

観光情報集約のためのユーザ検索行動に基づく スマートスクラップブックシステム

峯 祥平[†] 北山 大輔[†]

[†] 工学院大学情報学部コンピュータ科学科 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: tj110117@ns.kogakuin.ac.jp, ††kitayama@cc.kogakuin.ac.jp

あらまし 一般に、ユーザは観光の際に目的とする観光オブジェクトについての情報を Web やガイドブックを用いて収集する。この時、まとめる必要がある多数のオブジェクトの情報を収集し、あるオブジェクトの情報について調べるとき、ユーザによっては地図で場所を得たり、Web で詳細を調べるなど、定型的な行動をとることがある。このような定型的な行動を他のオブジェクトに対しても適応することで、手軽に観光情報を収集するスマートスクラップブックを提案する。

キーワード 情報集約, スマートスクラップブック, 観光情報, 検索行動パターン

1. はじめに

一般に、ユーザは観光の際に目的とする観光オブジェクトについての情報を Web やガイドブックを用いて収集することが多い。近年では、スマートフォンのアプリで「地球の歩き方」が出るなど、ガイドブックの電子化が進んでいる。あるオブジェクトの情報についてガイドブックの情報では不足している時、ユーザはオンライン地図で場所を得たり、Web 検索で詳細を調べるなどの行動をとる。この時、ユーザによっては複数のオブジェクトの情報に向けて、図 1 のようにスクラップブックを作成することがある。こうすることで自分専用の旅行のしおりを持つことができ、観光旅行へのモチベーションを高めることができる。しかしながら、Web やガイドブックによる情報収集が容易となっても、観光オブジェクトごとに必要な情報を収集し、スクラップブックを作成する場合、オブジェクトが増えるにつれて、その手間は増大する。

このようなスクラップブックを作成する際、地図ならば、最寄り駅や中心的な駅からの位置関係を表示、又、Web 情報ならオブジェクトの詳細情報や見どころといった特定のパターンに当てはまる情報をスクラップブックに追加すると考えられる。さらに、あるオブジェクトに対して追加した情報であれば別のオブジェクトに対しても類似する情報を追加する可能性は高いと考えた。そこで、地図で場所を得たり、Web で詳細を調べるなどの行動を定型的な行動とし、定型行動を他のオブジェクトに対しても適応することで、手軽に観光情報を収集できるスマートスクラップブックシステムを考えた。

例えば、ホテルに対して、ある駅との位置関係を示す地図をスクラップブックに追加するようなユーザは、他の観光オブジェクトに対しても同様にある駅との位置関係を示す地図を追加する可能性は高いといえる。

以下、本論文の構成を示す。まず 2 節では本研究の概要と関連研究について説明する。3 節では Web 操作パターン抽出について説明する。4 節では地図操作パターン抽出について説明

する。5 節ではそれぞれについての考察を行なう。

2. 本研究のアプローチと関連研究

2.1 研究の概要

本研究では、観光オブジェクトの情報収集をする際の定型的な行動を他のオブジェクトに対しても適応することで、手軽に観光情報を収集するスマートスクラップブックを提案する。他のオブジェクトに適用する定型的な行動の抽出方法として Web 操作パターン抽出と地図操作パターン抽出の 2 つがある。

Web 操作パターン抽出とはユーザがあるオブジェクトに関して手動で Web 検索結果をスクラップブックに追加した場合にそれを他のオブジェクトに適応させる手法である。ユーザが観光オブジェクトについて Web 検索を行い、必要な Web ページを選択したとする。選択されたページにはそれ以外のページにはない特徴が含まれていると仮定し、特徴を表すキーワードを抽出する。キーワード含有量と検索結果内のページの順位という指標を利用し、他の観光オブジェクトのページをスクラップブックに自動的に追加する。

地図操作パターン抽出とはユーザがあるオブジェクトに関して地図検索及び操作をし、スクラップブックに追加した場合にそれを他のオブジェクトに適応させる手法である。ユーザが観光オブジェクトについて地図検索を行い、ある地点に対してセンタリングやズームイン、パンなどの地図操作をし、地点が領域内に含まれるように地図を保存したとする。地点にはユーザの興味があると推定でき、興味を表すオブジェクトとして抽出する。他の観光オブジェクトにこの操作行動を適用する際、観光オブジェクトと地点が含まれる地図をスクラップブックに自動的に追加する。

2.2 関連研究

ある検索行動を別の検索行動に反映する研究について述べる。加藤ら [1] は入力された検索キーワード A とキーワード B との関係 R を抽出し、キーワード C が入力された時に関係 R を用いてキーワード D を出力するという、情報検索手法を提案し

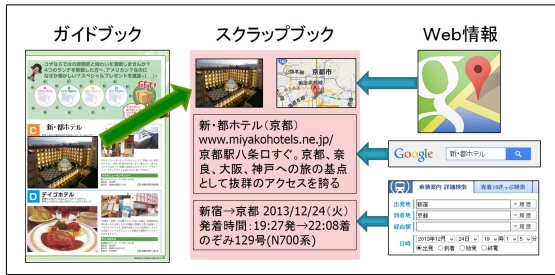


図1 ガイドブックとスクラップブックの関係図

ている。抽出したパターンを元に関係性から類似したワードに置き換え適用するといった点で本研究と類似しているが、本研究では、入力キーワードのみではなく、その検索結果に対するユーザの検索操作パターンから関係性を抽出するといった点で手法が異なる。

森ら[2]は他ユーザの検索ノウハウを自分の検索目的に適用する手法を提案している。ユーザの検索操作パターンを他のオブジェクトに対しても適用するといった点では類似しているが、森らは、検索クエリに着目し、その追加や削除、変更などの複数の操作からノウハウの生成を行っているのに対し、我々はユーザのページ選択行動からパターンを抽出する点で異なる。

中島ら[3]は相対的問い合わせとその質問処理方法である相対的マッピング処理方法を提案している。相対的評価に基づいて情報検索及び抽出を行い人間の主観に適した内容を発見するという点では類似しているが、本研究では、ユーザのWeb操作における順位とスニペットの内容から類似した条件を持つ、他オブジェクトの情報を抽出する点で異なる。

平元ら[4]は地図操作からユーザの意図を抽出し、その意図にあったWeb情報を提示するシステムを構築している。我々はこの研究における地図操作パターンからの意図抽出手法を用いて、地図操作パターン抽出を行っている。又、スクラップブックに適応するための地名を抽出するために拡張を行う。

石野ら[5]は旅行ガイドブックに対し、旅行ブログエントリや質問応答コンテンツを自動的に対応付けることで旅行ガイドブックの情報拡張を行い閲覧できるシステムを構築している。ガイドブックのみでは不足している情報を拡張するといった点やガイドブック、ブログ、質問応答コンテンツといったメディアを横断するシステムを利用するという点で類似しているが、本研究ではユーザの操作に基づき必要な情報を自動的に追加していくという点で異なる。

また、観光に関する先行研究は多々存在する。倉島ら[6]は行動範囲と興味情報に基づいて次の行動を予測し推薦するシステムを構築している。倉田ら[7]は提示した推薦プランに対するユーザの反応を元にプランを改訂し、これを繰り返すことで徐々にユーザの嗜好に適したものと改善していくシステムを構築している。上原ら[8]はWeb上に混在する観光情報からデータベースを構築し、ユーザのお気に入りの観光地を入力とし、複数の特徴ベクトルから観光地間の類似性を評価し推薦するシステムを構築している。いずれも観光地の推薦が目的であり、本研究は観光地の関連情報の抽出を目的としていることで

手法も異なる。

3. Web操作パターン抽出

本節では、ユーザがあるオブジェクトに関して手動でWeb検索結果をスクラップブックに追加した場合にそれを他のオブジェクトに適応させる手法を説明する。ユーザが観光オブジェクトAについてWeb検索を行い、必要なWebページを選択したとする。この時、検索結果の一覧から特定のページを選択したことに何らかの基準が存在すると考えた。我々はその基準をコンテンツの特徴と、検索結果の順位に依存すると考えた。コンテンツの特徴としては、選択されたページにはそれ以外のページにはないキーワードが含まれていると仮定し、TF-IDF法を用いてそのページに含まれる特徴的なキーワードを抽出する。TF値は選択されたページのスニペット内の単語の出現回数であり、DF値は検索結果内に対象の単語が含まれるページ数である。また、検索結果のページの順位に関しては、検索結果の上位に表示されるページを選択したユーザは、スニペットの内容を読まずに直感的にページの選択を行う可能性があり、下位を選択するユーザは、コンテンツの特徴によって選択したと考えられる。他の観光オブジェクトBのスクラップブックを作成する際、観光オブジェクトBの検索結果からキーワード特徴及び順位の特徴に合致するページを自動的に追加する。選択されたページのスニペット内の全ての単語ごとにTF-IDF値を求め、新たな観光オブジェクトのスクラップブック作成において検索結果中でTF-IDF値の合計が一番高いスニペットを持つページを保存する。なお、その際にページの順位も考慮する。検索結果スニペット中に全ての特徴語が含まれない場合は保存しない。

以下の数式によりページの合致度を算出する。

$$Score(r_t) = \alpha \times f(r_t) \times \frac{1}{f_m} + (1 - \alpha) \times \frac{1}{r_t} \quad (1)$$

$$f(r_t) = \sum_{x \in X(r_u)} TFiDF(x) \quad (2)$$

$$\alpha = \frac{r_u - 1}{2(s - 1)} + \frac{1}{2} \quad (3)$$

式中の $Score(r_t)$ はシステムが r_t ページ目を保存した際のスコアである。 α は $0.5 \leq \alpha \leq 1.0$ で、ユーザが s 個中 r_u 件目のページを選択した際に決まる、キーワード含有量とページ順位のバランスを調整するための重みである。ユーザにより選択されたページのスニペット内の単語集合は X 、その要素が x である。 f は X のTFiDF値の合計である。 f_m は f の値が最大となる時の r_t 件目の値である。

Web操作パターン抽出の具体的なアルゴリズムは以下である。

- (1) ユーザが、Webで「観光オブジェクトA」をクエリとしたWeb検索を行う。
- (2) ユーザが、検索結果から特定のページを選択する。
- (3) ユーザが、「観光オブジェクトA」のスクラップブックにこのページを保存する。
- (4) システムは、ユーザの選択したページのスニペット内のTF-IDF値を算出し、順位の重み α を決定する。

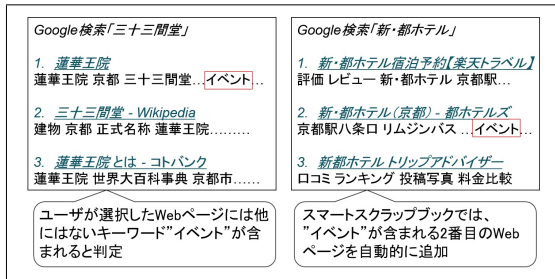


図 2 三十三間堂についての Web 検索行動

(5) 次にユーザが、「観光オブジェクト B」のスクラップブックを作成する。

(6) システムは、ユーザの操作パターンを元に検索結果スニペットに対し、式 1 によりスコアを算出し、最も高くなったページを抽出する。

(7) システムが、6 で検出されたページを自動的にスクラップブックに保存する。

図 2 は、Google で三十三間堂を検索した例である。選択されたページにはそれ以外のページにはない「イベント」というキーワードが含まれているため、これをキーワードとし抽出する。新・都ホテルにこの Web 検索行動を適用する時には、新・都ホテルによる検索で抽出した「イベント」を含むページをスクラップブックに自動的に追加する。

4. 地図操作パターン抽出

本節では、ユーザがあるオブジェクトに関して手動で操作を行い地図をスクラップブックに追加した場合にそれを他のオブジェクトに適応させる手法を説明する。ユーザが観光オブジェクト A について Google 地図検索 (図 3 の①) を行った後の動作に着目する。地図操作からのユーザ意図抽出に関して、平元ら [4] が絞込検索、比較検索、経路検索といった意図と地図操作パターンの関係について研究している。本手法でもそれらの地図操作パターンに基づき、ユーザの意図した地名を抽出する。絞込検索の操作パターンを例に説明する。中心点以外の別の地点 X に対してセンタリング (図 3 の②) およびズームインの操作 (図 3 の③) を行った後、パンやセンタリングで位置の調整を行い、観光オブジェクト A と地点 X が含まれる (図 3 の④) 画像を保存した。地点 X に対してセンタリングやズームインなどの操作を行ったことでそこにユーザの興味があると推定でき、さらに保存した領域内に存在するため、地点 X をユーザの興味を表すオブジェクトとして抽出する。他の観光オブジェクト B にこの操作行動を適用する際、観光オブジェクト B と地点 X が含まれる地図をスクラップブックに自動的に追加する。

地図操作パターン抽出の具体的なアルゴリズムは以下である。

(1) ユーザが、「観光オブジェクト A」をクエリとした地図検索を行う。

(2) ユーザが、地点 X をセンタリングする。

(3) ユーザが、地点 X に対してズームインをする。

(4) ユーザが、「観光オブジェクト A」と地点 X が地図上に含まれるようにパンやセンタリング、ズームアウトで調整する。

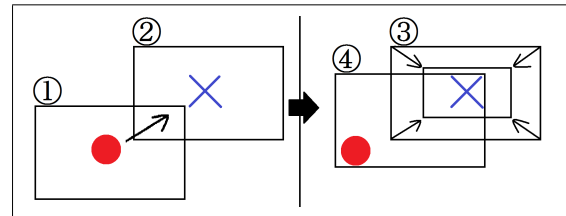


図 3 地図操作の模式図

(5) ユーザが、「観光オブジェクト A」のスクラップブックに地図を保存する。

(6) システムは、2, 3 の操作パターンから地点 X を抽出し、5 の時点の地図に地点 X が含まれるため、オブジェクト A に関する地点として X を抽出する

(7) ユーザが、「観光オブジェクト B」のスクラップブックを作成する。

(8) システムは、オブジェクト A におけるユーザの操作パターンを元に「観光オブジェクト B」と地点 X が含まれる地図を自動的に追加する

5. 実験と考察

5.1 Web 操作パターン抽出の実験

Web 操作パターン抽出に関して考察するために以下の実験を行った。実行環境は Google 検索エンジンである。データセットとして入力オブジェクト 10 件、適用オブジェクト 10 件とそれぞれのオブジェクトごとの順位とスニペットの情報、また入力オブジェクトごとに形態素解析によって得られた単語集合を用意した。それぞれのオブジェクトの種類の内訳は「宿泊」「社寺」「歩く」「グルメ」「土産」「見る」「遊ぶ」「カフェ」「甘味」の 9 つとなっている。異種のオブジェクトへの適用についても考察するため、多様なオブジェクトを用いた。あらかじめ、入力オブジェクトごとに適当なページと、それに対しての適用オブジェクトの出力ページに関しての正解 3 つを決定した。入力オブジェクト 1 件に対して適用オブジェクト 10 件の Score を求め、計 100 組み合わせの結果に対して、平均逆順位 MRR による評価を行う。本実験では 0.45 を基準に、それ以上となったものを正解とした。なお、0.45 は Score 降順に対して平均 2.2 位の順位の時の値である。その正解とシステムの結果の差異について考察を行う。

5.2 Web 操作パターン抽出の考察

実験結果は表 1 のようになった。100 組み合わせ中 70 組み合わせが基準値を上回る結果となった。そのため、内容と順位により保存するページを特定する手法は有用であると考えられる。基準値を下回る組み合わせに関しては、同名のオブジェクトが複数存在する場合である。本実験では「ル・タン」や「京都タワー」という名前を含んだ観光オブジェクトが他にも存在したために無関係のページスコアが高くなった。これは、Geo/Geo 曖昧性に関する問題であり、この曖昧性を解消することで、精度が上がると考えられる。

5.3 地図操作パターン抽出の実験

地図操作パターンによる地図保存に関して考察するために以

表 1 Web 操作パターン抽出の実験結果

	京都ホテルオークラ	三十三間堂	平安神宮	竹林の道	ル・タン	おたべ本館	京都タワー	東映太秦映画村	フランソワ喫茶室	加茂みたらし茶屋
新・都ホテル	0.57	0.55	0.47	0.28	0.61	0.57	0.23	0.57	0.48	0.55
金閣寺	0.49	0.61	0.46	0.48	0.53	0.53	0.23	0.58	0.40	0.57
清水寺	0.78	0.54	0.55	0.58	0.50	0.44	0.27	0.32	0.24	0.21
哲学の道	0.53	0.58	0.17	0.54	0.56	0.25	0.56	0.55	0.56	0.28
嵐山よしむら	0.57	0.22	0.42	0.53	0.32	0.58	0.47	0.61	0.56	0.53
祇園辻利	0.57	0.49	0.57	0.48	0.44	0.58	0.54	0.48	0.47	0.40
京都市美術館	0.58	0.61	0.48	0.21	0.56	0.61	0.57	0.44	0.56	0.28
京都国際マンガミュージアム	0.57	0.47	0.48	0.54	0.34	0.58	0.44	0.61	0.57	0.61
喫茶ソワレ	0.57	0.46	0.54	0.49	0.25	0.48	0.44	0.36	0.36	0.53
祇園きなな	0.61	0.48	0.36	0.54	0.56	0.53	0.46	0.33	0.32	0.43



図 4 三十三間堂と京都駅

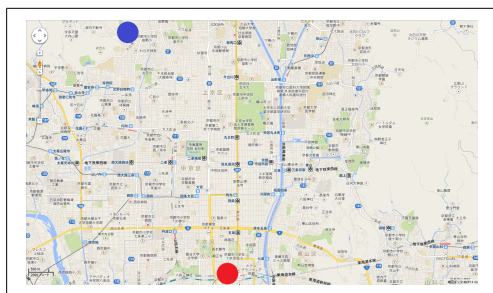


図 5 金閣寺と京都駅



図 6 三十三間堂と東寺宝物館

そのため、京都駅を抽象化し、観光オブジェクトとその最寄りの駅の地図というように、より適切な画像を保存できることが望ましい。また、異なる駅でそれぞれ地図が保存されている場合は、2つの駅間のアクセス方法を付与するなどのメディアを横断したシステムが必要である。次に、図6では、元々の検索オブジェクトである「新・都ホテル」(図6中の緑丸)が含まれ、「新・都ホテル」のスクラップブックと地図の内容が大きく重複している場合も考えられる。地図が重複している場合は枠で囲むなどしてスクラップブック上で表示する工夫が必要である。また、詳細検索以外の比較検索や経路検索のような他のパターンについてなどがあげられる。

6. まとめと今後の課題

手軽に観光情報を収集できるスマートスクラップブックシステムを提案した。他のオブジェクトに適用する定型的な行動の抽出方法として Web 操作パターン抽出と地図操作パターン抽出を開発した。Web 操作パターン抽出はユーザがあるオブジェクトに関して Web 検索を行い Web ページを保存した時のパターンからキーワード含有量とページの順位というデータを抽出し、次の観光オブジェクトに利用するものである。地図操作パターン抽出はユーザがあるオブジェクトに関して地図検索およびセンタリングやズームインなどの操作をし保存した時に、次の観光オブジェクトに利用するというものである。実験により、Web 操作パターン抽出としては、100 組み合わせの操作の適用において 70 組み合わせが基準値を上回る結果となり、有用性を確認した。地図操作パターン抽出の実験により、地点からの距離が近い観光オブジェクトへの適用を確認した。

Web 操作パターン抽出では、複数のキーワードを入力した場合や 2~3 個のページを見てから保存した場合について、2 回目以降の手動保存スクラップブックがある場合の抽出について、システムによって保存されてるデータの利用について、自動で

下の実験を行った。想定するユーザの操作として Google マップを用いてクエリ「新・都ホテル」による検索を行い、京都駅を地点とした場合と東寺宝物館を地点とした場合のそれぞれに対して以下を行う。地点に対してセンタリングおよびズームインの操作を行った後、パンやセンタリングで位置の調整を行い、新・都ホテルとそれぞれの地点が含まれる画像を保存した。その操作パターンを、京都駅を地点とした時の三十三間堂と金閣寺、東寺宝物館を地点とした時の三十三間堂の3つのオブジェクトに対して適応する。なお、新・都ホテルにとって三十三間堂は地理的に近い観光オブジェクト、金閣寺は地理的に遠いオブジェクトである。

それぞれの結果について述べる。図4, 5, 6はそれぞれ、京都駅を地点とした時の三十三間堂、京都駅を地点とした時の金閣寺、東寺宝物館を地点とした時の三十三間堂への適応結果である。

5.4 地図操作パターン抽出の考察

これらについて考察する。観光オブジェクトごとに地点との位置関係の地図を保存することができた。しかし、金閣寺と京都駅のように地理的に遠い場合も考えられる。この時、ユーザは手動で保存するならば金閣寺の最寄り駅を使うと考えられる。

保存されたデータをユーザが修正した際のレスポンスについてなどを今後の課題とする。また、観光に関連する Web ページのみを利用することで、Web 操作パターン抽出の精度向上が可能であると考えられる。地図操作パターン抽出では、オブジェクト間の距離が大きく離れている際の地点の抽象化について、詳細検索以外の比較検索や経路検索のような他のパターンについてなどを今後の課題とする。更に、Web と地図などのメディアを横断した操作におけるパターンを見つけ、スクラップブックの情報の追加や変更などができるシステムを構築する必要がある。

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。

文 献

- [1] 加藤誠, 大島裕明, 小山聡, 田中克己: 関係の類似性に基づく Web からのオブジェクト名検索, 情報処理学会論文誌データベース (TOD), Vol.2, No.2, pp.110-125, 2009.
- [2] 森悠太, 湯本高行, 角谷和俊: 集団の情報探索のための検索行動の抽出・適用方式, 電子情報通信学会 第 19 回データ工学ワークショップ (DEWS2008) 論文集, B7-3, 2008.
- [3] 中島伸介, 田中克之: 相対的マッピング処理に基づく相対的情報検索手法, 情報処理学会論文誌データベース (TOD), Vol.45, No.SIG04(TOD21), pp.63-75, 2004.
- [4] 平元綾子, 角谷和俊: オンライン地図におけるユーザ操作を用いた Web 検索方式, 電子情報通信学会論文誌. D, 情報・システム J90-D(2), pp.257-268, 2007.
- [5] 石野亜耶, 藤井一輝, 藤原泰士, 前田剛, 難波英嗣, 竹澤寿幸: 旅行プログエントリと質問応答コンテンツを利用した旅行ガイドブックの情報拡張, WebDB Forum 2012, 2012.
- [6] 倉島健, 岩田具治, 星出高秀, 高屋典子, 藤村 考: 行動範囲と興味の同時推定モデルによる地域情報推薦, 情報処理学会論文誌データベース Vol.6, No.2, pp.30-41, 2013.
- [7] 倉田陽平: 対話型観光プランニングシステムに向けて, 地理情報システム学会講演論文集 18, pp.309-312, 2009.
- [8] 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, 電子情報通信学会技術研究報告 信学技報 Vol.112, No.367, pp.13-18, 2012.