Representations of Sentences and Documents, In Proceedings of the 31th International Conference on Machine Learning, ICML2014, pp.1188-1196, 2014