

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

上原 尚[†] 嶋田 和孝[†] 遠藤 勉[†]

[†]九州工業大学情報工学部

〒 820-8502 福岡県飯塚市川津 680-4

E-mail: †{h_uehara,shimada,endo}@pluto.ai.kyutech.ac.jp

あらまし 本論文では、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する。本システムでは、入力をユーザのお気に入りの観光地とすることにより、観光地の幅広い特徴分析に対応し、同時にユーザの嗜好解析も実現する。観光地の特徴ベクトルは (1) 知恵袋・ブログ上での共起キーワードと (2) 時系列分布 (3) 知恵袋上でのカテゴリ構造 (4) 観光地周辺施設 (5) 地図画像から生成する。これらの特徴ベクトルからベクトル空間モデルの考え方のもと、コサイン類似度の算出を行い、類似性の高い観光地を推薦する。さらにどこが特徴的なのかを明確に表現するため、差分の可視化にも取り組む。キーワード 観光地推薦システム, 集合知, 類似度

Sightseeing location recommendation using tourism information on the Web

Hisashi UEHARA[†], Kazutaka SHIMADA[†], and Tsutomu ENDO[†]

[†] Faculty of Information, Kyushu Institute of Technology

Kawazu 680-4, Iizuka-shi, Fukuoka, 820-8502 Japan

E-mail: †{h_uehara,shimada,endo}@pluto.ai.kyutech.ac.jp

Abstract This paper describes a prototype system of sightseeing location recommendation. It integrates user's need and characteristics of each sightseeing location for the recommendation. The input of the system is a user's favorite location or facility. The system searches some locations related to the user's favorite on the basis of some similarity measures. We use five similarity measures; (1) co-occurrence keywords, (2) time sequence, (3) category information on Yahoo Chiebukuro, (4) surrounding area information and (5) map images. Our method visualizes the output of the recommendation.

Key words Sightseeing recommendation, Collective Intelligence, Similarity

1. ま え が き

近年、スマートフォンやノートパソコンの普及により、容易かつ迅速に大量の情報を得ることができ、なおかつ情報発信も行えるようになってきている。特に、観光情報ではそれが著しく、各地の観光施設や公共団体から公開されているものもあれば、旅行経験者によるブログの投稿、Q & A サイトでのお勧め観光地の質問・回答など、多種多様な観光情報が存在する。しかし、そこから得られる情報量が膨大であり、どの観光地が自分の嗜好にあっているかなどの、ユーザにとって正確な情報を素早く見つけ出すことは困難である。そのような背景から、観光地推薦システムの開発が盛んに行われている [1,2]。例えば、ユーザの旅行履歴と観光地の特性から観光地を推薦するシステ

ム [3] や、ユーザの観光動機・旅行目的から観光地を紐付けている観光コース自動作成システム [4] などがある。これらのシステムのようにユーザの嗜好を的確に捉えることは、推薦システムとして重要な項目の一つである。しかし、ユーザの嗜好に焦点を置いてしまうと、観光地の特徴表現の多様性に制約をかける結果になってしまう。ユーザの嗜好を考慮しながらも、観光地の特徴を最大限に引き出し、ユーザに提示することも推薦システムとして必要である。

そこで、本論文では、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する。本システムでは、入力をユーザのお気に入りの観光地とする。これにより、観光地の幅広い特徴分析に対応でき、なおかつ、ユーザの嗜好解析にも繋がる。

特定の特徴に依存しない柔軟な推薦システムを実現するため、各特徴から複数の独立した類似度を算出・評価することで観光地間の類似性を表現する。Web からの情報源として、日本版 Wikipedia^(注1)、Yahoo 知恵袋^(注2)（以下、知恵袋と呼ぶ）、ブログ、地図画像を用いる。観光地の特徴ベクトルは（1）知恵袋・ブログ上での共起キーワードと（2）時系列分布（3）知恵袋上でのカテゴリ構造（4）観光地周辺施設（5）地図画像から生成する。本研究では、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、さらにどこが特徴的なかを明確に表現するため、差分の可視化にも取り組む。本システムの概要図を図 1 示す。

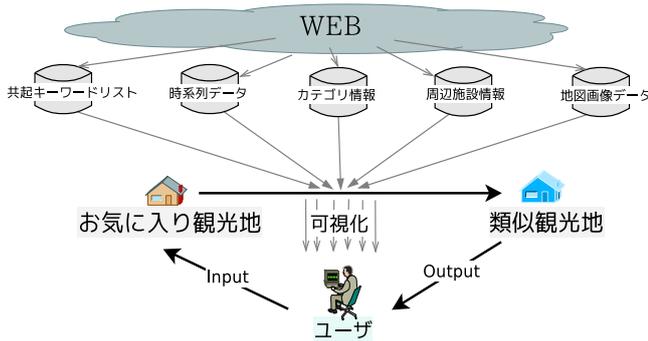


図 1 観光地推薦システムの概要図

2. 観光地の特徴表現

本研究では、観光地に関する 5 つの特徴表現を利用する。本節では、それらについて説明する。

2.1 共起キーワードリスト

共起キーワードリストの定義として、検索クエリを観光地名とした時に、Web から取得した文章中出现するキーワード（以下、共起キーワードと呼ぶ）に対して重要度を付加した集合体とする。本手法では、知恵袋とブログを対象に、観光地の共起キーワードを抽出し、重要度を算出することで、共起キーワードリストを作成する。知恵袋の情報は Yahoo 知恵袋質問検索 API^(注3)、ブログ情報は Yahoo ブログ検索 API^(注4)から取得する。知恵袋では、質問・回答が特定のカテゴリに分類されており、ある質問に対して最良の回答をベストアンサーと定義している。この知恵袋の特性に注目し、本手法では、知恵袋で最上層に位置する「地域、旅行、お出かけ」カテゴリ（以下、地域カテゴリと呼ぶ）内を対象に、任意の観光地を検索クエリとし、その際にヒットした質問のベストアンサーを活用する。また、ブログ情報ではヒットしたブログタイトル・サマリーを使用する（以下、これらから取得した文章のことを、取得文章と呼ぶ）。図 2 は、知恵袋のカテゴリ構造を大まかに示したものである。

(注1): 「日本版 Wikipedia」, <http://ja.wikipedia.org/wiki>

(注2): 「Yahoo 知恵袋」, <http://chiebukuro.yahoo.co.jp>

(注3): 「Yahoo 知恵袋質問検索 API」, <http://developer.yahoo.co.jp/webapi/chiebukuro/chiebukuro/v1/questionsearch.html>

(注4): 「Yahoo ブログ検索 API」, <http://developer.yahoo.co.jp/webapi/search/blogsearch/v1/blogsearch.html>

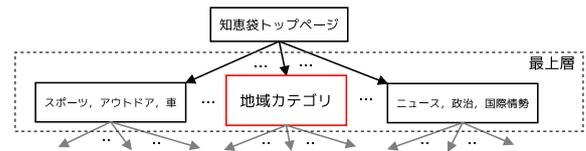


図 2 知恵袋のカテゴリ構造

キーワード抽出処理では、形態素解析器「MeCab」^(注5)とキーワード自動抽出用 Perl モジュール「TermExtract」^(注6)を使用することで、取得文章からキーワード抽出処理とキーワードの出現頻度に応じたランク付け処理を行う。しかし、これらを用いて得られたキーワードは、日本語として成立しない語や、観光とは全く関わりのない語までも含まれており、これらノイズの削除が必要となる。そこで、観光に関連したキーワードを集約した観光キーワード辞書の作成を行い、それを活用することで、ノイズ除去と、共起キーワードの重要度算出を行う。

日本版 Wikipedia の見出し語約 100 万件を記述したテキストファイルをもとに、観光キーワード辞書の構築を行う。観光キーワード辞書の作成手順として、まず最初に、英数字・記号だけで構成されている語は不要語とみなし、それらを取り除く。次に、残ったキーワード一つ一つに対して、知恵袋の地域カテゴリ内で検索を行い、ヒットしたキーワードと、そのカテゴリ内での総ヒット数を観光キーワード辞書に登録する。最後に、この登録されたキーワード一つ一つに対して、知恵袋の全カテゴリを対象に検索を行い、その総ヒット数も取得し、観光キーワード辞書に登録する。ここまでの手順を踏むことで、観光キーワード辞書の構築が完了する。今回、観光キーワード辞書に登録されたキーワードは 67644 語である。ここで、あるキーワード k に対する地域カテゴリの総ヒット数を L_k 、全カテゴリの総ヒット数を A_k とした時、キーワードが観光にどのくらい関連しているかどうかを示す度合い S_k （観光関連度）を、

$$S_k = L_k / A_k \quad (1)$$

と定義する。また、観光キーワード辞書のデータ構造とそのデータの例を表 1 に示す。

表 1 観光キーワード辞書のデータ構造とそのデータの例

キーワード k	地域カテゴリ L_k	全カテゴリ A_k
福岡	27647	164219
福岡タワー	212	325
明太子	450	4566
九州工業大学	2	711
東京	129547	886994
東京スカイツリー	2171	3984

ノイズ除去処理として、取得文章から得られたキーワードと、観光キーワード辞書内のキーワードとの文字列照合を行う。マッチしない場合は、不適切なキーワードと見なしてそのキー

(注5): 「形態素解析器 MeCab」, <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

(注6): 「キーワード自動抽出用 Perl モジュール TermExtract」, <http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html>

ワードを除外する．マッチングしたキーワードは共起キーワードとする．最後に，事前に取得してある共起キーワード k の出現頻度をもとに付加された値 F_k と，観光関連度の値 S_k から重要度 I_k の算出 (2) を行い，これをもとに共起キーワードリストを作成する．共起キーワードリスト作成が完了するまでの全体の流れを図 3 に示す．また，観光地を「東京スカイツリー」とした時に，ブログと知恵袋それぞれから取得した共起キーワードリストの一部を図 4 に示す．

$$I_k = F_k \times S_k \quad (2)$$

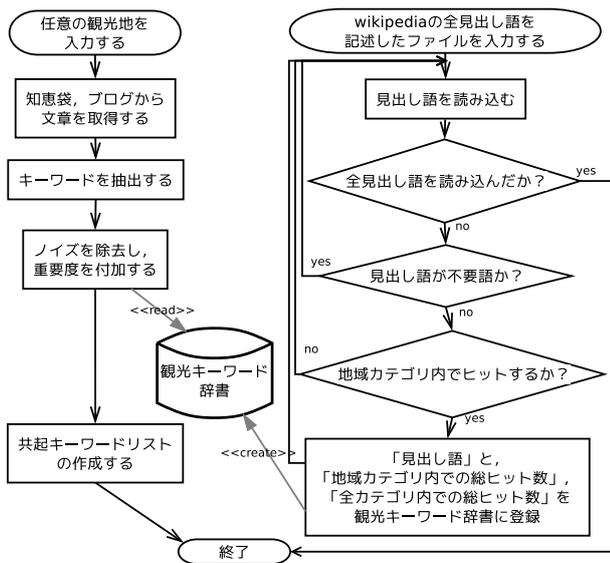


図 3 共起キーワードリスト作成までの全体フロー図

関連キーワードリスト (yahoo知恵袋)		関連キーワードリスト (yahooブログ)	
Rank	Keyword	Rank	Keyword
1	スカイツリー	1	スカイツリー
2	東京	2	東京
3	浅草	3	東京タワー
4	レジャー	4	東京ソラマチ
5	エンタメ	5	かめそば
6	台場	6	東京ディズニーシー
7	東京タワー	7	ディズニーシー
8	ホテル日航東京	8	東京ディズニーランド
9	ソラマチ	9	ソラマチ
10	アトラクション	10	展望台

図 4 「東京スカイツリー」の共起キーワードリストの一部

2.2 時系列データ

次に時系列データの有効性について議論する．観光地にはその観光地が活性化する時期のパターンという特徴があると考えられる．そこで，本システムでは対象となる観光地に関する投稿の時系列データを類似性判断のための特徴として利用する．ここでも，2.1 節同様に，知恵袋・ブログを情報源とする．知恵袋の情報からは，任意の観光地を検索クエリとした際にヒットした質問の投稿時刻を，ブログ情報からはヒットしたブログの更新時刻を，それぞれ最大で 1000 件の計 2000 件までの時系列データを収集する．

本研究では，その時系列データを，2 つのパターンで解析する．1 つ目の解析手法は，2004 年前期～2012 年後期の計 18 通りに分類する手法である．ここで，前期は 1 月～6 月，後期は 7 月～12 月と定義する．2 つ目の解析手法は，1 月前半～12 月後

半の計 24 通りに分類する手法である．ここで，前半は 1 日～15 日，後半は 16 日～31 日と定義する．観光地の活性化時期の事例として，「清水寺 (京都府)」と「天王寺公園 (大阪府)」とした時に取得した時系列データの一部をそれぞれ図 5 に示す．

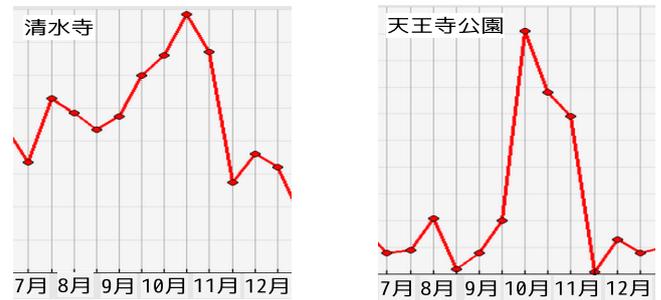


図 5 清水寺 (左) と天王寺公園 (右) の時系列データの一部

清水寺・天王寺公園ともに紅葉スポットとして有名な観光地である．その影響が，図 5 のグラフにも明確に現れており，10 月～11 月が著しく活性化している．これらのように，時系列データを用いることで，各時期からみた観光地の活性化の度合いを特徴として捉えることができる．

2.3 カテゴリ情報

カテゴリ情報は，知恵袋上でのカテゴリを使用する．2.1 節では，主に地域カテゴリに重点を置いていたが，ここでは，それ以外の最上層のカテゴリ (例えば，図 2 に示したようなカテゴリ) に着目する．これにより，観光地に対して幅広いジャンルからアプローチをはかることで，観光の属性とは違う視点からの特徴抽出を行う．手法としては，任意の観光地を検索クエリとして，知恵袋の地域カテゴリ以外の各カテゴリ内で検索を行い，それぞれの総ヒット数を取得する．観光地を「東京タワー (東京都)」にした時に，取得したカテゴリ情報を図 6 に示す．

ジャンル	件数		件数
教養と学問、サイエンス	508	生き方と恋愛、人間関係の悩み	144
スポーツ、アウトドア、車	42	ビジネス、経済とお金	37
エンターテインメントと趣味	424	コンピュータテクノロジー	3
ニュース、政治、国際情勢	402	子育てと学校	17
マナー、冠婚葬祭	24	職業とキャリア	20
暮らしと生活ガイド	114	おしゃべり、雑談	2
健康、美容とファッション	17	インターネット、PCと家電	305

図 6 「東京タワー」のカテゴリ情報

東京タワーのカテゴリ情報をみると，「教養と学問，サイエンス」カテゴリ内でのヒット数が多いことが分かる．これは，建築学の視点で東京タワーの構造の話題が頻繁に取り上げられているからである．類似した観光地を推薦する場合，このような点にも着目することで，より明確に類似性を評価することができる．

2.4 周辺施設情報

周辺施設情報とは，観光地の周辺 5km 以内にある施設のジャンル情報のことを定義する．本手法では地図データから周辺施設情報を取り出し特徴抽出を行う．地図データは，Yahoo ローカルサーチ API^(注7) (以下，ローカル API と呼ぶ) を活用する．

(注7): 「Yahoo ローカルサーチ API」, <http://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/localsearch.html>

ここでは、観光地の位置情報を入力として計 58 種類の周辺施設情報の取得を行う。東京スカイツリーの位置情報を対象に取得した周辺施設情報の一部を図 7 に示す。

和食	4356	ドラッグストア、市販薬	839
洋食	1356	家電、携帯電話	976
バイク	25	百貨店、ショッピングセンター	132
中華	849	コンビニ、スーパー	1293
アジア料理、エスニック	351	リサイクル、ディスカウントショップ	366
ラーメン	579	生活用品、インテリア	2526
カレー	113	趣味、スポーツ、工芸	642
焼肉、ホルモン、ジンギスカン	515	ファッション、アクセサリ、時計	2738
		食品、食材	1636

図 7 「東京スカイツリー」の周辺施設情報の一部

東京スカイツリーは、東京都の中でも比較的都心部に位置している。このような情報は、図 7 であげているような施設のヒット数に強く関連しており、この周辺施設情報を用いることで、観光地の地域情報を捉えることが可能となる。

2.5 地図画像データ

ここでも 2.4 節と同様に、事前に観光地の緯度経度の情報を取得しているものとする。ここで取り扱う地図画像としては、標準的な地図画像（以下、標準マップと呼ぶ）と、ミッドナイトモードの地図画像（以下、ナイトマップと呼ぶ）の 2 つのパターンを使用する。標準マップとナイトマップのそれぞれの特性は、最適な観光地特徴表現のツールとなる。標準マップでは、海や川、山などのように地理的な要素の違いで色合いが大きく変化する。それに対して、ナイトマップでは、周辺の建物・交通機関の数の変化で明暗の差がはっきりと分かれる。観光地を「金閣寺（京都府）」「福岡タワー（福岡県）」とした時の 2 パターンの地図画像をそれぞれ図 8 と図 9 に示す。



図 8 「金閣寺」(左：標準マップ, 右：ナイトマップ)



図 9 「福岡タワー」(左：標準マップ, 右：ナイトマップ)

金閣寺と福岡タワーは、地図画像のグラデーションの差が著しい。この差は、地理的条件から生じているもので、2.4 節で述べた周辺施設情報からは把握できない周辺情報である。また、これはユーザの嗜好にも大きく影響してくる要素（例えば、緑豊かな場所や海が近い場所など）でもあり、観光地の特徴表現として重要となるポイントであると考えられる。そこで、本研

究では地図画像の色情報の抽出を行う。ここで言う色情報は、HTML 色で表現されるカラーコードとその使用回数のことを指す。地図画像データは、yahoo スタティックマップ API^(注8)から取得する。色情報抽出の際には、PHP のクラスライブラリ ColorExtract^(注9)を活用する。

3. ベクトル空間モデル

本節では、ベクトル空間モデルの考え方のもと、2 節で取り上げた観光地の特徴表現を、特徴ベクトルに適用し、そのベクトルを用いた類似度計算処理について述べる。以下、対象とする観光地を s で表現する。

共起キーワードリストの特徴ベクトル（以下、共起キーワードベクトルと呼ぶ）への適用について述べる。共起キーワードベクトル K_s は、

$$K_s = \{k_{s_1}, k_{s_2}, \dots, k_{s_{n_k}}\} \quad (3)$$

と表現できる。ここで、 n_k は観光キーワードリストの登録されているキーワード数 67644 語となる。共起キーワードベクトル K_s 内の各要素は、各共起キーワードの重要度の値が入る。

時系列データの特徴ベクトル（以下、時系列ベクトルと呼ぶ）への適用について述べる。時系列ベクトル T_s は、

$$T_s = \{t_{s_1}, t_{s_2}, \dots, t_{s_{n_t}}\} \quad (4)$$

と表現できる。ここで、 n_t は、時系列データの特徴パターンである 2004 年前期～2012 年後期の計 18 通りと、1 月前半～12 月後半の計 24 通りの合計 42 通りとなる。時系列ベクトル T_s 内の各要素は、各時期でのヒット数の値が入る。

カテゴリ情報の特徴ベクトル（以下、カテゴリベクトルと呼ぶ）への適用について述べる。カテゴリベクトル C_s は、

$$C_s = \{c_{s_1}, c_{s_2}, \dots, c_{s_{n_c}}\} \quad (5)$$

と表現できる。ここで、 n_c は、知恵袋上での地域カテゴリを除く最上位のカテゴリ数 14 となる。カテゴリベクトル C_s 内の各要素は、各カテゴリ内でのヒット数の値が入る。

周辺施設情報の特徴ベクトル（以下、周辺施設ベクトルと呼ぶ）への適用について述べる。周辺施設ベクトル E_s は、

$$E_s = \{e_{s_1}, e_{s_2}, \dots, e_{s_{n_e}}\} \quad (6)$$

と表現できる。ここで、 n_e は、施設情報の第二階層ジャンル数 58 となる。周辺施設ベクトル E_s 内の各要素は、観光地 s の周辺 5km 以内で検索した時のヒット数の値が入る。

地図画像データの特徴ベクトル（以下、マップベクトルと呼ぶ）への適用について述べる。マップベクトル M_s は、

$$M_s = \{m_{s_1}, m_{s_2}, \dots, m_{s_{n_m}}\} \quad (7)$$

と表現できる。ここで、 n_m は、カラーコードのパターン数 16⁶

(注8): 「yahoo スタティックマップ API」, <http://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/static.html>

(注9): 「ColorExtract」, <http://www.phpclasses.org/package/3370-PHP-Extracts-the-most-common-colors-used-in-images.html>

となる．マップベクトル M_s 内の各要素は，カラーコードの使用回数の値が入る．

本研究では，類似度計算処理として，ベクトル空間モデルのコサイン類似度の算出を行う．2つのベクトル $A = \{a_1, a_2, \dots, a_n\}$ と，ベクトル $B = \{b_1, b_2, \dots, b_n\}$ のコサイン類似度 R_{AB} は，

$$R_{AB} = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}} \quad (8)$$

と定義されている．5つの特徴ベクトルをこの数式(8)に当てはめることで，観光地間の類似度を評価する．

4. 観光地データベースの構築

本システムを実現するにあたって，観光地の情報を収集した観光地データベースが必要となる．そこで，2.1節で作成・活用した観光キーワード辞書から観光地を抽出することで観光地データベースを構築する．この観光地抽出処理では，2.4節でも述べたローカルサーチAPIを活用する．ローカルサーチAPIでは，任意のキーワードを入力とした場合，そのキーワード名と関連性の高い施設情報を取得することができる．これを利用することで，観光地の抽出を実現する．また，その際に位置情報(緯度経度・都道府県名)も取り出し，これらを観光地データベースに登録する．これにより，4507件の観光地を登録した観光地データベースの構築が完了する．最後に，観光地データベースに登録されている観光地情報を対象に，2節に基づき特徴情報を全て収集し，観光地データベースに追加登録することでデータベースの再構築を行う．

5. 観光地推薦システム

本観光地推薦システムは，Webアプリケーションとして設計し，ユーザのお気に入りの観光地とその都道府県名を入力とすることで，最適な観光地を推薦する．本節では，入力から観光地を推薦するまでのシステムの処理の流れと，差分の可視化について述べる．

5.1 観光地の推薦

本システムのトップ画面を図10に示す．

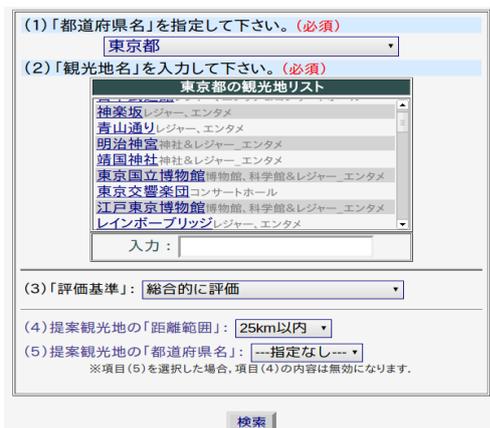


図10 観光地推薦システムのトップ画面

図10の項目(1)と，項目(2)を，本システムの基本クエリとする．項目(1)で「都道府県名」を指定，項目(2)でその都道府県内にある観光地を入力し，検索ボタンをクリックすることで，周辺(デフォルトでは25km以内)にある観光地を類似度が高い順にユーザに出力する．本システムでは，項目(2)で観光地リストを表示することで，ユーザがスムーズに入力を行えるように入力支援も行う．検索ボタンをクリックしてから，ユーザに観光地を出力するまでのシステムの処理フローを図11に示す．

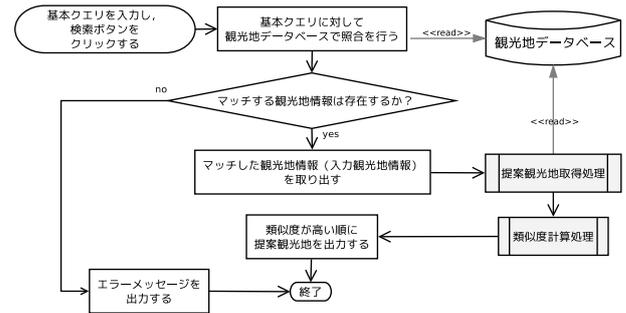


図11 基本クエリから出力までの処理フロー

本システムでは，オプションとして，図10の項目(3)の「評価基準」を指定することで，類似度計算処理の設定をユーザ自身でカスタマイズできる．評価基準の項目は，「共起キーワードリスト」「カテゴリ情報」「時系列情報」「周辺施設情報」「地図画像データ」の5種類あり，これらは2節で述べた特徴表現に対応している．また，項目(4)を選択することで，提案観光地の周辺距離の範囲を指定することができ，項目(5)の都道府県名を選択することで，提案観光地の都道府県を指定することができる．類似度計算処理フローを図12に示し，提案観光地取得処理フローを図13に示す．

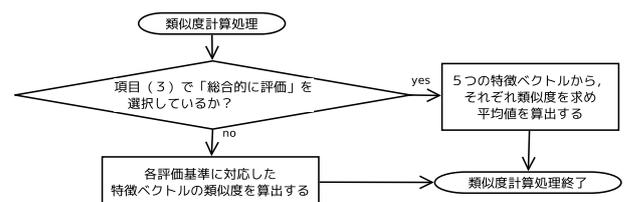


図12 評価基準の選択項目(上)と類似度計算処理フロー(下)

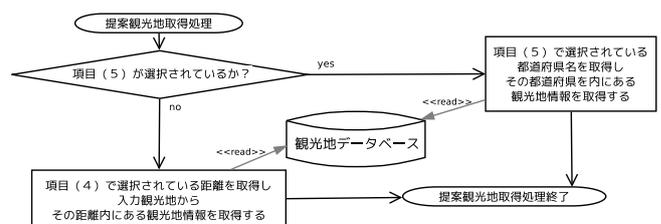


図13 提案観光地取得処理フロー

出力結果の例として，項目(1)に「東京都」を指定し，項目(2)で「東京スカイツリー」，項目(3)で「総合的に評価」，項目(5)で「福岡県」を指定した時の出力結果の一部を図14に示す．

1 2 3 4 5 6 7 8 次へ >>			
[1]	福岡タワー	[福岡県]	【類似度:0.543】 レジャー、エンタメ
[2]	アイランドシティ	[福岡県]	【類似度:0.439】 レジャー、エンタメ
[3]	福岡市動植物園	[福岡県]	【類似度:0.435】 レジャー、エンタメ

図 14 入力を「東京スカイツリー」で福岡県内を対象とした出力結果の一部

5.2 差分の可視化

本システムでは、観光地を推薦するだけでなく、入力した観光地と推薦した観光地の差分の可視化も行う。差分の可視化を行う画面への遷移は、推薦観光地をクリックすることで実現する。ここで、差分の可視化の初期画面を図 15 に示す。

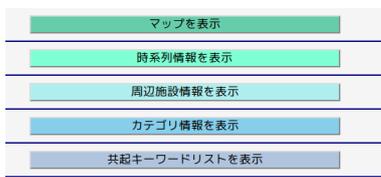


図 15 差分可視化の初期画面

図 15 の「マップを表示」を選択した場合は、2つの観光地の標準マップ・ナイトマップを表示する。「時系列情報」を選択した場合は、両観光地の時系列データから生成した折れ線グラフを重ね合わせて表示する。「周辺施設情報」「カテゴリ情報」では、両観光地のデータから生成したレーダーチャートを重ね合わせて表示する。「共起キーワードリスト」では、両観光地のキーワードリストに対して、両者共通のキーワードに色の変化を加えて表示する。

ここで、観光地間の差分の可視化の事例として「広島平和記念資料館（広島県）」と「ひめゆりの塔（沖縄県）」を取り上げて説明する。これらのカテゴリ情報のレーダーグラフを図 16（塗りつぶされている方が「ひめゆりの塔」）に示す。

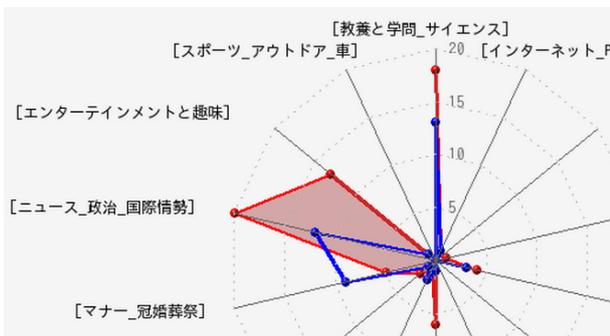


図 16 広島平和記念資料館とひめゆりの塔のカテゴリ情報

広島平和記念資料館は広島原爆の惨状を後世に伝えるための施設として開館した施設であり、ひめゆりの塔は沖縄戦末期に置かれた慰霊碑のことを指す。このように両者ともに歴史的に似たような背景をもっている。この影響による両者の類似性は、図 16 のカテゴリ情報での「ニュース、政治、国際情勢」「教養と学問、サイエンス」カテゴリへの傾きから把握することができ

きる。

次に、両者の時系列データをグラフ化したものを図 17 に示す。

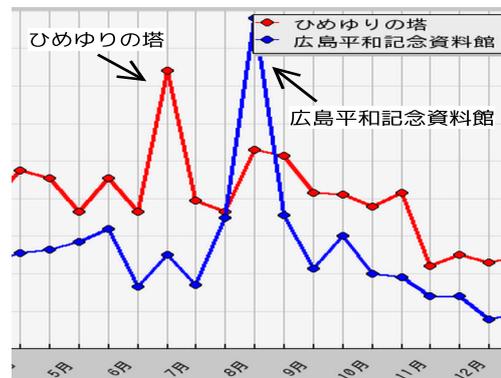


図 17 広島平和記念資料館とひめゆりの塔の時系列データ

図 17 より、両者の観光地の活性化時期にズレが生じているのが見て分かる。これは、お互いがそれぞれ関与している歴史的な詳細要素が関係している。ひめゆりの塔は、沖縄終戦日である 6 月 23 日に強く影響を受けており、それに対して、広島平和記念資料館は、広島に原爆が投下された日の 8 月 6 日の影響を受けている。この要素は、類似性の中での差分を明確に示している部分であると言える。これらのように、差分を可視化することで、類似・非類似の境界を際立たせると同時に、観光地自身が持っている特徴を最大限に引き出すことができる。

6. まとめ

本論文では、Web 上に混在する観光情報を抽出し、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。特徴ベクトルは (1) 知恵袋・ブログ上での共起キーワードと (2) 時系列分布 (3) 知恵袋上でのカテゴリ構造 (4) 観光地周辺施設 (5) 地図画像から生成した。観光地間の類似度計算処理では、ベクトル空間モデルの考え方のもと、特徴ベクトル間のコサイン値を算出することで実現した。また、本観光地推薦システムの基盤となる観光地データベースの構築にも取り組んだ。さらにどこが特徴的なのかを明確に表現するため、観光地間の差分の可視化にも取り組んだ。

今後の課題として (1) 観光地の新たな特徴表現の検討 (2) 利用者によるアンケート調査の実施などが挙げられる。

文献

- [1] 川村秀憲, 鈴木恵二, 山本雅人, 松原仁: 観光情報学, 情報処理学会デジタルプラクティス, Vol.3, No.4, October 2012
- [2] 斎藤 一: Web における観光情報提供と分析, 人工知能学会誌, Vol.26, No.3, 2011/5
- [3] 樽井勇介: 協調フィルタリングとコンテンツ分析を利用した観光地推薦手法の検討, 上武大学経営情報学部紀要, 2011, 第 36 号, p.1-14
- [4] 倉田陽平, 奥貫圭一, 貞広幸雄 (2000): 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, 9, 199-202