DEIM Forum 2019 P1-141

珍スポット検索のためのランキング手法の検討

堀内　進次†　　山本　祐輔†

† 静岡大学情報学部　〒 432–8011 静岡県浜松市中区城北 3-5-1
E-mail: †{horiuchi,yamamoto}@design.inf.shizuoka.ac.jp

あらまし　本稿では, 珍スポットのランドマーク検索手法の提案を行う．提案手法は, 既に珍スポットとわかっているランドマーク名でウェブ検索した結果の文書集合のスニペットおよび, タイトルの形態素解析を行い, 文書中に含まれる形容詞の割合が大きいものを珍スポット特有の形容詞とする．その上で, ランドマーク名でウェブ検索した時の文書集合中に珍スポット特有の形容詞が含まれる割合を分析し, その割合が大きいものを珍スポットとなる可能性が高いと見なす．本研究では, 珍スポットが多いとされる伊豆を取り上げ, 提案システムの評価実験を行った．比較手法として“ランドマーク名+珍スポット”でウェブ検索したときの検索結果ヒット数順のリスト,“ランドマーク名”でウェブ検索したときの検索結果ヒット数順のリストを用意した．その結果提案システムが他 2 手法よりもランキングの精度が高いことを明らかにした．

キーワード　珍スポット，ランドマーク検索

1 はじめに

ある地域を訪問した時あるいは訪問する時にその地域の有名な観光地を知りたいというニーズが存在する．このようなニーズに応えるために，観光ガイドやランドマーク検索が広く利用されている．観光ガイドはある地域の有名な名所や観光地を知るのには有益である．しかし，その地域に住む人あるいはその地域を観光し慣れた人にとっては，有名な名所や観光地を知っても新鮮味がない．むしろ一般的な観光ガイドに掲載されていないような，意外なランドマークを知りたい可能性がある．実際,「ワンダー JAPAN 日本の不思議な異空間」という書籍に代表されるように，一般的な観光客が訪問することが少ないマニアックなランドマークを掲載した雑誌やウェブサイトは一定数存在しており，そのようなランドマークを知りたいというニーズは一定数存在する．本研究では，一般には見かける機会が少ないと思われるランドマークを「珍スポット」と定義し，珍スポットを検索するためのランキング手法について提案する．

提案手法は珍スポットが言及される際によく用いられる形容詞に着目する．その上で，ランドマーク集合からランドマークをウェブ検索したときのタイトルとスニペットに含まれる珍スポット特有の形容詞を含む文書の割合を特徴量とし，ランドマークの珍スポット度を算出する．算出された珍スポット度をもとに，ある地域に存在するランドマークをランキングすることで，珍スポット検索を実現する．

Google マップや Yahoo 地図など，ある地域のランドマークを検索するための検索エンジンは多数存在するが，珍スポット順にランドマークをランキングする仕組みは，我々の知る限りない．珍スポットを検索する手法として一般的なランドマーク検索エンジンに「珍スポット」というキーワードを入力することも考えられるが，メタ情報に「珍スポット」と記載されているランドマークは多くない．また，仮にメタ情報に珍スポットと記載されていたとしても，そのランドマークが実際に珍スポットでない可能性もある．本稿で提案する手法によって，未知の珍スポットを検索できるようになることが期待される．提案手法を活用することで，意外な観光地を探している人や自動車などでの移動時に退屈している人に一風変わったランドマークを推薦することが可能となる．

2 関連研究

2.1 画像や位置情報に着目したランドマーク検索の研究

画像の類似度や位置情報に着目したランドマーク検索の研究はいくつか行われている [1] [3] [4] [5] [6] [7]．川久保らは VisualRank を用いて位置情報付き画像のランク付けを行った．VisualRank とは画像の類似度行列に PageRank アルゴリズムを適用してものである．画像の視覚的特徴と位置情報からバイアスベクトルを作成し，キーワードをクエリとして与えたときの地域の代表画像を抽出に成功している [1]．川久保らは，Web 上から取得した画像を用いて画像の視覚的特徴と単語概念に加え，単語とジオタグの関係性についても分析も行っている [7]．

一方で，帆足らは，SNS や Flickr などの写真共有サイトの携帯電話で撮影した位置情報付きの画像を想定し，人々が興味を示す画像 (POI 画像) の抽出を行った．提案手法では，先行研究においてすでにわかっている位置情報付きの POI 画像の近傍に位置するランドマークを検索対象の POI 画像候補とし，得られた検索対象ランドマーク候補を Web 検索する．さらに，Web 検索において得られた候補群の画像と POI 画像の content-based 類似度を算出し，その結果に基づく POI 画像を推測し，POI 画像にランドマーク情報を付与する手法の提案を行っている [5]．

その他に山本，中澤らはランドマークの場所性と象徴性を特徴量としてシステムに与え，定量化することで強いランドマーク検索の実現，タブレットなどの端末への応用を実現し，実装

可能性，妥当性，有用性を明らかにした．強いランドマークとは，場所性と象徴性の両特性が人に強く認識されるものである．例えば，駅や市役所が強いランドマークに属する [4] [3].

また，中林は画像をクエリとした類似度計算手法として SIFT 特徴量 (Scale Invariant Feature Transform) を用いた検索が有用であることを明らかにした [6].

2.2 Web 上のテキストを利用したランドマーク検索の研究

他方では，テキストベースでのランドマーク検索や地域性の算出に関する研究が行われている [8] [9]．倉島は blog から地域性やランドマークなどの街の話題抽出を行っている. 提案手法では,「へ/に」などの助詞からランドマークや地域の抽出，サ変名詞や動詞から体験の抽出を提案している [9].

一方で，土田は Word2vec を用いて地域やランドマークの意味演算を行っている Word2vec は単語をベクトルに変換し定量化することができ,「東京-日本+フランス= パリ」のような地域やランドマークの加算，減算が可能である．提案手法では，twitter のテキストを分析対象として Word2vec を適用し，観光分野に応用可能な情報抽出を提案し，“「東京」と「スカイツリー」” の関係性が “「大阪」と「あべのハルカス」” との関係性に近いなどといった地域とランドマークの関係性について明らかにした [8].

上記のようにランドマークや地域性の研究は様々なアプローチから行われている．

2.3 本研究との比較

本研究の関連研究としていくつか述べたが，珍スポットを検索に関する研究は筆者が知る限り行われていない．本研究の提案手法では，都市名を入力としてデータベースに問い合わせヒットしたランドマーク集合を珍スポット順に並べる．その際，節 2.2 のように Web 上のテキストを利用し，ランドマークを Web 検索したときのタイトルとスニペットを形態素解析を行う．形態素解析の結果の形容詞に着目し，珍スポットに特有の形容詞を抽出する．抽出した形容詞を用いて，ランドマーク集合の各ランドマークに比較指標を与える．

3 提 案 内 容

本章では，珍スポット検索を行うためのランキングの方法について述べる．

提案手法では，道の珍スポットの発見のために，既知の珍スポットが言及される文書に特徴的に現れる形容詞に着目する．具体的には，事前に既知の珍スポットが言及される文書に特有の形容詞を抽出する．その後，未知のランドマークについて，ランドマーク名でウェブ検索を行い，得られたウェブ検索結果に含まれるタイトルとスニペットに含まれる語を分析する．事前処理で抽出済みの既知の珍スポットに特有の形容詞が含まれる文書の割合が大きい場合，対象ランドマークは珍スポットであるとみなす．以降では，提案手法の各ステップについて具体的に述べる．

3.1 珍スポットを言及する文書に特有の形容詞の抽出

珍スポットを言及する文書に特有の形容詞の抽出方法について述べる．本研究では，Bing Search API[1]を使用し，既知の珍スポットと分かっているランドマークをウェブ検索し，ウェブ検索結果を 50 件取得する．その後，取得したウェブ検索結果リストに含まれるタイトルおよびスニペットを形態素解析し，検索結果リストに含まれる形容詞が珍スポットを言及する際に特徴的に用いられる語かどうかを分析する．この際，既知の珍スポットランドマーク集合 L 中の各ランドマーク名でウェブ検索したときに得られるウェブ文書のタイトルとスニペットに含まれる形容詞 w_a を含む割合を計算し，w_a が珍スポットに特有の形容詞である度合い $f(w_a, L)$ を計算する．$f(w_a, L)$ の定義を以下に記す：

$$f(w_a, L) = \frac{1}{|L|} \sum_{l \in L} \frac{H(l, w_a)}{H(l)} \tag{1}$$

ここで，$|L|$ は L の要素数を表す．$H(l)$ はランドマーク l で検索したときの文書の数を表す.

$f(a, L)$ を用いて珍スポットに特有の形容詞のランキングを作成する．本研究では既知の珍スポットとしてワンダー JAPAN [2] から計 100 件を抽出し，珍スポットに特有の形容詞の抽出に用いた．作成した珍スポット特有の形容詞のランキングから一般的なランドマークにも共通している形容詞を除外するために一般的なランドマークについても同様に上記の手法を用いて一般的なランドマークに有する形容詞のランキングを作成する．本研究では，一般的なランドマークとして “ 伊豆の有名観光地 ” として取り上げられているランドマーク計 40 件を用いる [2]．作成した 2 つのランキングのリストを用いて，珍スポットを言及するウェブページに頻出する形容詞リストと一般的なランドマークを言及するウェブページに頻出する形容詞リストを比較する．双方のリストに出現する形容詞について，上式で計算した度合いの差を求め，差の絶対値が 0.01 以上のものを珍スポットに特有の形容詞とした．最終的に得られた珍スポット特有の形容詞を表 1 に示す.

3.2 珍スポット検索のためのランキング手法

前節で述べた手法によって抽出した珍スポット特有の形容詞リスト W_a を用いてランドマーク集合を珍スポット順に並べる手法について述べる．

提案システムは，入力として都市名を受け取るとランドマークデータベースから指定された都市に属するランドマーク集合を得る．次に，得られたランドマーク集合を珍スポット度を算出する関数を用いてソートする．本研究では，ランドマークをウェブで検索したときのタイトルとスニペットの文書に含まれる珍スポット特有の形容詞を含む文書の割合が大きいほど．珍スポット度は高いと考える．そこで，任意のランドマーク l と

1 : https://azure.microsoft.com/ja-jp/pricing/details/cognitive-services/search-api/

2 : https://retrip.jp/articles/7216/

表 1　珍スポットに特有の形容詞

形容詞	割合
怪しい	0.01566
ぼろい	0.01157
黒い	0.00828
おおい	0.00270
うまい	0.00191
暗い	0.00174
太い	0.00169
っぽい	0.00168
嬉しい	0.00150
愛しい	0.00146
くろい	0.00145
あやしい	0.00142
永い	0.00116
ぽい	0.00107
名高い	0.00106
もったいない	0.00106
細い	0.00106
臭い	0.00105
恐ろしい	0.00105
くい	0.00104
ありがたい	0.00102
たまらない	0.00102
稚い	0.00102
正しい	0.00089
明るい	0.00084
弱い	0.00083

珍スポットに特有の形容詞集合 W_a が与えられたとき，l の珍スポット度 $Rank(l, W_a)$ を下記式で定義する

$$Rank(l, W_a) = \frac{1}{|W_a|} \sum_{w_a \in W_a} \frac{H(l, w_a)}{H(l)} \quad (2)$$

本研究では任意のランドマークの珍スポット度を求める際，ウェブ検索で得る検索結果数を最大 100 件とした．また，使用する珍スポット特有の形容詞を表 1 中の形容詞の上位 3 つとした (怪しい，ぼろい，黒い)．

4　評 価 実 験

本章では提案手法の評価実験について示す．

4.1　デ　ー　タ

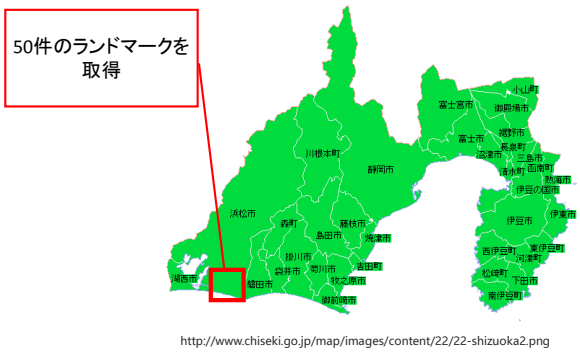
今回の実験で扱うランドマークの情報は，FourSquare API[3] を用いて Foursquare から取得した．FourSquare API で 1 度に取得できるランドマークは 50 件である．そのため，指定範囲（正方形の矩形）をずらしながら伊豆地方のランドマークを取得する．

ランドマーク情報取得のイメージ図を次の図 1 に示す．

作成したデータベースに“伊豆”というクエリを与えて該当するランドマークを収集した．結果，伊豆のランドマークを合計 1214 件収集した．

3：https://developer.foursquare.com/

図 1　ランドマーク情報取得のイメージ図



http://www.chiseki.go.jp/map/images/content/22/22-shizuoka2.png

4.2　比 較 手 法

ランキング評価のために提案手法を用いて作成した珍スポットのランキング手法のほかに 2 つの手法を用いて，ランキング手法の性能比較を行った．ベースラインとなる手法は以下の 2 つである：

- “ランドマーク名+珍スポット”でウェブ検索したときの検索結果ヒット数順に並べる手法
- “ランドマーク名”でウェブ検索したときの検索結果ヒット数順に並べる手法

4.3　既知の珍スポット用いたランキングの評価

既知の珍スポットを抽出することができるか検証比較を行うために伊豆のランドマークをランダムに取り出した 30 件のランドマークにワンダー JAPAN [2] から無作為に抜粋した伊豆半島の 5 件の珍スポットを挿入したリストを用意した．計 35 件のランドマークが入ったリストを提案手法のランキングのほか 2 つの手法を用いてランキング化し，それぞれのランキングを評価した．

挿入したランドマークを次の表 2 に示す．

表 2　挿入したランドマーク

珍スポット名
怪しい少年少女博物館
伊豆極楽苑
ねこの博物館
熱海秘宝館
伊豆長岡温泉赤線跡

それぞれのランキングを評価する尺度として P@k を定義する．P@k はランキングの k 件目までに存在する正解珍スポットの割合である．

4.4　ユーザ目線の珍スポットランキングの評価

データベースから“伊豆”とクエリを与えたときにヒットしたランドマーク 1214 件に対して提案手法と他 2 つの手法を適用し，性能を比較した．作成した各リストの上位 30 件を抜き出し，30 件までの P@k を求めた．

ランキング中のランドマークが実際に珍スポットであったか

どうかを判別するために，大学生 5 人に次の全てのランドマークを珍スポットかどうか判別してください．判別する際の珍スポットの定義は，普段見かけないような一般的でない像，施設または観光地としますという問いをかけた．珍スポットである場合には 1，珍スポットでない場合は 0 と各リスト内の全てのランドマークに評定をつけてもらった．5 人のうち 3 人が珍スポットと答えたものを珍スポットとした．このデータを基にランキング評価を行った．

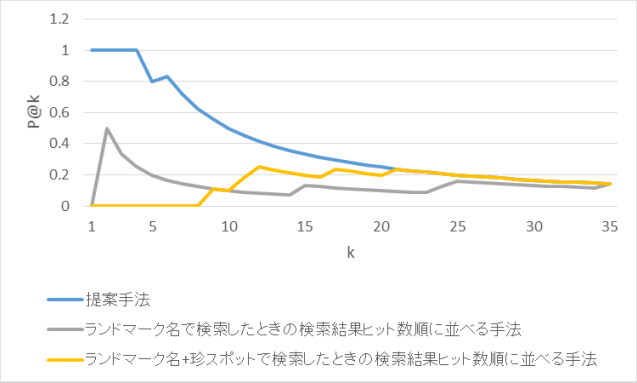
5 結　　果

本章では，4 で述べた実験の結果について記す．

5.1 既知の珍スポットを用いたランキング評価実験の結果

次に伊豆のランドマーク 30 件と既知の珍スポット 5 件の各手法におけるランドマークリストの P@k グラフを図 2 に示す．

図 2　既知の珍スポットを用いた各におけるランドマークリストの P@k のグラフ



提案手法における P@k のグラフは k の値がいくつになってもその他 2 つの手法よりも下回ることはなかった．提案手法は，ランキングの 6 位までに挿入した珍スポット 5 件をランクインさせた．一方，ベースラインとなる 2 手法は，既知の珍スポット全 5 件をランキングにランクインさせるのにそれぞれ 35 位，21 位までランキングを走査する必要があった．

提案手法によるランキング結果を次の表 3 に示す．

挿入した 5 件のランドマークが上位 6 件にランクインした．また，13 件目以降の珍スポット度が 0 と珍スポット特有の形容詞を含まないランドマークが存在した．実際のランキングの 7 位にランクインした「土肥金山山上社」はワンダー JAPAN [2] に載っているランドマークである．

5.2 ユーザ目線の珍スポット判定を用いたランキングの評価実験の結果

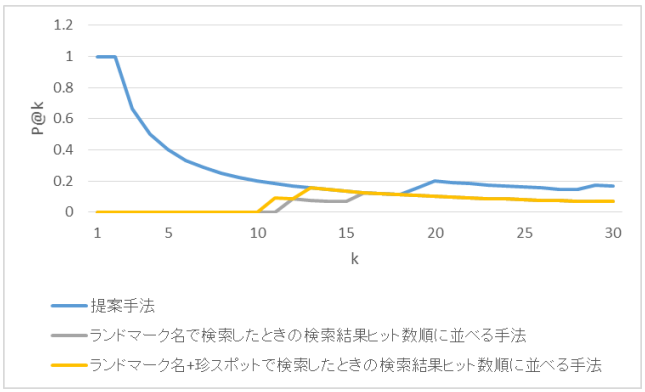
実際にユーザが珍スポットだと感じるランドマークの検索性能を評価するために，4.4 節の評価データを用いて，手法評価を行った．各手法の P@k のグラフを図 3 に示す．

3 手法の各ランキングの上位 30 件にランクインした珍スポットと判別されたランドマークは合計 9 件存在した．

表 3　既知の珍スポットを用いた実際のランキング

ランドマーク名	珍スポット度
怪しい少年少女博物館	0.393
伊豆極楽苑	0.023
ねこの博物館	0.011
熱海秘宝館	0.007
天城越え歌碑	0.004
伊豆長岡温泉赤線跡	0.0024
土肥金山 山神社	0.00227
浄蓮の滝	0.00224
カフェレストラン マシェリ	0.00126
世界一の巨大金塊	0.00120
伊豆の佐太郎	0.001122
ホタル公園	0.001118
駿河湾洋上	0.00116

図 3　ユーザ目線の珍スポット判定を用いた P@k のグラフ



提案手法における P@k のグラフは k の値がいくつになってもその他 2 つの手法よりも下回ることはない．

提案手法，他 2 手法でランキングした結果，検索結果上位 30 件までにランクインしたランドマークのうち珍スポットと判別されたランドマークはそれぞれ 5 件，2 件，2 件であった．提案手法は上位 2 位に珍スポットと判別されたランドマークをランクインさせた．他の 2 手法ともにランキングの上位 10 件に珍スポットと判別されたランドマークがランクインすることはなかった．

提案手法と他 2 手法のランキングの上位に表示されるランドマークの画像の例を次の図 4 に示す．また，画像の下に引用 URL を示す．

6 考　　察

6.1 提案手法の有用性

実験結果 5.1 ではランダムに取り出した伊豆のランドマーク 30 件に加えて既知の珍スポット 5 件の計 35 件において提案手法と他 2 手法でランク付けを行った．

ランキングの評価のグラフは，提案手法における P@k の値がその他 2 つの手法よりも k の値がいくつになっても下回ることはない．そのため，提案手法のランキングは，珍スポットランキングとしてその他手法より優れているといえる．また，提

図 4　各手法の上位画像の例

4)https://tblg.k-img.com/restaurant/images/Rvw/78764/150x150_square_78764621.jpg
5)https://tblg.k-img.com/restaurant/images/Rvw/71664/150x150_square_71664199.jpg
6)https://tabelog.com/shizuoka/A2205/A220503/22002537/dtlphotolst/4/smp2/
7)https://ja.foursquare.com/v/js-bar/5144721fe4b0b6e406dffd33/photos
8)http://www.minami-izu.jp/entry.html?id=11163
9)https://bqspot.com/photo2006/2006-02-14-01.jpg
10)http://www.funkygoods.com/hai/oohito/oohito_a.html
11)https://www.facebook.com/pages/%E3%81%86%E3%82%81%E3%81%AF%E3%82%89-%E3%82%89%E3%81%81%E9%BA%BA/153782681652191 8

案手法の上位 10 件には挿入した珍スポット 5 件がランクインした．したがって，提案手法のランキング性能は最も高いと考える．

ランキング評価のグラフから，ランドマーク名+珍スポットでウェブ検索したときの検索結果ヒット数順のリストがその次に優れているといえる．この結果から，珍スポットであるランドマークを珍スポットというキーワードを加えてウェブで検索して絞り込むことは有効であるといえる．

提案手法ではワンダー JAPAN から抜粋した伊豆以外も含む既知の珍スポットについて珍スポット特有の形容詞の抽出を行った．抽出された形容詞を使用してランキングを作成した際，挿入したランドマークが上位にランクインした．このことから全ての珍スポットをウェブ検索したときのタイトルとスニペットの文書に珍スポットを表す語として共通した形容詞が用いられることが多いと考えられる．したがって，本研究の提案手法は伊豆以外の他の地域にも適用できる可能性が高いと考える．

実ユーザによる珍スポットランキング評価においても，提案手法は上位に珍スポットと判別されたランドマークをランクインさせたことや上位 30 件のランキングに 5 件の珍スポットをランクインさせたことから他の 2 手法より優れているといえる．

6.2　提案手法の今後の課題

実験を行う際に珍スポット度を珍スポット特有の形容詞を 3 つの要素を用いて算出した．その際，珍スポット特有の形容詞の重みを全て同一であるとしたが，形容詞の頻出度合いから重み付けをすることでより多くの珍スポットが上位にランク付けされると考える．

また，本実験でランドマークをウェブ検索したときの文書集合中に珍スポット特有の形容詞を含む文書の割合をランドマーク比較指標として与えたが，視覚的特徴などの他の観点からのパラメータを与えることで全てのランドマークに珍スポット度を与えることが可能であり，よりランキングの精度を高めることができると考える．

7　おわりに

本研究ではランドマークをウェブで検索したときのタイトルとスニペットの文書集合中の珍スポット特有の形容詞が含まれる割合に注目し，ランドマークに特徴量を与えてランドマーク集合のランキング化する手法の提案を行った．

既知の珍スポットを正解とするデータセットおよびユーザ目線の珍スポットを正解とするデータセットを用いた比較実験を行った．その結果，提案手法の有用性を示すことができた．

今後の課題として，珍スポット度を決定するときに用いる形容詞の重み付け，ランドマークの視覚的特徴量を考慮したランキング関数の設計などが挙げられる．

謝辞　本研究は JSPS 科研費 JP17K17832，JP18H03494，JP18KT0097 の助成を受けたものです．ここに記して謝意を表します．

文　献

[1] 川久保秀敏 (2010)「位置情報を考慮した VisualRank による地域別代表画像の選出」,MIRU2010.
[2] 三才ブックス (2016)『ワンダー JAPAN 日本の不思議な《異空間》800 三才ムック vol.638』三才ブックス
[3] 中澤優一郎 (2012)「象徴性と相対的場所性に基づく強いランドマーク検索システムの実現方式」,DEIM2012
[4] 山本 隆徳 (2012)「道路の認知度と道路までの距離に基づいたランドマークの特徴量算出方式の実現とランドマーク検索への応用」,DEIM2012
[5] 帆足 啓一郎 (2009) 画像類似度を利用した位置情報付き写真データへのランドマーク情報付与手法」, 情報処理学会論文誌　データベース Vol.2 No.3 pp.41-52
[6] 中林 拓馬 (2013)「画像をクエリとした建築情報検索システムの実現に向けた類似画像検索技術の検証」, 日本建築学会技術報告集　第 19 巻　第 42 号,pp.775-778
[7] 川久保 秀敏 (2009)「Bag-of-Features 表現を用いたエントロピーによる単語の視覚性の分析」,2009-CVIM-166 pp.267 - 274
[8] 土田 崇仁 (2016)「Word2Vec を用いた地域やランドマークの意味演算」,DEIM2016
[9] 倉島健 (2005)「Blog からの街の話題抽出手法の提案」,DEWS2005