

旅行目的の関連性に基づく地理オブジェクト集合抽出手法

池田洋平[†] 北山大輔[†]

工学院大学情報学部[†]

1 はじめに

近年、旅行先を決める上で web を利用することが多くなった。例えば旅行先で評判の良い温泉に行きたいなら、じゃらん net などでもランキングを見て、上位の温泉がある場所を旅行先を選ぶことが出来る。しかし、実際に旅行に行く時は1つの目的だけではなく複数の目的があることが多い。一般的に、こうした1つの目的を満たした情報を得ることは容易でも、複数の目的を満たした情報を得ることは容易ではないのが現状である。

観光ルートを推薦する研究 [1][2] が行われている。これらの研究では、ユーザの嗜好や希望度、その他の制約を満たす最適なルートを推薦する。しかし、この前段階であるどこを目的地にするべきかどうかということには答えてくれない。そこで、本研究ではユーザが入力する複数の目的を満たした旅行地を推薦するシステムを提案する。オブジェクトの web 上での人気や移動時間、関連性を用い、旅行先となるオブジェクト集合の統合的なスコアを算出する。まずそのために、旅行時における複数の目的からオブジェクト集合の生成を行い、次に得られたオブジェクト集合の評価を行う。良い評価を得られる観光地は、人気のオブジェクトが含まれていることと、オブジェクト間の距離が近いこと、そしてそういったオブジェクト間の関連性が高いことであると考えた。よって関連性の高さ、人気の高さ、距離の近さからオブジェクト集合の評価を行う。図1は本手法の概念図を示す。

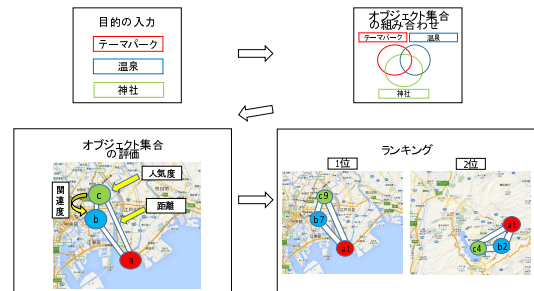


図 1: 地理オブジェクト集合の概念図

評価を行う際計算量が膨大になる。そこで、目的ごとに後述する人気度の上位 n 件のオブジェクトを選択し、それらのオブジェクトの半径 m [km] にあたる範囲からオブジェクト集合を生成する。選択したオブジェクトの半径 m [km] 以内に他の選択したオブジェクトが存在し対象とする地域が重なってしまう場合は、順位が下位のオブジェクトを選択から外し新たに $n+1$ 位以降のオブジェクトを選択する。これら n 箇所の地域で目的ごとに1つずつオブジェクトを選択することでオブジェクト集合を生成する。

2.2 オブジェクト集合の評価

オブジェクト集合の評価は関連度、人気度、距離スコアからオブジェクト区間のスコアを求め、それらの平均を求めることでオブジェクト集合の評価 O を求める。式中の N はオブジェクトの数、 $S_{i,j}$ はオブジェクト i と j の区間におけるスコアである。

$$O = \frac{S_{1,2} + S_{1,3} + \dots + S_{N-1,N}}{n C_2} \quad (1)$$

$$S_{i,j} = R_{i,j} \times P_{i,j} \times D_{i,j} \quad (2)$$

関連度 $R_{i,j}$ は目的間の関連性とし、オブジェクト i,j の2つの目的 A,B をキーワードとした旅行レビューの検索結果数から Jaccard 係数を用いた。2つのキーワードが旅行レビューの同じページ内に存在していれば、セットで行く場所であり関連性が高いとした。

2 オブジェクト集合の生成と評価

2.1 オブジェクト集合の生成

本研究ではオブジェクト集合を各目的に対応したオブジェクトが1つずつ選択された集合と定義する。オブジェクト集合の生成は本来総当りで行うが、集合の

An Extraction Method of A Set of Geographical objects based on Relevance among Travel Purposes
 YOHEI IKEDA[†]DAISUKE KITAYAMA
 Faculty of Information Studies, Kogakuin University
 163-8677, Shinjuku, Japan

j109011@ns.kogakuin.ac.jp, kitayama@cc.kogakuin.ac.jp

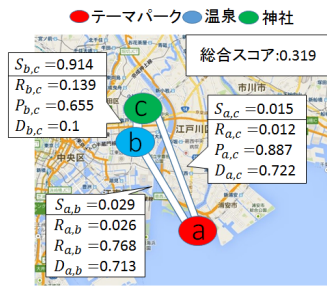


図 2: オブジェクト集合の評価の例

人気度 $P_{i,j}$ は施設の人気の高さとし、foursquare の訪問数を用いた。オブジェクト i および j の訪問数の平均を求め、それをその地域における最大の訪問数 Max_p で割ることで、人気度の値が 0 より大きく 1.0 以下の値を取るようにした。

$$P_{i,j} = \frac{(\log_{10} p_i + \log_{10} p_j) / 2}{\log_{10} Max_p} \quad (3)$$

距離スコア $D_{i,j}$ はオブジェクト i, j 間の距離とし、距離の逆数を用いた。距離 $d_{i,j}$ は緯度経度の差と地球の半径から直線距離を求め、単位は km とした。なお底が 10 の log の場合、距離が 1.0km 未満のとき負の値をとってしまう。そのため、距離が 1.0km 未満のときは例外的に距離を 1.0km として扱うこととした。

$$D_{i,j} = 1 - \frac{\log_{10} d_{i,j}}{\log_{10} Max_d} \quad (4)$$

3 結果と考察

旅行目的をテーマパーク・温泉・神社、対象地域を浦安市・箱根市・日光市とし、その 3 地域から各目的ごとに 30 件ずつ foursquare からオブジェクトの情報を取得し、30km の範囲内でオブジェクト集合の生成及び評価を行った。関連度は表 1 とした。表 2 は、各地域で最も高くなったオブジェクト集合の評価 O であり、1 が浦安市、2 が箱根町、3 が日光市の結果である。

結果としては、日光市では訪問数も多く、距離的にも近い集合が得られた。その一方、箱根町では地域内で一番訪問数の多いテーマパーク「鈴廣 かまぼこの里」では無く、4km ほど温泉と神社に近いが訪問数が非常に少ない「Forest Adventure」が含まれた集合が得られた。浦安市でも、温泉と神社の距離は非常に近いが、どちらも訪問数は非常に少ないオブジェクトが含まれた集合が得られた。よって、全体として訪問数の多さよりも距離の近さが重視された結果となった。より実用的な結果を得るためには、人気度と距離スコアが同

表 1: 関連度

テーマパーク-温泉	テーマパーク-神社	温泉-神社
0.026	0.012	0.139

表 2: オブジェクト集合の評価

	テーマパーク	温泉	神社	O
1	ディズニーランド	天然温泉	亀戸天神社	0.0307
2	Forest Adventure	ユネッサン	箱根神社	0.0320
3	日光江戸村	日光やしおの湯	日光東照宮	0.0269

じ程度考慮されるように、これらの値の設定を行う必要がある。

4 まとめ

本論文では、ユーザが入力する複数の目的を満たした旅行地を推薦するシステムを提案し、オブジェクトの web 上での評価や移動距離、関連性を用い、旅行先となるオブジェクト集合の統合的なスコアを算出した。

今後の課題としては、多数の目的を入力した場合に、それら全てを含む地域が存在するとは限らないため、全ての目的オブジェクトを含んでいる訳ではない地域を扱える必要がある。また、食事処など同じ目的のオブジェクトが複数あってもよいオブジェクト集合の評価方法などを考える必要がある。さらに、目的ごとに自分で重みをつけることでよりユーザの満足度の高い結果が得られると考えられるため、今後目的ごとの重みを考慮した仕組みを考えていく事も課題である。

謝辞

本研究の一部は、平成 25 年度科研費若手研究 (B)(課題番号: 24700098) によるものです。ここに記して謝意を表すものとします。

参考文献

- [1] 丸山 敦史, 柴田 直樹, 村田 佳洋, 安本 慶一, 伊藤 実, P-tour:観光スケジュール作成支援とスケジュールに沿った経路案内を行うパーソナルナビゲーションシステム, 情報処理学会論文誌, vol.45, No.12, PP2678-2687, 2004
- [2] 倉田 陽平, 奥貫 圭一, 貞広 幸雄, 個人嗜好に応じた観光コース自動作成システムの開発, 地理情報システム学会講演論文集, vol.9, PP199-202, 2000