

# 歴史に基づく観光スポット推薦手法の提案

11371120 森尾 健太郎（灘本研究室）

あらまし：本研究では観光雑誌や観光地サイトなどの、今までの情報源では知り得ることができなかった新たな観光スポットや既存の観光スポットから、その過去の情報を我々が提案する跡地探索キーワードを用いて抽出し、歴史に基づく観光スポットの推薦を手法の提案を行う。

## 1. はじめに

現在、観光雑誌やインターネット上の観光サイト、旅行関連のテレビ特番など様々な観光スポットに関する情報源が溢れている。人々はそれらの情報を基に計画を立て、旅行に行くと考えられる。しかしながら、観光雑誌や観光サイトなどに掲載されている情報は有名な場所やありきたりな場所がほとんどである。これらの場所に何度も行くと同じ観光スポットばかりで飽きてしまう恐れがある。そこで訪れたことのない場所や、訪れたことがある場所であっても、その場所の過去の情報を知る事により新たな発見をすることができる。しかしながら、観光雑誌や観光サイト等の情報源では有名な観光スポットが多い為、このような観光スポットを発見する事は困難である。新たな観光スポットや既存の観光スポットの過去の情報を発見する手法はいくつか考えられるが、本研究では観光スポットの歴史的価値に着目し、今までの情報源では知り得ることができなかった新たな観光スポットや既存の観光スポットの過去の情報の推薦手法を提案する。具体的には旅行に行った際にユーザの周囲のランドマークを抽出し、これらの歴史的な過去の情報を検索し推薦する。

## 2. 関連研究

難波<sup>[1]</sup>らは、学習アルゴリズムブートストラップ法を使用することで、Web 上から新旧地名の抽出を行うためのパターンを大量に収集する手法を提案している。張<sup>[2]</sup>らは、Web 上の情報、郵便番号データ、位置参照情報ダウンロードサービスを用いて、過去から現在にいたる人物にアクセスするための地図インタフェースの開発を提案している。それに対して、本研究では歴史に着目した新たな観光スポットを推薦する点と、Web 上からの情報抽出の際に独自のクエリを用いることで、歴史的情報を含む Web ページを抽出する点異なる。

## 3. 提案手法の流れ

図1と以下に提案手法の処理の流れを示す。

① ユーザの入力したランドマーク名と我々が提案する定型文を用いてランドマークに関する歴史的な情報を含む Web ページ群を取得する。

② 取得した Web ページ群から本文を抽出し、形態素解析により過去のスポット名候補を取得する。

③ 過去のスポット名候補から出現回数を用いて過去のスポット名を決定する。

④ 決定したスポットの情報をユーザに提示する。

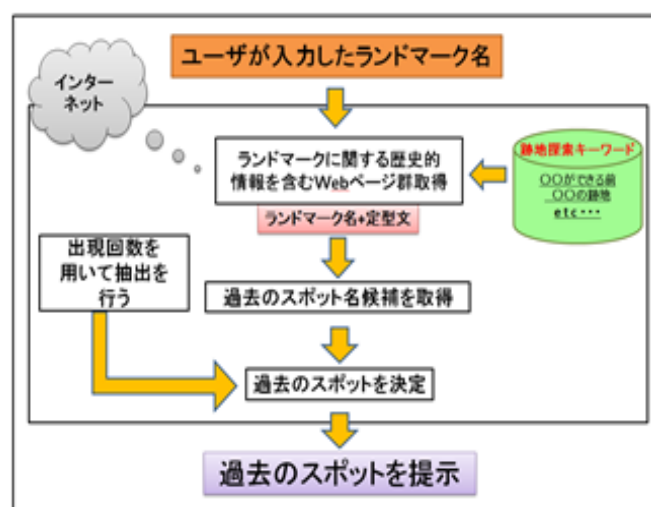


図1：システムフロー

## 4. 観光スポット推薦手法

### 4. 1. Web ページ群の抽出

「ユーザの入力したランドマーク名」と定型文を組み合わせたものをクエリとして、歴史的な情報を含む Web ページ群を取得する。定型文は歴史的な情報を含む文によく用いられるキーワードであり、手作業で決定した。この定型文を本研究では「跡地探索キーワード」と呼ぶ。跡地探索キーワードは全部で 25 種類からなる。表 1 に跡地探索キーワードの一部を示す。

表 1：定型文の例

〇〇の前は	〇〇の昔	〇〇の歴史
〇〇の過去	〇〇の前は	〇〇の概要
〇〇の謂われ	〇〇の跡地	〇〇の沿革

この跡地探索キーワードを用いて各キーワードとユーザの入力したランドマーク名を含む Web ページを取得し、本文を抽出する。

#### 4. 2. 過去のスポット名候補の抽出

「後楽園球場」や「皇居前広場」のように過去のスポット名は固有名詞を含む名詞の複合からなる場合が多い。その為、本研究では抽出した歴史的な情報を含む Web ページから固有名詞含む名詞が連続しているものを過去のスポットとする。そこで、歴史的な情報を含む Web ページを形態素解析し、名詞が連続しているものを連結し過去のスポット名候補とする。形態素解析器には juman<sup>3)</sup>を用いる。

#### 4. 3. 過去のスポットを決定

ユーザが検索に用いたランドマーク名を過去のスポット名候補から除き、残った候補内の出現回数を求める。最も多い出現回数の過去のスポット名候補を過去のスポット名と決定する。表 2 にランドマーク名が「新宿御苑」と「皇居」の例を示す。

表 2：過去のスポット名候補

ランドマーク名	
	新宿御苑
1	新宿御苑:1423
2	内藤家:215
3	名無しさん:193
4	江戸時代:192
5	1周年:182
6	四谷:170
7	日本庭園:166
8	内藤新宿:151
9	甲州街道:149
10	新宿区:130
11	玉川上水:117
12	新宿駅:105
13	専用バスルーム:102
14	電気ポット:96
15	ヘアドライヤー:94
16	冷蔵庫:94
17	お店:94
18	薄型テレビ:94
19	カーペットフロア:90
20	タオル付:90

表 2 からわかるように「新宿御苑」ならば「内藤家」、「皇居」ならば「江戸時代」が過去のスポット名となっている。

### 5. 評価実験

#### 5. 1. 実験内容

提案手法を用いて、過去のスポット名が正しく抽出できているかを評価するために無作為に選んだランドマーク 20 件を対象に実験を行った。ランドマークの条件として、現在と過去での変化が最も大きいと考えられる東京都のランドマークを用いた。

実験の手順は、実験に用いるランドマーク名と跡地探索キーワードを入力して Web ページを抽出する。得られた Web ページから提案手法を用いて、過去のスポット名を決定し、それらが正解であるか

の判定を行う。表 3 にその結果を示す。

表 3：実験結果

	現在のランドマーク名	正解の過去のスポット	提案システムの結果
1	六本木ヒルズ	森ビル	森ビル
2	両国国技館	日本大学講堂	蔵前国技館
3	北の丸公園	江戸城	江戸城
4	尾久の原公園	旭電化工業跡地	荒川区
5	飛鳥山公園	旧渋沢庭園	北区
6	乃木公園	旧乃木邸	乃木神社
7	東京都庁	淀橋浄水場	東京府
8	東京ドーム	後楽園球場	後楽園球場
9	東京タワー	紅葉館跡地	東京スカイツリー
10	東京スカイツリー	葉平橋貨物線跡	浅草寺
11	池袋サンシャイン	巣鴨プリズン	巣鴨プリズン
12	清水谷公園	大久保屋敷跡	江戸時代
13	新宿御苑	内藤家屋敷跡	内藤家
14	国会議事堂	広島藩浅野家屋敷跡	帝国議会
15	皇居	江戸城	江戸城
16	恵比寿ガーデンプレイス	エビスビール工場	恵比寿駅
17	お台場	第 00 台場	第 3 台場
18	よみうりランド	多摩丘陵	稲城市
19	井の頭公園	御殿山	吉祥寺
20	外務省	福岡藩黒田家屋敷跡	日本人

#### 5. 2. 考察

ランドマーク 20 件に対して 7 件は正確に過去のスポット名を抽出することができた。適合率は 35% である。抽出できた過去のスポット名とできなかった過去のスポット名を比較すると知名度に差があり、その差によって抽出できるかが変わるのでないかと考えられる。また、正しく抽出できた過去のスポット名の結果を見ると、過去のスポット名が多数の固有名詞と共起している事がわかった。その為、係り受け解析を行うことで、精度の向上につながるのではないかと考えられる。

#### 6. まとめと今後の課題

本研究では、歴史に着目し跡地探索キーワードを用いることで、歴史的な情報を含む Web ページを取得し、現在のランドマークから過去のスポット名を抽出する手法の提案を行った。今後の課題として、過去のスポット名候補に係り受け解析を行うことで、提案手法の精度が向上するかの実験を行い、抽出した過去のスポット名を地図上にマッピングし、リアルタイムで観光スポット情報の推薦を行う研究を進める。

#### 参考文献

- [1]平山拓実, 難波英嗣, 竹澤寿幸, “新旧地名・施設名対の抽出による文書の時空間マッピング” DEIM Forum 2016 B1-4
- [2]張鋼, 村上晴美, “昔の住所を持つ人物の地図上への表示” FIT2015(第 14 回情報科学技術フォーラム)
- [3] JUMAN: <http://nlp.ist.i.kyoto-u.ac.jp/>