

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

潘 健太[†] 北山 大輔^{††}

[†] 工学院大学大学院工学研究科情報学専攻 〒163-8677 東京都新宿区西新宿 1-24-2

^{††} 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: [†]em18011@ns.kogakuin.ac.jp, ^{††}kitayama@cc.kogakuin.ac.jp

あらまし 観光スポットを決める時に Web 上の観光情報を活用して計画を立てることが多くなっている。しかし、膨大な情報から自分の嗜好に合う観光スポットの情報を得ることは簡単ではない。ユーザは観光スポットを決める時、先入観に基づいて決定することも少なくない。そこで本研究では、ユーザの観光スポットのレビュー選択によってユーザの嗜好に合う観光スポットを検索する手法を提案した。本手法では、ランキングやおすすめ情報によって左右されないために、各ユーザの嗜好の抽出を、画面に表示された観光スポットレビューから好みのレビューを選択するという簡単な作業のみで行う。本稿では、プロトタイプシステムを構築し、レビューに季節とタイプ特徴ベクトルを加える場合の検索精度の変化について評価実験と考察を行う。さらに、概要、レビューと観光スポット名の関連性についての実験し、考察を行う。

キーワード 検索, 観光スポット, レビュー

A Consideration on User's Review Selection and Feature Extraction for Search Engine of Sightseeing Spot

Kenta HAN[†] and Daisuke KITAYAMA^{††}

[†] Graduate School of Informatics, Kogakuin University Nishishinjuku 1-24-2, Shinjuku-ku, Tokyo, 163-8677 Japan

^{††} Faculty of Informatics, Kogakuin University Nishishinjuku 1-24-2, Shinjuku-ku, Tokyo, 163-8677 Japan
E-mail: [†]em18011@ns.kogakuin.ac.jp, ^{††}kitayama@cc.kogakuin.ac.jp

Abstract When planning sightseeing spots, tourists often plan using tourist information on the Web. However, tourists are not easy to obtain information on sightseeing spot that match their preference. Furthermore, users often decide visiting sightseeing spot based on some biases. Therefore, in previous work, we proposed a method to search sightseeing spot suitable for user's preference by review selection. With this method, in order not to be influenced by ranking and recommendation information as a bias, extraction of preference of each user is carried out only with a simple operation of selecting a favorite review from the sightseeing spot review. In this paper, we evaluate the search accuracy of the season and the type feature vector to review. Furthermore, we conduct an experiment and discussion on the relation between summary, review and sightseeing spot.

Key words search engine, sightseeing spot, reviews

1. はじめに

旅行先を決定する時、旅行者は観光スポット検索サイトや観光情報に関連する書籍を見て観光スポットを選び、旅行計画を立てる。しかし、膨大な情報から自分の好みに合う観光スポットを見つけることは簡単ではない。行きたい観光スポットが決

まっていない場合ではランキングやおすすめ情報を見て観光スポットを決めることが多くなると考えられる。この時、自分の観点からではなく、観光スポットの来場者数や人気度による先入観が決定に与える影響は大きくなると考えられる。実際に訪問する地域は偏っており、現在、旅行者の滞在先の 48 % は、東京、京都、大阪の主要 3 都市に集中している [1]。この問題に対

し我々は、観光スポット名を隠した状態でユーザがレビューを選択することによって、先入観の影響を減らして、観光スポットを検索する手法を提案している [2]。本稿では、レビューに季節特徴ベクトルと同行者タイプ特徴ベクトルを加える場合の検索精度の変化の評価実験と、概要、レビューと観光スポット名の関連性についての評価実験を行い、被験者実験の結果を踏まえ、ユーザのレビュー選択と特徴抽出について議論する。

本論文の構成は下記のとおりである。2 節では関連研究について述べる。3 節では提案手法の概要について述べる。4 節ではレビューに季節特徴ベクトルと同行者タイプ特徴ベクトルを加えるときの検索精度の変化の評価実験の結果を踏まえて述べる。5 節では概要、レビューと観光スポット名の関連性の評価実験の結果を踏まえて述べる。最後に 6 節では今後のまとめと課題について述べる。

2. 関連研究

クチコミを使用した地理情報の検索および推薦に関する研究は数多く行われている。廣嶋ら [3] は地理情報検索の際のクエリ入力支援として、提示する特徴語の抽出手法について研究を行った。この手法では、各ブログ記事から特徴語候補の抽出および地点の特定を行った。具体的には、特徴語の候補を Wikipedia の見出し語に限定し、ボアソン確率を用いて特徴語抽出を行った。また、地点の特定のための地名表現の抽出および緯度経度への変換の方法としては平野ら [4] の手法を用いた。

観光地検索するとき、松本ら [5] はクチコミから特徴語を抽出して利用する研究を行なった。抽出対象を任意の名詞として、4 種類の手法、TFIDF、ATF(Average Term Frequency)、ボアソン確率、エントロピーのうちどの手法が特徴語抽出に適しているのか検討を行なった。また、抽出した特徴語を利用した検索支援システムを試作し、実験を通して特徴語提示の効果を検証した。従来のレビューを利用する手法では、クチコミを分析して、それぞれ嗜好や価値観を表す単語を重要視して、ユーザには単語レベルで提示されることが多い。

3. 観光スポット検索手法

我々が [2] においてすでに提案している手法の概要を説明する、詳細については論文を参照されたい。単語レベルの提示では、レビューの文章の雰囲気損なうため、ユーザの入力としては、レビューそのものに対して興味の有無を入力することを考えた。本手法では、ユーザがレビューを読むことで行きたい観光スポットを想起できる点に着目して、ユーザの選択レビュー、季節、同行者タイプの 3 つの観点からユーザの特徴を捉え観光スポットを検索する。ユーザは提示された 20 件のレビューから 3 件のレビューを選択する。レビューや観光スポットの特徴として、レビュー文や観光スポットの文書特徴、季節による特徴、同行者のタイプによる特徴のそれぞれの特徴ベクトルを用いる。ユーザの選択レビューの文書ベクトルは、非選択レビューに対する TFIDF 値を特徴量とし、観光スポットの文書ベクトルは、他の観光スポットに対する TFIDF 値を特徴量とする。選択レビューおよび、観光スポットの季節、同行者タイプの特徴ベク

トルの生成は、単語出現確率を用いる方法を定義する。単語出現確率を用いる特徴量の代表的な手法である、カルバックライブラー情報量 (KLD:Kullback Leibler Divergence) を参考に単語の特徴量を算出する。

本手法では、各季節や各タイプの中の単語が全体の出現確率から見て、多ければ特徴的とするために、カルバックライブラー情報量の個々の単語についての部分を使う。たとえば、春の季節の特徴量を算出する場合、単語 j に対する全季節の中における出現確率は $P_{all}(j)$ 、春の単語 j に対する春の文書中における出現確率は $P_i(j)$ とし、各単語の出現確率を求める。これを式 1 と定義する。ここで i は季節やタイプを表す。

$$w_{i,j} = P_i(j) \log\left(\frac{P_i(j)}{P_{all}(j)}\right) \quad (1)$$

4. 季節特徴とタイプ特徴の効果

4.1 実験内容

先行研究 [2] の実験では、全ての特徴ベクトルを使う条件でしか実験できていない。そのため、各特徴ベクトルの効果を明らかにする実験を行う。クラウドソーシングのサービスである、CrowdWorks^(注1)を利用して被験者を集め、じゃらんで取得した観光スポットの文書、季節、タイプを使って、被験者の選択による観光スポットの検索を行った。文書ベクトルに季節特徴ベクトルとタイプ特徴ベクトルをそれぞれ加えた場合での実験を行った結果について考察する。実験は以下の 4 パターンについて行う。

- (1) レビューのみ (Review only: R)
- (2) レビュー+季節 (Review & Season : RS)
- (3) レビュー+タイプ (Review & Type : RT)
- (4) レビュー+タイプ+季節 (All : RST)

4.2 実験結果と考察

本実験では、187 件の使用可能なデータを集めた。実験中被験者は検索された観光スポットが実験開始時に書いた 3 つの要求を満たしているか、観光スポットが既知かどうか、実際に行ってみたかについて回答した。結果を表 1 に示す。「既知・未知」、「行ってみた・みたくない」、満たされた要求の数について、各手法に差は見られない。また、t 検定を行なったが、有意差が見られなかった。そのため、各特徴ベクトルを加えることによって検索精度を上げる効果はないと考えられる。

5. レビュー選択行動の特性分析

5.1 実験内容

本研究では、観光スポットに対する先入観に影響されないユーザの嗜好を抽出するためにレビューを用いている。そこで、スポット名、スポットの概要文、レビューのそれぞれのみから行きたい観光スポットを選んだ時のユーザ行動について実験を行い、先入観に関する特徴を考察する。本実験では、3 ジャンル合計 12 個の観光スポットを用いる (表 2)。ジャンルは、公園、自然景観、神社の 3 つを使用した。観光スポットは東京都

(注1) : <https://crowdworks.jp/>

表 1 季節特徴とタイプ特徴の効果の実験結果

		R	RS	RT	RST
全スポット数		450	470	470	480
行ってみたい スポット数		156(35%)	126(27%)	145(31%)	139(29%)
行ってみたい ない スポット数		294(65%)	344(73%)	325(69%)	341(71%)
行って みたい	1 つ以上	111(71%)	77(61%)	102(70%)	95(68%)
	2 つ以上	65(42%)	43(34%)	65(45%)	58(42%)
	3 つ	37(24%)	13(10%)	30(21%)	30(22%)
行って みたい ない	1 つ以上	155(53%)	112(33%)	167(51%)	166(49%)
	2 つ以上	53(18%)	38(11%)	44(14%)	41(12%)
	3 つ	14(5%)	5(1%)	10(3%)	10(3%)

内の観光スポットであり、じゃらんのサイトにおいて各ジャンルでのランキング上位 20 位から 2 つを有名な観光スポットとして、100 位以下から 2 つを無名な観光スポットとして用いた。この時、概要文とレビューがあるものを選んだ。レビューは以下 (1)～(5) の条件を満たしたものを用いる。

- (1) 3 月～8 月に書かれている
- (2) タイプの属性が一人である
- (3) 観光スポット名がレビューの中には出現しない
- (4) 観光スポット特徴を表現している
- (5) その観光スポットに対して肯定的な内容である

(2) の同行者タイプはじゃらんのレビューの属性であり、一人、カップル・夫婦、家族、友達同士、その他の 5 つである。原則として、1 件のレビューを用いるが、100 字未満の場合は 100 字以上になるように複数のレビューを用いた。(1)～(5) の条件を満たすレビューがない場合、(1)、(2) の条件を満たさないもの、(3)～(5) の条件を満たさないものという順で条件を緩めた。以上の条件に当てはまる観光スポットの概要とレビューを使って観光スポット名のみを示した場合に被験者が行きたいと感じる順位付け(観光スポット名のみ)、スポット 1 のようにスポット名は伏せて、レビュー文のみを示した場合に被験者が行きたいと感じる順位付け(レビューのみ)、同様に概要文のみを示した場合の順位付け(概要のみ)を比較する。同じ被験者が回答する場合それぞれのタスクの順番による影響が考えられるため、順番を入れ変えた。以下の 4 つのパターンのタスクを用いて、それぞれ 50 人づつ、別の被験者が回答した。この時被験者は、先に行ったタスクと後に行ったタスクで同じスポットが含まれていると感じたか、感じた場合にその個数を回答した。

- (A) レビューのみ→観光スポット名のみ
- (B) 観光スポット名のみ→レビューのみ
- (C) 概要のみ→観光スポット名のみ
- (D) 観光スポット名のみ→概要のみ

(A) と (B) によって被験者にレビューを見せる前や後に観光スポット名を見せることで、観光スポット名をレビューに与える影響について調べることができる。(C) と (D) によって被験者に概要を見せる前や後に観光スポット名を見せることで、観光スポット名を概要に与える影響について調べることができる。また、この 2 つの組み合わせからレビューと概要の関係性も調

表 2 実験に用いた観光スポット

	公園	自然景観	神社
有名な 観光スポット	新宿御苑	雲取山	明治神宮
	皇居東御苑	千鳥ヶ淵	深大寺
無名な 観光スポット	仙川公園	都立東大和公園	梶森神社
	牟礼の里公園	南沢緑地	熊川神社

表 3 同じ観光スポットを含むと回答された個数の平均

レビューからスポット名	スポット名からレビュー
4.23	4.74
概要からスポット名	スポット名から概要
3.89	4.55

表 4 各観光スポットの順位の標準偏差の平均

レビューからスポット名		スポット名からレビュー	
レビュー	3.07	スポット名	3.09
スポット名	3.24	レビュー	3.08
概要からスポット名		スポット名から概要	
概要	3.21	スポット名	3.05
スポット名	3.27	概要	3.36

べることができる。

5.2 実験結果と考察

合計 400 件^(注2)のデータのうち使用可能なデータ数は 380 件であった。

表 3 の各値は、「提示した 1 から 12 の観光スポットのレビュー/概要の中に、観光スポット名で提示した A から L の観光スポットと同じ観光スポットがあると思いましたか？」の質問に対しての同じ観光スポットを含むと回答された個数の平均である。レビュー、概要共に、後に観光スポット名を示した場合よりも先に観光スポット名を示した場合の方が、同じ観光スポットがあると感じた数が多かった。観光スポット名が未知である場合、被験者はレビューや概要を具体的な観光スポット名を意識することなく読んでいたと考えられる。被験者が先に観光スポット名のタスクを行なった場合、被験者は観光スポット名を印象に残したままレビューや概要を読み、特定の観光スポットを意識している可能性がある。

観光スポット名のみ、概要のみ、レビューのみのそれぞれ自体が、順位決定のばらつきを与えているかどうか確かめるために、それぞれの条件におけるそれぞれの観光スポットにつけられた順位の標準偏差を求めた。表 4 は各条件における標準偏差の平均である。その結果、特に高い、もしくは特に低い標準偏差は見られなかった。そのため、各条件自体が順位決定にばらつきを与えていないことがわかった。

表 5 はスポット名のみ、レビューのみ、概要のみをそれぞれ先または後に行なった。6 種類の順位の組み合わせ関してのスピアマンの順位相関係数である。表 6 は、各順位において、有名な観光スポットが出現した順位である。観光スポット名(先)と観光スポット名(後)の順位相関が高いことから、観光スポッ

(注2) : 1 人につき 1 オブジェクトの回答で、各オブジェクトで 2 つの回答が得られる

表 5 スピアマンの順位相関係数

	b	c	d	e	f
a: スポット名 (先)	0.84	0.29	0.44	0.17	0.36
b: スポット名 (後)	-	0.58	0.05	0.36	0.35
c: レビュー (先)		-	0.06	-	-
d: レビュー (後)			-	-	-
e: 概要 (先)				-	0.62
f: 概要 (後)					-

表 6 有名な観光スポットの順位の割合

レビュー	トップ 4 入り	0.75(6/8)
	ワースト 4 入り	0.25(2/8)
概要	トップ 4 入り	0.38(3/8)
	ワースト 4 入り	0.25(2/8)
スポット名	トップ 4 入り	1.00(16/16)
	ワースト 4 入り	0.06(2/16)

ト名に関して他の提示情報の影響を受けずに、一般的に行きたい順位によって決定されていると考えられる。また、有名な観光スポット名ほど上位に来やすいことから、被験者は予め知識として持っていた先入観の影響を受けた選択である可能性があることが考えられる。

概要 (先) と概要 (後) の順位相関が高いことから、概要に関しては他の提示情報の影響を受けずに、一般的に行きたい順位によって決定されていると考えられる。概要 (先) と観光スポット名 (先) の順位相関が低いことから、概要から選ばれる観光スポットとスポット名から選ばれる観光スポットには違いがあることがわかる。また、概要による選択では有名な観光スポットが上位にくる傾向がないことから、知名度のような先入観の影響を受けていない選択である可能性があると考えられる。

レビュー (先) とレビュー (後) の順位相関が低いことから、レビューについては先に提示されたスポット名の順位付けの影響を受けて決定されたり、後に提示される観光スポット名の順位付けに影響を与えた可能性があると考えられる。実際、それぞれの順位相関は 0.58, 0.44 と強い相関を示している。特に、先にレビューを提示して後に観光スポット名を選択する場合の順位相関が 0.58 と高く、レビューのみ時の選択基準が、観光スポット名のみ時の選択に影響を与えていると考えられる。一方、レビュー (先) と観光スポット名 (先), レビュー (後) と観光スポット名 (後) のような直接の関係がない組み合わせでは順位相関は高くないことから、レビューによる選択は知名度のような先入観の影響を受けないことがわかる。これらのことより、レビューのみを提示し、それを選択させることは、一般的な選択傾向や、知名度による選択傾向のような先入観の影響を低減し、ユーザの嗜好にあう体験を選択できる可能性があると考えられる。

被験者は「観光スポット名/レビュー/概要から観光スポットの雰囲気イメージしやすさについて」という質問に対して 5 段階の評価をした (表 7)。レビューや概要に比べ、観光スポット名から観光スポットの雰囲気をイメージすることは非常に難しいことがわかった。また、概要とレビューを比べると、概要

表 7 観光スポットの雰囲気のイメージしやすさの 5 段階評価の平均

レビュー	3.22
概要	3.16
スポット名	2.50

の方が被験者にとって雰囲気が少しイメージしやすいことが分かった。概要では、歴史的な知識や内容が書かれている場合がある。それに対してレビューでは、より日常生活に関係する内容が書かれているため、被験者はより簡単に観光スポットの雰囲気をイメージすることができる。また、概要は観光スポットの全体的な紹介を表している。しかし、ユーザが投稿するレビューでは、ユーザが観光スポットを訪れた季節や同行者のタイプによって体験できることが異なるため、レビューの内容も季節やタイプによって異なる。よって選択したレビューによって、ユーザの要求する体験や季節、同行者のタイプを抽出することができると考えられる。

6. まとめと今後の課題

本研究では、ユーザがレビューを選択することによって、観光スポットを検索する手法を提案し、レビューに季節特徴ベクトルとタイプ特徴ベクトルを加えるときの検索精度を評価した。また、観光スポット名、概要、レビューとユーザの選択について実験を行い、先入観による観光スポットの選択の影響を考察した。実験の結果、先入観を減らすことができるユーザの入力として、レビュー選択の妥当性を確認した。また、季節やタイプの影響に関しては検索精度への影響がないことがわかった。今後、レビューを分析する時に新たな重み付けの方法を検討する必要がある。また、実験時に得られたユーザの観光スポットに対する要求の詳細な分析、要求の判断基準の設定とその妥当性についての評価を行う予定である。

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] アンドレ・アンドニアン, 桑原 祐, 山川 奈織美, 石田 遼, 日本の観光の未来 2020 年への持続可能な成長に向けて, マッキンゼー・アンド・カンパニー日本支社, 交通・運輸・物流研究グループ, pp.1-48, 2016
- [2] 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, 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 (DEIM Forum 2018), H1-2, 2018
- [3] 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, 第 26 回人工知能学会全国大会, Vol.26, 1C1-R-5-6, 2012
- [4] 平野徹, 松尾義博, 菊井玄一郎, 地理的距離と有名度を用いた地名の曖昧性解消, 情報処理学会全国大会, Vol.70, No. 人工知能と認知科学, pp.85-86, 2008
- [5] 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, 2013