

# ユーザのレビュー選択に基づく観光スポット検索手法

潘 健太<sup>†</sup> 北山 大輔<sup>†</sup>

<sup>†</sup>工学院大学情報学部コンピュータ科学科 〒163-8677 西新宿 1-24-2

E-mail: †j114095@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。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する。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。実際、旅行者の滞在先の 48% は東京、京都などの主要都市である。そのため、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は、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。また、プロトタイプシステムを構築し、検索性能と簡易性の評価実験を行う。

キーワード 観光スポット, 検索, レビュー, TFIDF, 単語出現確率

## 1. はじめに

2020 年、日本はオリンピック、パラリンピックが開催することが決定し、文化、観光資源を世界へアピールすることの重要性が増してきている。近年、観光業は、訪日外国人の増加による“爆買”現象など、日本の経済の向上効果が期待される。

しかし、訪問する地域が偏っている問題点があり、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]。これは、外国人旅行者のみの問題だけではない。現状は、海外旅行の元気さばかりが目につき、国内旅行については空洞化の危機が懸念されている [2]。また、国内でも観光客は有名な観光地に集中している傾向がある。これらの事によって、都会と地方の差が生まれ、旅行者が主要都市に集中している [3]。

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。また、検索結果には有名な観光スポットが多く表示されるため、ユーザの興味に合致するが知名度のないスポットはそもそも検索結果に表示されないという問題がある。一般的な観光スポット検索では、ユーザがエリア、ジャンル、キーワードなどを指定するため、興味を詳細に入力しようとすれば、時間や手間がかかる問題がある。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する。図 1 は提案手法の概念図である。

本稿の構成を以下に述べる。2 章でレビューによる推薦や検索に関連する研究の紹介を行う。3 章では提案手法の概要について述べる。4 章は単語の重み付け手法とその予備実験につ

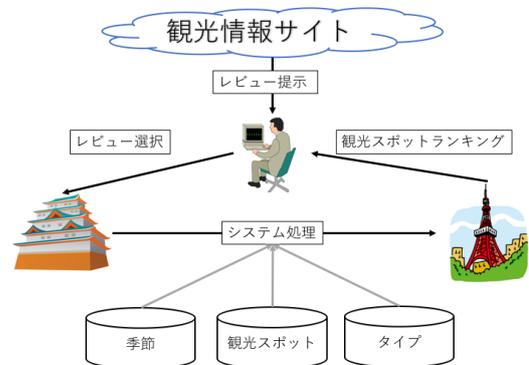


図1 レビューに基づくスポット検索手法

て述べる。5 章では提案手法の評価について述べる。最後に 6 章でまとめと今後の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 [4] 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、ポ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 [5] の手法を用いた。

観光地検索するとき、松本ら [6] 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)、ポアソン確率、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。

上原ら [7] は Web から観光情報を抽出し、複数の特徴ベクト

ルから観光地間の類似性を評価することで、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は、知恵袋・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、地図画像から生成した。これらの特徴ベクトルから観光地間の類似度の測定を行い、類似度の高い観光地を推薦した。

野守ら [8] は日本全国の観光地のクチコミデータを用いて、観光客が話題にする観光テーマを確率的に抽出し、そのテーマを軸として各観光地の特徴を定量的に評価した。またクチコミのテキストデータにテキストマイニングを実行して表現を抽出し、観光地ごとにその表現の出現頻度を集計したクロス集計表にPLSAを実行することで、観光客のクチコミだけに基づいた観光テーマの抽出と観光地の特徴分析を行なった。

地理情報以外でレビューを活用した研究について紹介する。佐藤 [9] はコスメアイテムに対してのクチコミからアイテムの特徴を抽出した。また、それぞれのユーザのクチコミからアイテムに対する不満に相当する項目を抽出し、これらを基に各アイテムの特徴を可視化することで不満を解消するような新しい商品を提案した。林ら [10] は、レビュー中の評価表現とその評価対象に基づき、ユーザが書いた映画レビューから他者が書いた映画レビューを推薦する手法を提案した。評価表現とその評価対象に基づく推薦を実現するために、レビュー中の「良い」や「良くない」などの肯定的または否定的な評価表現とその評価対象に着目した。また、服部ら [11] はレビューを分析して、その価値観と繋がり深い要素としてユーザの「こだわり」に着目し、その結果に基づくユーザモデリング手法について提案した。

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。我々は、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。本研究では、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、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。

### 3. 観光スポットの類似度の算出方法

我々は、ユーザのレビュー選択に基づき観光スポット検索手法を提案する。具体的にはまず、ユーザがランダムに提示されるレビューから好むレビューを選択する。次に、選択されたレビューと類似する特徴を持つ観光スポットを検索する。この時、レビューや観光スポットの特徴として、レビュー文やスポットの特徴、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。特徴ベクトルの生成は、TFIDFを用いる方法と単語出現確率を用いる方法を定義し、予備実験により適切なものを選択する。

#### 3.1 観光スポットのレビューデータの定義

本研究では、スポット、季節、タイプの3種類の特徴を用いる。スポットに関しては、そのスポットについて書かれたすべ

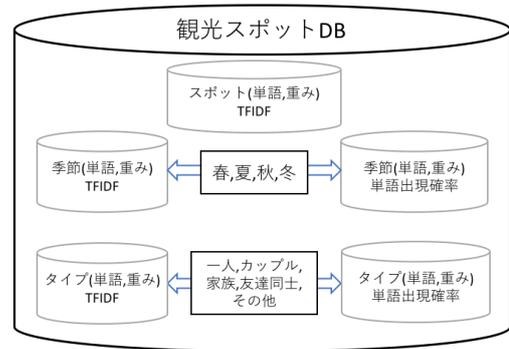


図2 観光情報データベース

表1 形態素解析の例

レビュー文書	園内も広く、気分転換に散歩したりするのにちょうどよい。きれいに清掃などもされていて、気分がよいです。
形態素解析	園内 広い 気分転換 散歩 ちょうどよい きれいな 清掃 気分 よい

てのレビューを1つの文書として扱う。季節は3月から5月を春、6月から8月を夏、9月から11月を秋、12月から2月を冬の時期に分類する。タイプは、一人、カップル・夫婦、家族、友達同士、その他の5つに分類し、その時期に書かれたレビュー集合を1つの文書として扱う。図2のように、各種類の単語とその重みをデータベースとして管理する。

これらのレビューはすべて形態素解析器「MeCab」[12]を使用することで、単語抽出処理を行う。しかし、これらを用いて得られた単語は、日本語として成立しない語が含まれており、これらノイズの削除が必要となる。具体的には、助詞、助動詞、連体詞、記号を削除する(表1)。本稿に置いて、レビューデータはじゃらん<sup>(注1)</sup>から取得したものをを用いる。

#### 3.2 TFIDFによる特徴ベクトル生成

ある観光スポットのレビュー集合を文書*i*とし、*i*に対する単語*j*の出現回数を $TF_{i,j}$ 、単語*j*が出現するスポットの文書数を $DF_j$ 、スポットの文書総数を $|D|$ とした時、そのスポットにおける単語の特徴量は、式1で定義される。

$$w_{i,j} = TF_{i,j} \times IDF_j \quad (1)$$

$$IDF_j = \log\left(\frac{|D|}{DF_j}\right) \quad (2)$$

本手法では入力として、ユーザがレビューを*m*件中*n*件選択する。それをまとめて1つの文書と見なし、他の選択されなかった個々のレビューも文書とみなすことで、式1, 2によってTFIDF値を算出し、選択レビューの特徴ベクトルとする。

季節に関しては、各季節のレビューをまとめて季節ごとに1つの文書と見なす。すなわち、文書の数の合計は4件として、式1, 2によって季節の単語特徴量を算出し、特徴ベクトルを生成する。

(注1) : <https://www.jalan.net/kankou/>

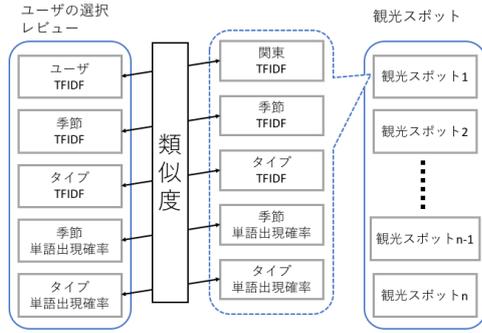


図3 類似度計算の概念図

タイプに関しては、各タイプのレビューをまとめてタイプごとに1つの文書と見なす。すなわち、タイプの文書数は5件となる。これらを式1, 2によりタイプの単語特徴量を算出し、特徴ベクトルを生成する。

### 3.3 単語出現確率による特徴ベクトル生成

季節とタイプに関しては式2の $|D|$ の数が小さいため、わずかな $DF$ 値の差が結果に非常に大きく影響する。そのため、季節とタイプにおいてTFIDFを使うことが適切ではないと考えられる。そこで、季節とタイプは単語出現確率を使って単語の重み付けを行う。

本手法では、単語出現確率を用いる特徴量の代表的な手法である、式3に示すカルバックライブラー情報量(KLD:Kullback Leibler Divergence) [13]を参考に単語の特徴量を算出する。 $N$ は全単語集合である。

$$D_{kl}(P||Q) = \sum_{j \in N} P(j) \log\left(\frac{P(j)}{Q(j)}\right) \quad (3)$$

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(式3)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4と定義する。ここで $i$ は季節やタイプを表す。

$$w_{i,j} = P_i(j) \log\left(\frac{P_i(j)}{P_{all}(j)}\right) \quad (4)$$

### 3.4 コサイン類似度による類似度算出

TFIDFと単語出現確率はL2ノルムを使って正規化を行う。式5は、式1および式4で計算した結果を使って、文書 $i$ の中の単語 $j$ の正規化した重み $n \cdot w_{i,j}$ を求める。 $N$ は全単語集合である。

$$n \cdot w_{i,j} = \frac{w_{i,j}}{\sum_{k \in N} (w_{i,k})^2} \quad (5)$$

3.2節と3.3節によって求めた特徴ベクトルから、ユーザーの選択レビューと観光スポット間の類似度を求める。本研究では、ユーザーの選択レビューに注目するため、ユーザーの選択レビュー

とスポット、選択レビューに含まれる単語の季節ベクトルとスポットの季節ベクトル、選択レビューに含まれる単語のタイプベクトルとスポットのタイプベクトルの類似度(図3)を求める。

ユーザーの選択レビュー $v_i$ と検索された観光スポットの全レビュー $v_j$ の各特徴ベクトルの類似度計算には、コサイン類似度(式6)を用いる。すなわち、ユーザーの選択レビューと観光スポットの全レビューとの類似度、季節、タイプ同士の類似度それぞれを計算する。

$$\text{sim}(v_i, v_j) = \frac{v_{i1}v_{j1} + v_{i2}v_{j2} + \dots + v_{in}v_{jn}}{\sqrt{v_{i1}^2 + \dots + v_{in}^2} \times \sqrt{v_{j1}^2 + \dots + v_{jn}^2}} \quad (6)$$

### 3.5 平均値による観光スポット検索

まず、ユーザーの選択レビュー、季節、タイプの3つの特徴ベクトルに関して、それぞれTFIDFもしくは単語出現確率による類似度を計算する。次にそれを重みつき平均値により統合し、観光スポットを検索する。整理すると、以下の2通りになる。

(1) ユーザーの選択レビューベクトルとスポットベクトルの類似度、TFIDFによる季節特徴ベクトルとタイプ特徴ベクトルそれぞれの類似度の重みつき平均値による観光スポット検索方法

(2) ユーザーの選択レビューベクトルとスポットベクトルの類似度、単語出現確率による季節特徴ベクトルとタイプ特徴ベクトルの類似度の重みつき平均値による観光スポット検索方法

各類似度の重みに関する予備実験を行なった。被験者の要求、季節、タイプの平均値を取るための重みの比を1:1:1にする場合では、検索された観光スポットの結果は被験者の要求に傾き、被験者が選択した季節とタイプは全体的に弱い傾向になっている。平均値を決定するために、実験を行った、被験者の選択レビューを使ってさまざまな平均値の比を用いて検討した結果、平均値の比を2:5:5にすると、被験者の要求、季節、タイプの3つの特徴をバランス良く取ることができる。

## 4. 重み付け手法に関する実験

### 4.1 実験内容

TFIDFによる特徴ベクトルの生成(3.2節)と単語出現確率による特徴ベクトルの生成(3.3節)に検索された観光スポットを比較し、精度の良いものを決定する。

3人の被験者に以下の4つの手順で実験を行った。表示するレビュー数 $m$ は20、選択するレビュー数 $n$ は3で行った。

(1) 被験者が5つのタイプ(一人、カップル・夫婦、家族、友達同士、その他)から1つを選ぶ。選択したタイプと実験時の季節である冬に合致するレビューの中からランダムに20件のレビューを表示する。

(2) ランダムに表示した20件のレビューから被験者が3つのレビューを選択する。

(3) 被験者が選択したレビューに応じて3つの要求を書く。

(4) 被験者が検索された観光スポットに対し、要求、季節、タイプが合っているかを判断する。

また、本実験における被験者の動作および、システムの処理の流れを図4に示す。

被験者はTFIDFによる特徴ベクトルの生成と単語出現確率

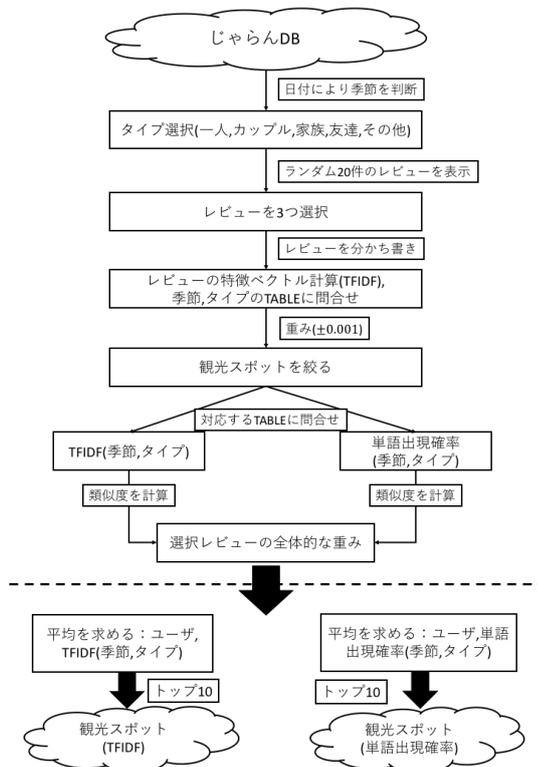


図 4 重み付け手法に関する実験

表 2 被験者が選択したレビューの例

レビュー 1	〇〇にこんな場所があったのかと思うほど静かな雰囲気です。都会の喧騒から離れて心を落ち着けたいときにおすすめです。
レビュー 2	12月の展望台からの景色が最高です。空気は冷たく澄んでいるし。昼に行ってもいいですが、夜のほうがお勧めです！
レビュー 3	温泉の名前の通り、本当にお肌つるつるになります！入浴中ですぐ実感できました。しばらく時間が経っても体ポカポカです。

による特徴ベクトルの生成によって検索された観光スポット 10 件について、被験者の選択レビューに対する 3 つの要求、季節、タイプについてそれぞれの観光スポットが満たされているか、被験者がチェックボックス式で評価した。評価の結果から TFIDF と単語出現確率による特徴ベクトルの生成について考察する。表 2 は評価実験で季節が冬、タイプが友達同士で、被験者が選択したレビューの例である。また、被験者が観光スポットの具体的な場所を特定できないように、都道府県市区町村名を伏せ、「〇〇」に置き換えている。被験者が選択したレビュー全体に対する 3 つの要求は以下であった。

- 要求 1: 静かな雰囲気
- 要求 2: 景色が最高
- 要求 3: 温泉

表 2 の 3 つのレビューを入力とした場合の実行例を示す。TFIDF による観光スポット検索の上位 10 件と単語出現確率による観光スポット検索の上位 10 件および、被験者の回答は、それぞれ、表 3、表 4 のようになる。

表 3 検索された TFIDF による観光スポット上位 10 件の例

観光スポット	要求 1	要求 2	要求 3	友達同士	冬
関宿城跡	○	○		○	○
八幡岬公園	○	○			○
平成の森公園	○	○			
勝浦つるんつるん温泉	○	○	○	○	
稲荷山公園	○	○			
般若院	○	○			
ふれあい交流館			○		○
底倉温泉	○		○		
天然温泉平和島	○	○	○	○	○
妙顕寺	○				

表 4 検索された単語出現確率による観光スポット上位 10 件の例

観光スポット	要求 1	要求 2	要求 3	友達同士	冬
関宿城跡	○	○		○	○
天然温泉平和島	○	○	○	○	○
渋川温泉きらめきの湯	○	○	○	○	○
スカイテルメ渋川					
勝浦つるんつるん温泉	○	○	○	○	
般若院	○	○			
水戸市植物公園	○	○		○	
八幡岬公園	○	○			○
ふれあい交流館			○		○
亀城公園	○	○			
白岩滝	○	○			

## 4.2 実験結果と考察

表 5 は被験者 3 人の結果である。上位部分に絞った場合も考察するために、上位 1 から 5 件と、1 から 10 件に分類している。また、表中の数値は、全体、要求別、季節別、タイプ別に合致したという被験者の回答の割合 (満足度) を示している。全体的に見ると、単語出現確率の方の被験者の満足度は TFIDF より少し高い。また、要求、季節、タイプを別々に、被験者の満足度を見ると、単語出現確率による特徴ベクトルの生成の方が良い結果を出していることが分かる。よって、単語出現確率による特徴ベクトルの生成を使って評価実験を行う。

## 5. 評価実験

### 5.1 実験内容

従来のジャンルによる観光スポットを検索する方法と提案手法のレビューによる観光スポットを検索する方法を比較する。

クラウドソーシングのサービスである、CrowdWorks<sup>(注2)</sup>を利用して 38 人被験者を集め、じゃらんで取得した観光スポットのジャンル 18 個、タイプ、季節を使って、被験者の選択による観光スポットの検索と、被験者がレビューを選択することによる観光スポットの検索の比較を行なった。

A. レビュー選択に基づく観光スポットを検索 (提案手法)

B. ジャンル選択に基づく観光スポットを検索

まず、被験者は 3 つ要求を記述した。要求について、具体的なジャンル、具体的な場所や感情的な要求がある。たとえば、

(注 2) : <https://crowdworks.jp/>

表 5 重みづけ手法に関する実験結果

		全体		要求		季節		タイプ	
		TFIDF	単語出現確率	TFIDF	単語出現確率	TFIDF	単語出現確率	TFIDF	単語出現確率
A	1~10	0.52(26/50)	0.56(28/50)	0.53(16/30)	0.53(16/30)	0.2(2/10)	0.4(4/10)	0.8(8/10)	0.8(8/10)
	1~5	0.48(12/25)	0.64(16/25)	0.53(8/15)	0.60(9/15)	0.0(0/5)	0.6(3/5)	0.8(4/5)	0.8(4/5)
B	1~10	0.34(17/50)	0.42(21/50)	0.27(8/30)	0.40(12/30)	0.3(3/10)	0.2(2/10)	0.6(6/10)	0.7(7/10)
	1~5	0.48(12/25)	0.48(12/25)	0.33(5/15)	0.40(6/15)	0.6(3/5)	0.4(2/5)	0.8(4/5)	0.8(4/5)
C	1~10	0.66(33/50)	0.74(37/50)	0.67(20/30)	0.73(22/30)	0.6(6/10)	0.6(6/10)	0.7(7/10)	0.9(9/10)
	1~5	0.72(18/25)	0.84(21/25)	0.73(11/15)	0.80(12/15)	0.6(3/5)	0.8(4/5)	0.8(4/5)	1.0(5/5)

表 6 観光スポットのジャンル

アウトドア	ウォータースポーツ・マリンスポーツ
雪・スノースポーツ	その他スポーツ・フィットネス
エンタメ・アミューズメント	レジャー・体験
クラフト・工芸	果物・野菜狩り
ミュージアム・ギャラリー	神社・神宮・寺院
伝統文化・日本文化	自然景観・絶景
乗り物	動・植物
風呂・スパ・サロン	ショッピング
観光施設・名所巡り	祭り・イベント

神社、公園、東京、のんびり、自然豊かなどがある。要求の種類によって満足度の差異が発生する。具体的な場所やジャンルと感情的な要求を比べるとさまざまな観光スポットが合致する感情を書いた方が被験者の満足度が高いと考えられる。今回の評価実験では、被験者に要求として感情的な要求を記述するように指示している。次に、被験者が5つのタイプ(一人、カップル・夫婦、家族、友達同士、その他)から1つを選ぶ。以下の2つのシステムがランダムな順番で実行される。

Aのレビュー選択に基づく観光スポットを検索する方法の実験手順は以下のような流れで行う。

- (1) 事前に選択したタイプと実験時の季節である冬に合致するレビューの中からランダムに20件のレビューを表示する。
- (2) ランダムに表示した20件のレビューから被験者が3つのレビューを選択する。
- (3) 被験者が検索された観光スポットに対し、事前に書いた要求にあっているか、既知か未知かについて判断する。

また、Bのジャンル選択に基づく観光スポットを検索する方法の実験手順は以下のような流れで行う。

- (1) 被験者が18個のジャンル(表6)から1つを選択する。
- (2) 被験者が検索された観光スポットに対し、事前に書いた要求にあっているか、既知か未知かについて判断する。

検索した後、被験者は以下の2つの設問について回答した。また、それらの選択理由について自由記述で答えた。

(1) レビュー選択とジャンル選択の検索結果はどちらの方が良いかを選択してください。

(2) 普段の生活中ではレビュー選択とジャンル選択のどちらの方が使いたいかを選択してください。

すべての被験者でAとBの観光スポット検索方法によって

表 7 要求において観光スポットが満たす要求の数

	1つ以上	2つ以上	3つ
すべて(レビュー)	0.71(364/490)	0.36(178/490)	0.13(64/490)
すべて(ジャンル)	0.71(312/440)	0.39(170/440)	0.16(69/440)
既知(レビュー)	0.67(20/30)	0.43(13/30)	0.23(7/30)
既知(ジャンル)	0.62(88/142)	0.35(50/142)	0.15(21/142)
未知(レビュー)	0.71(326/460)	0.36(165/460)	0.12(57/460)
未知(ジャンル)	0.75(224/298)	0.40(120/298)	0.16(48/298)

合計49回<sup>(注3)</sup>の実験データを取得し、これらのデータを基つき、レビュー選択に基づく観光スポット検索とジャンル選択に基づく観光スポット検索について考察する。

## 5.2 実験結果と考察

提案手法では、490件の観光スポットが検索され、そのうち30件が既知のスポットであった。ジャンル選択では、440件の観光スポットが検索され、そのうちの142件が既知のスポットであった。レビュー選択に基づく観光スポット検索の方がより多くの被験者が知らない観光スポットを表示することができるといえる。

表7に評価実験の結果を示す。レビュー選択とジャンル選択によって検索された観光スポットがいくつの要求を満たしていたかを示している。両手法ともに要求を満たす精度は同等であるといえる。すべての観光スポットの結果に関してはジャンル選択による観光スポット検索の方が満足度が要求の個数にかかわらず高い。既知の観光スポットの結果に関してはレビュー選択による観光スポット検索の方が3つともに良い結果を出している。基本的にレビュー選択で検索される観光スポットは既知の観光スポットが少ないが、既知の観光スポットが出てきた場合は満足度が高くなることがわかった。また、未知の観光スポットの結果に関してはレビュー選択による観光スポット検索の方の満足度が3つともにジャンル選択より低くなっている。

次にレビュー選択とジャンル選択の差異について考察する。今回の実験では要求として、被験者はゆっくり、のんびり、楽しい、自然豊かを記述した場合が多かった。しかし、要求の内容が同じでも被験者の考えによって選択するジャンルが異なる。たとえば、のんびりという要求について8つのジャンルが選択されている(表8)。また、のんびりに関するジャンルを選択するとき、自身に関連し、かつ室外体験できるジャンルがよく選ばれた。具体的には、アウトドアとレジャー・体験であった。一

(注3)：評価実験の回答数は被験者1人につき2最大件までである。

表 8 要求がのんびりのときに選択されたジャンルと回数

ジャンル	回数
アウトドア	3
レジャー・体験	3
観光施設・名所巡り	2
雪・スノースポーツ	1
動・植物	1
神社・神宮・寺院	1
風呂・スパ・サロン	1
ショッピング	1

表 9 要求がのんびりのときに選択されたレビューのキーワードと回数

キーワード	回数
人	8
イルミネーション	7
山	6
綺麗	5
地元	5
神社	5

方、要求が同じのんびりのときでも、レビュー(表9)では、イルミネーションや綺麗といったジャンルでは表現しにくい要素を含むレビューが選択されている。また、ジャンルでは1回しか選択されていない神社が上位にきているなどジャンルとは異なる特徴が見られた。このことより、レビュー選択によりジャンルでは表現できない被験者のより詳細な要求を汲み取れる可能性がある。

被験者の要求を1つでも満たす観光スポットは461件であった。そのうち、レビュー選択によって検索された観光スポット数は325件、ジャンル選択によって検索された観光スポット数は136件となっている。ジャンル選択で検索された観光スポットの数が少ない原因として、被験者がジャンルを選択するとき、観光施設・名所巡り、自然景観・絶景、レジャー・体験を選ぶ場合が多く、かつ表示した観光スポットは選択したジャンルのレビューの数が上位のものとなっているため、ジャンル選択によって検索された観光スポットの数が少なくなっている。提案手法のレビュー選択によって観光スポットを検索する手法が被験者により多様な観光スポットを表示し、要求を満たすことができた。また、ある要求に対して検索された観光スポットすべてが合致しなかったケースが、レビュー選択では13回に対し、ジャンル選択では27回となっている。レビュー選択とジャンル選択の差は約2倍となっている。提案手法のレビュー選択によって検索された観光スポットはジャンル選択で反映できない要求を幅広く満たすことができた。

検索後の設問に対する回答について述べる。レビュー選択の方が良いと回答した件数は16件(表10)となっている。また、レビュー選択の方が普段の生活中で使用したいと回答した件数は13件(表11)となっている。被験者の回答より、得られた両手法のメリットをまとめる。

レビュー選択による観光スポット検索では、観光スポットの情報を把握しやすいというメリットがある。これに対し、ジャ

表 10 被験者がレビュー検索の方が良いと回答した例

自分の考えと似た人のレビューが数個あれば、かなり自分好みの検索結果が出る。
レビューにより一目で内容を把握でき、情報収集しやすかったから。
レビュー選択だと、自分の知らない場所ばかりが検索結果として表示され、自分の中で新たな発見がありそうだから。

表 11 被験者がジャンル検索の方が良いと回答した例

レビューを一つ一つ読むのは大変です。
目的の情報にたどり着くのが速いから。
深く考える必要がないから

ンル選択による観光スポット検索では、レビューを選択するために悩む必要がないというメリットがある。

一方、被験者が回答した記述から3つの問題点が見つかった。1つ目は、レビューをランダムに表示しているため、長文のレビューを表示する場合があります。この場合レビューを選ぶ被験者に手間がかかるという問題点がある。2つ目は、レビューでは人によって感覚が異なるため、表示したレビューの受け取り方が人によって異なってしまふ。3つ目は、レビュー選択は被験者が観光スポットに対する目的は明瞭でない時は良い結果を出す。被験者の目的が明瞭である場合は、ジャンル選択の方がより適した結果を出している。これらより、表示するレビューの決定方法や要約方法、また、レビューの個人化手法が必要とする。

## 6. まとめと今後の課題

本研究では、ユーザが行きたい観光スポットが決まっていな場合、ランキング、おすすめ情報やジャンルなどに観光検索情報を使用することによって、検索した観光スポットが偏る問題に着目した。具体的には、ユーザがこれらの観光スポット検索情報により観光スポットに対するイメージの範囲が狭くなる問題点がある。そこでユーザが観光スポットを決める際に、ユーザがレビューのみを選択することによって、ランキングやおすすめ情報などの先入観を左右されないで、観光スポットを検索する手法の提案を行なった。

予備実験で、TFIDFによる特徴ベクトル生成と単語出現確率による特徴ベクトル生成の比較を行い、単語出現確率によって検索された観光スポットの要求、季節、タイプがTFIDFの結果よりもユーザに満たされていることが分かった。

評価実験では単語出現確率を用いてレビュー選択に基づく観光スポット検索とジャンル選択に基づく観光スポット検索の比較を行なった。結果、ジャンル選択に基づく観光スポット検索の方がユーザの要求に対して満たされる場合が多い結果となった。一方、提案手法を用いることでユーザが観光スポットに対する要求が満たされた上で、知らない観光スポットに対する興味と関心を引き出すことができる可能性があることを確認した。

今後の課題としては、関東圏内の観光スポットを使っているため、有名でない観光スポットは東京都以外の県に集中している。次回では、範囲を東京都内に変更して、都内の有名でない

かつユーザに満足する観光スポットを検索できるかについて調査する。ユーザが観光スポットを検索するために、レビュー選択する時に新たな分析方法があげられる。今回の実験では、選択したレビューを分析する時、形態素解析を使っているため、レビューは単語に分けられ、単語に重み付けを行うことに対して、ユーザは句のキーワードを選択している場合がある。結果、検索された観光スポットは選択したレビューの中で重みの高い単語を重要視した観光スポットになる。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 29 年度科研費若手研究 (B)(課題番号: 15K16091), および挑戦的萌芽研究 (課題番号: 16K12536) 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] 室谷正裕, 観光地の魅力度評価—魅力ある国内観光地の整備に向けて—, 運輸政策研究, Vol.1, No.1, pp.14-24, 1998
- [3] 公益財団法人日本交通公社, DBJ・JTBF アジア・欧米豪訪日外国人旅行者の意向調査 (平成 28 年版), 株式会社日本政策投資銀行, pp.1-44, 2016
- [4] 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
- [5] 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
- [6] 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
- [7] 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, 社団法人 電子情報通信学会, 信学技報, Vol.112, No.367, pp.13-18, 2012
- [8] 野守 耕爾, 神津 友武, 口コミデータに PLSA を適用した観光客目線による観光地分析, 第 29 回人工知能学会全国大会, Vol.29, 1J2-OS-18a-2, pp.1-4, 2015
- [9] 佐藤 美里, 関 亜紀子, レビューの分析に基づく化粧品推薦方法の提案, 日本大学生産工学部第 49 回学術講演会講演概要, Vol.49, pp.713-714, 2016
- [10] 林 貴宏, 尾内 理紀夫, Web 上のレビューを利用した映画推薦システム, 人工知能学会論文誌, Vol.30, No.1, SP1-E, 2015
- [11] 服部 俊一, 毛中 杰, 高間 康史, 価値観に基づく情報推薦におけるレビュー分析の利用可能性に関する検討, 人工知能学会 インタラクティブ 情報アクセスと可視化マイニング研究会 (第 1 回), SIG-AM-01-01, 2012
- [12] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237 (2004.)
- [13] 杉山 将, 確率分布間の距離推定:機械学習分野における最新動向, 日本応用数理学会論文誌, Vol.23, No.3, pp.439-452, 2013