

個人写真の一般物体認識結果に基づくインタラクティブな観光地情報推薦

北村理紗¹⁾ 伊藤貴之¹⁾(正会員)

1) お茶の水女子大学大学院人間文化創成科学研究科

Interactive Tourist Spot Recommendation Applying Generic Object Recognition to Personal Photos

Risa Kitamura¹⁾ Takayuki Itoh¹⁾

1) Graduate School of Humanities and Sciences, Ochanomizu University

{risak, itot}@itolab.is.ocha.ac.jp

概要

デジタルカメラやスマートフォンなどの普及に伴い、誰もがいつでもどこでも手軽に写真を撮影することが可能となった。このような写真をライフログの一種として分析することで、個人の嗜好を推測できるのではないかと考えた。この考えに基づいて我々は、ユーザ自身によって過去に撮影した旅行写真からユーザの旅行の嗜好を推測し、その嗜好に合った観光地を推薦することを目的とした観光地推薦手法を研究している。本手法では、過去に撮影された旅行写真に一般物体認識を適用することで、よく撮影する被写体のキーワードを集め、そのキーワードを利用して特定の地域に関する観光地情報を抽出する。さらに、本手法では物体認識結果より得られたキーワード間の共起関係をグラフ構造化してユーザに提示する。このユーザインタフェース上での写真群の取捨選択により、ユーザは的確に旅行目的を絞り込みながら観光地情報を検索できる。本論文では提案手法の処理手順と実行例・評価結果を示し、本手法の有効性を議論する。

Abstract

Thanks to the recent spread of smartphones, tablets and digital cameras, people can take photos easily anytime, and anywhere. We can estimate patterns of actions and movements of people by analyzing their photos treating them as a life log. Based on this discussion, we are developing a technique to recommend tourist spots based on the estimation of users' preferences for traveling plans from their past personal travel photos by using generic object recognition. This technique applies the generic object recognition to a set of travel photos to extract preferred keywords of users' travel, and retrieves information using the keywords as query words. Also, the technique provides a visual user interface for structuring a large number of travel photos based on subject information and displaying a list of photos for each subject according to the purpose of travel. When a user selects photos of particular interest on the user interface, this user interface gathers keywords corresponding to the selected photos, extracts tourist spot information related to these keywords and recommends it to the user. This paper introduces the processing flow of the presented technique, and then a case study and evaluation with real travel photos.

1 はじめに

デジタルカメラやスマートフォン、タブレット端末などの普及によって、誰もがいつでもどこでも手軽に写真を撮影できる環境になった。現代社会ではライフログの一種ともいえる個人写真を分析することで、写真撮影者自身の嗜好を分析することが可能となる。一方で、個人旅行の多様化にともなって、観光地情報の推薦技術の研究が活発化している。このような背景から我々は、過去に撮影した写真を参照した旅行情報推薦について検討を進めてきた。なお本論文では広域な地名（例えば北海道、東京、京都など）を「地域」と称し、地域内での狭域の建築物・イベント開催地・写真名所などを「観光地」と称する。

写真を参照した旅行情報推薦の一手法として我々は、特定の個人が過去に撮影した旅行写真から被写体のキーワードを抽出し、それらを利用して特定の地域において写真撮影者が興味を持ちそうな観光地情報を検索・推薦する手法を開発している。本手法ではまず旅行写真群に一般物体認識を適用することで、旅行写真群の被写体を表すキーワード群を抽出する。続いて、グラフ可視化手法 Koala [1] に実装されたグラフ構築およびクラスタリング手法を適用して、旅行写真およびキーワードを構造化する。この旅行写真をウェブブラウザ上で木構造風に表示することで、過去の代表的な旅行写真が分類された形でユーザに提示される。ユーザが関心のある写真を選択すると、本手法は選択された写真に対応するキーワードを集め、そのキーワード群に関連する観光地情報を検索してウェブブラウザ上で提示する。このようなユーザインタフェースによりユーザは、過去の旅行履歴を振り返りながら次の旅行で訪問する観光地の情報を閲覧できる。

本研究のターゲットユースケースの例として、過去に旅行などで写真をたくさん撮影しているユーザが、特定の都市や地域を訪れることになったときに、その場所に関する興味を持ちそうな情報を簡単に収集したいという用途を想定している。2章にて後述するように、写真を利用した旅行推薦手法は既に数多く研究されている [2, 3, 4, 5, 6, 7]。しかし、写真共有サイトに投稿された写真のジオタグ情報を使用したものが多く、位置情報が付与されていない写真は使用できないという問題があ

る。本手法では、個人が所有する旅行写真に一般物体認識を適用することで、ユーザがよく撮影する風景や物体のキーワードを集め、そのキーワードを検索ワードとして、特定の地域に関する観光地の情報を抽出する。

ところで、個人旅行者に推薦すべき情報は以下の3種類に分類可能である。

観光地情報 A: 個人旅行者が過去に経験していないイベントや観光名所に関する情報。

観光地情報 B: 個人旅行者が過去に旅行以外の機会に経験しているが、旅行中には経験していない情報。野球観戦が好きで日頃よく行くが、過去の旅行先で野球を観戦したことはない、というような事象が該当する。過去の旅行以外のイベントで撮影した個人写真からの情報検索が可能である。

観光地情報 C: 個人旅行者が過去に旅行を通して経験したイベントや風景のうち、次の旅行で再度楽しみたい事象。以前の旅行でも海に行ったが次の旅行先でも海を見たい、というような事象が該当する。過去の旅行で撮影した個人写真からの情報検索が可能である。

現時点で我々は、上記の「観光地情報 C」を対象として提案手法を開発している。一方で、提案手法の入力はユーザ自身が撮影した写真群であり、技術的には旅行写真に限定する必要はない。つまり旅行以外のイベントを含めた全ての個人写真を入力情報とすることが技術的には可能であり、これによって観光地情報 B, C ともに検索可能となる。しかし旅行以外のイベントを含む全ての個人写真を対象としてキーワードを抽出することで、旅行に全く関係ないキーワードが多数抽出される可能性がある。そこで現時点で我々は観光地情報 C のみを対象とする。

本論文では我々が開発した旅行情報推薦手法を提案する。2章では関連研究を紹介し、3章で提案手法の処理手順を示す。4,5章で提案手法の実行例と評価を示し、6章で今後の課題を論じる。

2 関連研究

2.1 インタラクティブな旅行情報推薦システムの研究

倉島ら [2] は、Flickr に投稿された写真のジオタグ情報を人々の旅行履歴として利用した旅行ルート推薦手法を提案している。この手法では、ユーザの現在地から行

きやすい場所とユーザの興味に合致した場所に移動しやすいと仮定し、行動モデルを生成している。各ユーザのジオタグ付き写真集合は、時間情報でソートすると個人の旅行履歴とみなすことができると考え、ジオタグ情報を利用してユーザの行動モデルを生成している。Chengら [3] は、倉島らと同様に Flickr に投稿されたジオタグ付きの写真を利用してユーザの人物属性を考慮したトラベルルート推薦手法を提案している。この手法ではユーザの属性を性別、年齢、人種の顔属性に重点を置き、その属性ごとに行動モデルを生成している。

Gaoら [4] が提案する W2Go では、ユーザが関心のある旅行先の特定の都市や地域を入力すると、Flickr の写真に付与されているタグとジオタグ情報、Yahoo Travel Guide を利用してユーザの興味を考慮したランドマークをランク付けして提示する。Caoら [5] が提案するシステムでは、ユーザが所望のキーワードを入力すると、撮影場所をもとに写真をクラスタリングし、各クラスタにおける代表写真と、ユーザの関心のある写真もしくは場所を表すキーワードを照合して提示する。また、Luら [6, 7] は、ユーザの旅行先、旅行日程、訪問時間、好きな旅行スタイルを考慮したルートを推薦するインタラクティブな旅行推薦システム Photo2Trip を提案している。この手法では、Panoramio から収集したジオタグ付き写真から人気のあるランドマークを抽出し、旅行ブログや旅行サイトを利用して、ユーザの好みに応じてカスタマイズされた旅行ルートプランを推薦している。

これらの手法はいずれも写真のジオタグ情報を前提とした推薦手法であり、位置情報が付加された写真を利用することが前提となっている。そのため、位置情報が付加されていない写真を利用できないという問題がある。そこで我々は、ジオタグ情報を利用せず、旅行写真における被写体情報からユーザの旅行の嗜好を推測し、その好みにあった観光地を推薦する手法を開発している。

また、これらの手法が搭載するユーザインタフェースでは、推薦に反映させるユーザの入力はテキスト情報のみであり、写真群を入力していない。そこで本研究では、写真群とキーワード群の両方を入力とするユーザインタフェースを提案する。

2.2 大量画像ブラウザ

近年の急速なデジタル端末の普及に伴い、大量の写真を整理したり、ユーザが効率よく閲覧できる技術はユー

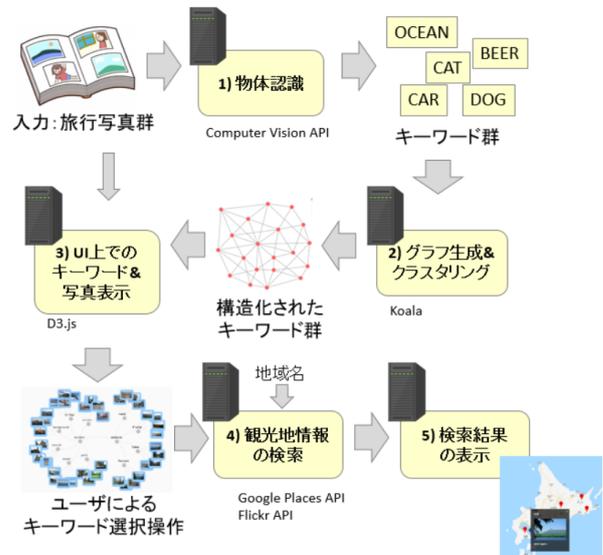


図1 提案手法の処理手順

ザにとって非常に有用であり、数多くの研究がなされている。大量画像ブラウザは「構造型」[8, 9]と「非構造型」[10, 11]に大別できると考える。これらの手法の多くは、過去に撮影された大量の画像の全容を一目で把握し、ユーザが関心のある特定の写真をズームインするといった操作方法を得意とする。本手法では、ユーザの旅行目的に近い意味を有するキーワードや写真から、それに関連するより具体的なキーワードやそのキーワードが付与された写真を表示することで、ユーザの旅行目的を絞り込めるよう、木構造を導入したユーザインタフェースを提案する。本手法は構造型の画像ブラウザの一種であるが、従来手法 [8, 9] の多くは空間充填型の画像配置手法を採用しているのに対して、本手法は写真やキーワードを稜線で接続するノードリンク型の表現手法を採用している点異なる。

3 提案手法

提案手法の処理手順を図1に示す。本手法は写真所有者の過去の旅行に関係あるキーワードの中から検索ワードとなるキーワードを選択させて観光地情報を検索する仕組みとなっている。そこで本手法では、まず写真所有者が撮影した旅行写真群を入力し、各写真に一般物体認識を適用する(図1(1))ことでキーワード群を得る。この結果として得られるキーワード群は数百～数千におよぶ場合があることから、キーワード群をノードとする

グラフを生成し、これにクラスタリングを適用することで、キーワード群を構造化する(図 1(2))。続いて構造化されたキーワード群および写真群をウェブブラウザ上でユーザに提示し(図 1(3))、これから検索したい観光地情報に関係あるキーワードを選択させる。そして選択されたキーワードと地域名を検索ワードとして観光地情報を検索し(図 1(4))、検索結果をウェブブラウザ上にて地図とともに表示する(図 1(5))。なお本手法では、旅行の目的地となる地域は決まっているが、その都市や地域の中で具体的にどのような観光地を訪れるか決まっていないう状況想定して、その地域における観光地情報を推薦する。

3.1 キーワードの付与

前章で紹介した関連研究ではジオタグ情報が付与されている写真を使用することが前提となっているが、我々の日常生活において必ずしも位置情報が付与された写真を撮影するとは限らない。そこで本研究では、ユーザが所有する過去の写真を分類する手がかりとして、位置情報の代わりに一般物体認識結果を用いる。我々の実装では Computer Vision API^{*1} という Microsoft 社の画像認識 API を用いて一般物体認識の情報を取得する。Computer Vision API では各写真に対して一般物体認識結果を物体名のリストで返し、その各々について範囲 [0,1] の実数で確信度を返す。なお、一般物体認識サービスは Microsoft 社以外の各種企業からもリリースされている。4 章にて紹介する旅行写真群に対して数社の一般物体認識サービスを適用したところ、Computer Vision API が最も正解率が高かったことから、われわれの実装では Computer Vision API を採用した。

3.2 物体認識結果のグラフ構造化

続いて本手法では、キーワードをノードとしたグラフを構築し、さらにノード群に対してクラスタリングを適用することで、グラフを構造化する。以下、ノードの集合 (= キーワードの集合) をクラスタと称する。 i 番目の写真の j 番目のキーワードの確信度を c_{ij} 、キーワードの総数を m 、写真の総数を n とし、以下の処理によってグラフを生成しクラスタリングを適用する。

- j 番目のキーワードに対応するノードを n 次元ベク

トル (c_{1j}, \dots, c_{nj}) で表現する。

- 任意の 2 ノード間についてベクトルの内積を算出し、内積が閾値以上であれば 2 ノードをエッジで接続する。この処理は同一写真での共起度の高いキーワードをエッジで接続することに相当する。
- 同一キーワードと共起する傾向にある 2 つのキーワードをできるだけ同一クラスタに所属させる。この処理によって、上位概念となる同一のキーワードに共起する下位概念のキーワード群が同一クラスタに所属されるようなクラスタリング結果を得る。

我々の実装では、グラフ可視化手法 Koala[1] に適用されたグラフ構築およびクラスタリング処理を上記の処理に適用している。Koala に搭載されたクラスタリング手法には「多くの共通ノードとエッジで連結されたノード群を同一クラスタに所属させる」という基準が採用されている。この基準は上記の「上位概念となるキーワードに共起する下位概念のキーワード群が同一クラスタに所属される」という結果を得るのに有利である。そこで本手法ではグラフの構築およびクラスタリングに Koala を採用している。

3.3 旅行目的を選択するためのユーザインタフェース

過去に撮影した大量の写真の全貌を必ずしもユーザ自身が詳細に記憶しているとは限らないし、またその写真から抽出されるキーワード群をユーザ自身が詳細に連想できるとも限らない。そこで本手法では、旅行写真の被写体情報にもとづいて大量の旅行写真およびキーワードを構造化し、ユーザの旅行目的や嗜好を的確に思い返しながらかキーワードを選出するためのユーザインタフェースを搭載する。

現時点での Koala の実装はウェブブラウザ上での可視化結果の表示に対応していない。そこで本手法では、Koala により構造化されたグラフを JSON 形式で出力し、D3.js^{*2} の Force Layout というグラフ可視化機能を適用したユーザインタフェースを提供する。D3.js はグラフ可視化機能を搭載した JavaScript 言語のライブラリとしては我々が知る限り最も普及しているものである。我々は D3.js のグラフ可視化機能の一つである Force Layout の他に cola.js^{*3} のグラフ可視化機能も試

^{*1} <https://azure.microsoft.com/ja-jp/services/cognitive-services/computer-vision/>

^{*2} <https://d3js.org/>

^{*3} <https://ialab.it.monash.edu/webcola/>

みたが、ノード配置の妥当性、対話操作の滑らかさなどの点で D3.js のほうが本研究の目的に合致していると主観的に判断し、D3.js の Force Layout を採用した。

D3.js を用いて構築するユーザインタフェースでは、Koala で得られた 1 つのクラスタを 1 つのノードとし、以下のようにルートノードおよびハブノードを定義する。

- エッジ数の最も多いクラスタを「ルートノード」とする。
- 閾値以上のエッジ数でエッジ数の最も多いクラスタとエッジが張られているクラスタを「ハブノード」とする。

また各ノードを表示する際には、ノードに属するキーワードをあわせて表示する。さらに本手法では、キーワードが付与された写真を複数枚選び、各ノードからエッジで接続して表示する。本章ではノードからエッジで接続された写真を「写真ノード」と呼ぶ。ただし、ルートノードには写真ノードを接続しないものとする。この写真ノードを新たに追加することで、ユーザは被写体で分類された写真群を確認できる。

ここで本手法では、各ノードのキーワードが付与された写真のうち代表的なものを写真ノードとして自動選出する。代表的な写真の選出に際して我々の実装では、各ノードに関連付けられた i 番目のキーワードに対するスコア $score_i$ を以下の式

$$score_i = conf_i / appear_i \quad (1)$$

により算出し、その合計値 $\sum score_i$ の高い順にあらかじめ設定された枚数の写真を選出する。なお、 i 番目のキーワードに対して確信度 $conf_i$ と出現回数 $appear_i$ が与えられているとする。

ユーザに最初に提示する初期画面では、ルートノードとハブノード、およびハブノードに対応するキーワードが付与された写真ノードを表示し、それ以外のノードは表示しない。そして、ルートノードとハブノード以外のノードは、初期画面で提示したノードと写真ノードからユーザが関心のあるノードをクリック操作で選択した場合に表示する。以上の表示方法を図 2 に示す。

エッジ数が多いノードであればあるほど、そのノードに対応するキーワードは多様なキーワードと共起して写

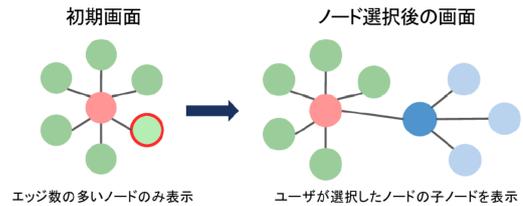


図 2 ノードの表示

真に付与される確率の高いキーワードである。このようなキーワードは、極めて一般性の高い、上位概念に属するキーワードであると考えられる。そこで、クラスタのエッジ数で表示するノードを制限し、木構造のように表示することで、ユーザの旅行目的を絞り込むのに適したユーザインタフェースを提供する。これにより、まず旅行目的に近い意味を有する上位概念をノードをユーザに選択してもらい、それに接続されているノードを同時に表示する。ここで表示されたノードは、より具体的な、下位概念と考えられるようなキーワードに対応するノードである。このような操作により、ユーザの旅行目的に沿った形でキーワードを絞り込む。逆に、旅行目的に沿わないノードを対話操作によって観光地推薦の対象から除外することもできる。

3.4 観光地の検索

続いて本手法では、前節で示したユーザインタフェース上での対話操作により絞り込んだキーワードを、あらかじめ指定された地域名とともに検索ワードとして、観光地情報を検索する。我々の実装では、観光地情報の抽出に以下の 2 種類の API のいずれかを切り替えて採用している。

Places API (Google Maps Platform):^{*4} Google が提供するサービスで、地域名に関係ある施設やサービスなどの情報を検索するのに用いる。なお Places API の中でも我々の実装では、検索対象とキーワードとの照合が可能で、また完全な住所が与えられてなくても漠然とした地域名からの検索が可能な Text Search requests を使用する。

Flickr API:^{*5} Flickr が提供するサービスで、Flickr に投稿された写真に付与されたメタデータと検索ワードの照合により該当写真群を検索するのに用いる。我々の実

^{*4} <https://cloud.google.com/maps-platform/places/>

^{*5} <https://www.flickr.com/services/api/>

装では flickr.photos.search メソッドを用いて、以下の手順で観光地情報を抽出する。

1. 特定の地域名から、キーワードに関連する写真を取得する。
2. 取得した写真の撮影場所（緯度・経度）周辺で撮影された写真の枚数を調べる。
3. 以下のいずれかの基準を満たす場所を観光地とみなす。
 - キーワードとの関連を示す値が一定以上とされている。
 - 一定枚数以上の写真が撮影されている。
4. 観光地とみなした場所に関する検索結果を返す。

本手法ではユーザインタフェース上で関心のあるノードもしくは写真ノードを選択させ、選択されたノードに含まれるキーワード（もしくは選択された写真に付与されたキーワード）を観光地検索に使用するキーワードとする。

さらに我々の実装では、Maps JavaScript API *6 を利用して、観光地情報を Google マップ上に表示する。

4 実行例

本章では、あるユーザの国内・国外の旅行写真合計 2,581 枚に対して一般物体認識を適用し、抽出されたキーワードから構築されたグラフを画面配置し、その結果をもとに Force Layout でグラフを描画した例を紹介する。物体認識により得られたキーワードの総数は延べ約 145,000 個である。

キーワード群を Koala でグラフ構造化し、ハブノードにあたるキーワードを抽出した例を図 3 に示す。また、特定の旅行や観光のタイプに関連するキーワードを図 4 に示す*7。図 3 から、エッジ数の多いハブノードには outdoor や sky などの一般性の高い単語が多いことがわかる。図 3 の (d) は最もエッジ数の多いクラスターでルートノードに相当する。また図 4 では、赤い枠線で囲まれた範囲内に、特定の旅行のカテゴリに関するキーワード

が配置されていることがわかる。例えば、図 4 の (a) は山に関連する単語群、(b) は動物園や水族館に関連する単語群、(d) は海や川に関連する単語群である。

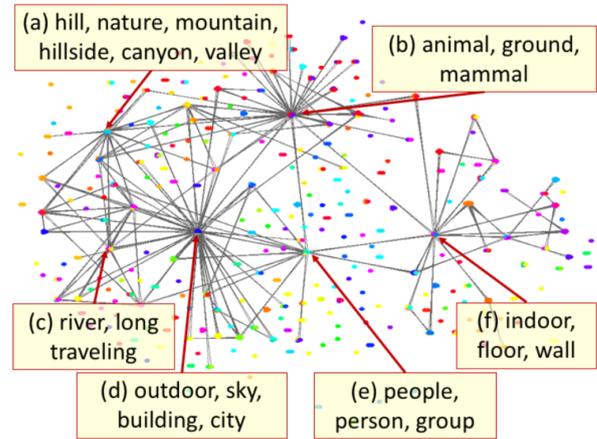


図 3 ハブノードに対応するキーワード群

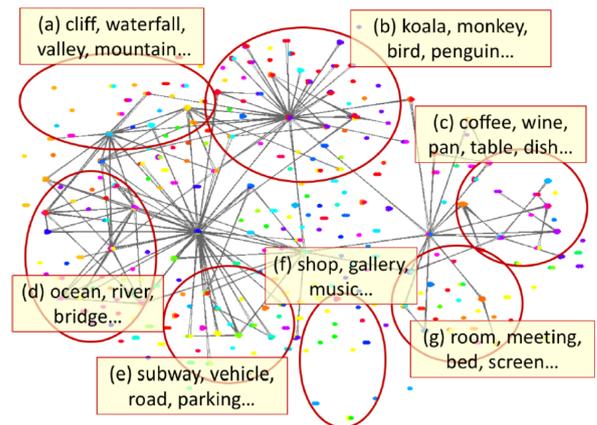


図 4 旅行カテゴリを表すキーワード群

続いて、Koala で構築されたグラフ構造を JSON 形式で出力し、D3.js の Force Layout で表示した初期状態を図 5 に示す。この例では、最もエッジ数の多いルートノードを中央に配置する同時に、ルートノードからエッジが張られているハブノードに対して、各ハブノードのキーワードが付与され写真を 4 枚選んでノードとして配置している。ノードの横に表示しているキーワードは、各ハブノードを代表すると考えられるキーワードを手動で選んでおり、ノードをクリックするとクリックしたノードに属する全てのキーワードが表示される。なお、この例ではキーワードの登場回数が 5 個以下のキーワードを除外している。例として、図 5 の初期状態が

*6 <https://developers.google.com/maps/documentation/javascript/tutorial>

*7 図 3,4 に示す可視化結果は、実行結果の検証のために用いるものであり、ユーザに直接提示される可視化結果ではない点に注意されたい。

ら nature が属するハブノードを選択し、その子ノードを表示した例を図 6 に示す。nature の子ノードとして、nature の下位概念のキーワードと考えられる forest や lake などのキーワードが表示されていることがわかる。このように旅行目的に近い上位概念と考えられるノードを選択してもらい、その下位概念と考えられるノードを表示することで、ユーザの旅行目的を絞り込めるようなインタラクションを実現する。

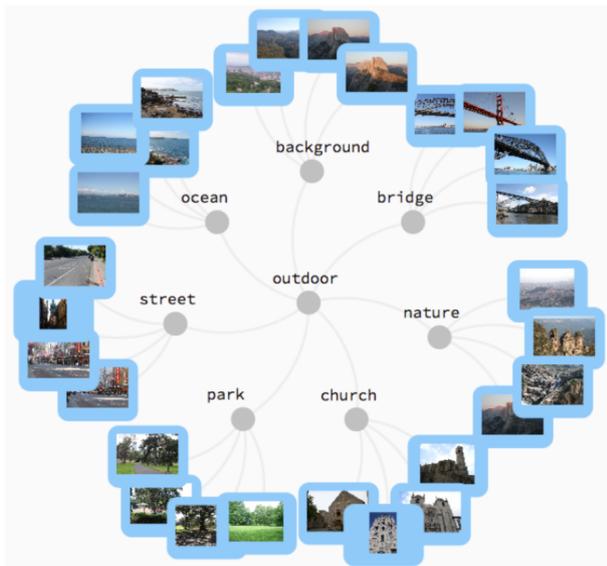


図 5 グラフ表示の初期状態の例

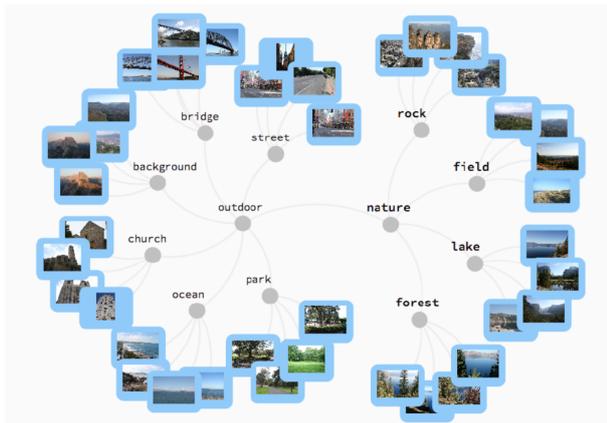


図 6 キーワードノードをクリックした例

このグラフ構造を操作するウェブブラウザ上のユーザインタフェースのスナップショットを図 7 に示す。写真ノードをマウスオーバーすると、写真ノードに関連するキーワード群が描画領域の左上端に表示される。また

「SHOW CHILD NODE」ボタンを押すと図 6 のように子ノードを表示した状態となる。「SEARCH NODE BY KEYWORDS」ボタンを押すとキーワードを入力できる。「SELECT NODE」ボタンを押すと写真を選択できる。これらの操作によってキーワードまたは写真を選択した上で「SEARCH SPOTS」ボタンを押すと、検索ワードが API に入力される。我々の環境および本章で用いるデータにおいて、全ての操作に対して 1 秒以内になめらかに画面が切り替わる。

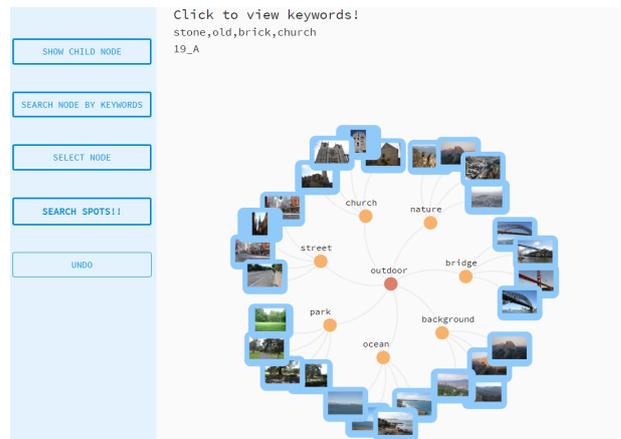


図 7 ウェブブラウザで使用するユーザインタフェース

観光地推薦のための API から得られた観光地推薦結果の表示例を図 8 に示す。この表示では、Google マップ上に推薦スポットをマーカーで表示し、マーカーをクリックすると情報ウィンドウを表示している。この例では旅行先を「北海道」とし、図 5 の初期画面から nature が属するハブノードの子ノードに属するキーワードの lake を選択した場合の観光地推薦結果である。

5 評価

前章で紹介した国内・国外の旅行写真を用いて評価実験を実施した。具体的には、写真所有者以外の被験者 10 人による評価と、写真所有者による評価を実施した。本章では評価実験の内容と結果について紹介する。

5.1 被験者による評価

旅行で訪れてみたい特定の都市や地域における興味を持ちそうな情報を簡単に収集し、具体的な訪問地（観光地）を決めることを想定した評価実験を実施した。前章で紹介した旅行写真を用いて、写真所有者以外の 20 代女性被験者 10 人を評価実験に参加させた。客観性の高

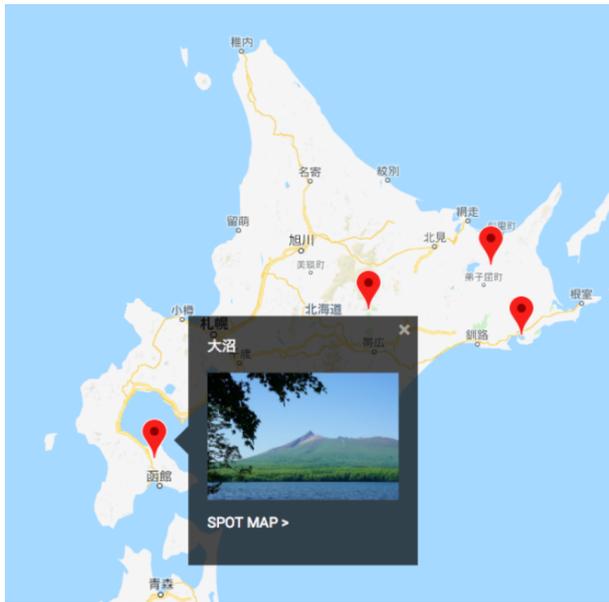


図 8 観光地推薦結果の例

い実験結果を求めるといふ観点から、写真所有者を被験者に含めずに実験を実施した。

まず代表写真の選出方法について評価した。具体的には、Computer Vision API を用いた選出されたキーワードを単位として、式 (1) を用いて写真を選出することの妥当性を評価した。本評価実験を実施するにあたり、以下の 2 種類の写真選出方法を用意した。

写真選出 A: 式 1 で算出したスコアに基づく選出

写真選出 B: ランダムな選出

2 種類の写真選出結果を表示したユーザインタフェースを操作させ、写真周辺のキーワードに対して適切に選択されていると思われる写真の枚数を回答させた。表示された写真の総数はいずれも 122 枚であった。表 1 に 10 人の回答の平均値を示す。この結果より、キーワードの出現頻度と確信度による代表写真選出が効果的であることが伺える。

表 1 写真選出結果の評価

	写真選出 A	写真選出 B
写真枚数の平均値	88.0	68.7

続いて情報推薦結果について評価した。具体的には、キーワードと地域名から観光地情報を検索することの意義、情報検索手段として採用した API による検索結果の

満足度の違い、複数の API を切り替えて利用することの意義などを検討するための評価実験を実施した。本実験では地域を「沖縄」と「北海道」とし、被験者にキーワードノードと写真ノードを 1 つずつ選択してもらい、その各々に対して提示された情報推薦結果を 5 段階評価してもらった。情報検索手段として以下の 3 種類を試みた。

検索手段 1: Places API による推薦 (ユーザレビューの高い順)

検索手段 2: Flickr API による推薦 (キーワードとの関連度の高い順)

検索手段 3: Flickr API による推薦 (撮影された写真の枚数が多い順)

表 2 に 5 段階評価の平均値を示す。この結果から、第三者による過去の旅行写真を Flickr API で取得する方法のほうが本手法との親和性が高いことが伺える。また、本手法において必ずしもユーザレビューにもとづく情報推薦が望ましいとは限らないことが示唆される。被験者のコメントにも「キーワードに sunset を選択して夕焼けを見るスポットを検索したつもりが、「サンセット」を施設名に含むユーザレビューの高いホテルが表示されていた」といった事例が紹介されており、ユーザレビューにもとづく観光地情報提示の難しさがみられる。また検索手段 2 と 3 の評価は甲乙つけがたい状況にあり、複数の検索手段をサポートすることの有効さを示唆させる結果となった。しかし残念ながら被験者のコメントからは、検索手段 2 と 3 の評価の違いを理由づけるコメントは見当たらなかった。

表 2 検索手段別の情報推薦結果の評価

検索手段	1	2	3
沖縄・キーワードノード	2.0	4.2	4.5
沖縄・写真ノード	2.8	3.9	3.7
北海道・キーワードノード	1.7	4.1	3.9
北海道・写真ノード	2.9	3.8	3.7

最後に被験者に「ユーザインタフェースの操作性」「情報推薦結果の表示手段」について 5 段階評価してもらった。表 3 に 5 段階評価の平均値を示す。この結果から、本手法が被験者から一定の評価を得られていることがわかる。被験者によるコメントも「写真の撮影日時と場所の表記機能、撮影された観光地の関連ウェブサイトの表

示機能が欲しい」「気に入った写真をマークするなどして、それに類似する写真を強調表示する機能があると嬉しい」といった追加機能に関するアイデアのみで、「わかりにくい」「使いにくい」「面倒くさい」「応答が遅い」などの不満は一切見られなかった。

表3 提案手法の5段階評価の回答者数

評価	1	2	3	4	5
ユーザインタフェースの操作性	0	0	0	9	1
情報推薦結果の表示手段	0	0	2	6	2

5.2 写真所有者による評価

前章で紹介した写真の所有者に、Koalaによるグラフ構造およびユーザインタフェース上で示した木構造を提示した。続いて、地域を「沖縄」と「北海道」に設定した際の観光地推薦結果を提示した。そして、キーワードの選出結果、およびキーワードのクラスタリングや木構造構築の結果、ユーザインタフェース上での表示結果、といった点について主観評価してもらった。さらに、観光地推薦結果の満足度についても主観評価してもらった。

写真所有者から総論として、Koalaによるクラスタリング結果も、ユーザインタフェースで表示した木構造も、総じて適切な構造であり、撮影写真の傾向に応じて旅行の目的をうまく分類できている、とのコメントを頂いた。

また、今回使用した旅行写真では、物体認識結果として得られるキーワードの粒度に偏りが見られるというコメントを頂いた。例えば本実験の実行結果では動物に関するキーワードが多かった。一方で、ユーザインタフェースで示した木構造では動物のキーワードが多く表示されているわけではなく、ユーザインタフェース上でのほうがバランスよい表示結果になっているというコメントを頂いた。これは、ユーザインタフェース上で表示するノードをクラスタのエッジ数で制限しているためであり、むしろこの制限によってキーワード表示のバランスがよくなったとも考えられる。

さらに写真所有者からは、ルートノードに対応するキーワード outdoor に関する観光地は検索しやすいが、キーワード indoor に関する情報が目につきにくい、という意見を頂いた。エッジ数が最も多いクラスタを単一のルートノードとしてユーザインタフェース上で表示しているという仕様が原因であると考えられる。今後の課題として、ルートノードを複数個提示するといった改善

を試みたい。

続いて写真所有者に、近いうちに「沖縄」と「北海道」に旅行に行く想定して、本手法で推薦された観光地情報がどのように満足できるものであるかについて尋ねた。その結果として、いくつかのキーワードについて以下のようなコメントを得られた。

beach, ocean: 「沖縄」「北海道」全般にわたって網羅的に海岸が紹介された。自分のドライブコースに合わせて海岸を選ぶことができる。写真を1クリックするだけでこれだけの情報が得られるのは簡単で有用である。

mountain: 「沖縄」「北海道」全般にわたって網羅的に山が紹介された。ただし山といっても非常に多彩な楽しみ方がある。山だけでなく他のキーワードとの組み合わせをユーザインタフェース上で選ぶことにより、直観的に観光地情報を得られる。

このようなコメントから、いくつかのキーワードについて、キーボードで単語を打って入力する一般的な検索操作と比較して、ユーザインタフェース上の手軽な操作によって満足できる検索結果を得られるとの評価をいただいた。一方で別のいくつかのキーワードについて以下のように問題点が指摘された。

plant: 植物に関する場所と工場が雑多に表示された。同じスペルの全く意味の異なる単語への対策が必要である。

nature: 単語の抽象度が高いせいか、検索結果が雑多すぎた。旅行の参考にならなかった。

church: もともと国内旅行ではあまりいい検索結果は期待できないと感じた。この単語を表示する必要性を感じなかった。

この結果から、物体認識結果として得られるキーワードには、観光地情報の検索に向いているキーワードとそうでないキーワードがあることが示唆される。そこで今後の課題として、ユーザインタフェースで優先的に表示するキーワードを選別する手法を検討したい。例えば、観光地情報検索において満足度の高い単語と低い単語を調査し、満足度の高い単語を優先的に提示するといった方法がある。

5.3 既存の旅行推薦サービスとの比較

本手法が既存の旅行推薦サービスと比較してどのような優位性を有するかについて我々自身で比較を行った。我々は既存の旅行情報サイトの例として TripAdvisor^{*8} および JTB^{*9} をしばらく閲覧した。そして、これらのウェブサイトの旅行推薦機能について本手法がどのような優位性を有するかについて議論した。その結果を以下にまとめる。

本手法では過去の旅行写真を表示することから、ユーザの嗜好に合った情報の検索が容易となる。このことから、旅行目的や訪問地を絞り込むのが容易であり、また地域名の選択・写真の選択という少ない操作で観光地を検索できる、といった点で有利である。さらに、過去の旅行写真にグラフ構造化を適用することで、写真所有者自身の過去の旅行の内容にもとづいて写真が分類表示される。このことから本手法は、観光地情報を検索するためのユーザインタフェースをパーソナライズされた形で提供できる手法であると言える。

それに対して、例えば TripAdvisor で観光地情報を検索する場合は、特定の都市や地域を入力したあとに非常に多くの旅行カテゴリが提示される。例えば「大阪」といった特定の地域名を入力すると、「ショッピングモール」や「レストラン予約」、「文化体験」など 48 の旅行カテゴリが提示される。あるいは、JTB のサイトで観光地や観光ツアーを検索する場合、「ハネムーン」「スポーツ観戦・参加ツアー」「クルーズ」などのカテゴリを最初を選ぶ必要がある。これらのカテゴリは一般的な旅行者を想定して汎用的に構造化されたものであり、パーソナライズされたものではない。この点で本手法のほうが個々の旅行者に合ったユーザインタフェースを提供可能であると考えられる。

6 まとめと今後の課題

本論文では旅行写真の被写体情報を利用した観光地推薦の一手法を提案した。本手法では、過去に撮影された旅行写真に一般物体認識を適用することで、よく撮影する被写体のキーワードを集め、そのキーワードを利用して特定の地域に関する観光地情報を抽出する。また本手法では、旅行目的に応じて被写体ごとに写真を一覧表示

し、過去の旅行を振り返りながら次の旅行先の訪問地を決定するユーザインタフェースを提供する。本論文では本手法の実行例と評価結果を示し、その有効性と課題について議論した。

今後の課題として、評価実験の結果を踏まえてユーザインタフェースを改良したい。また、キーワードによっては推薦される情報の満足度が低い傾向もみられるため、キーワードの重み付けを計算する手法を検討したい。

本手法で採用している Koala の実装に関する制約の解消も今後の課題としてあげられる。例えば現状の Koala の実装では、エッジ生成の条件となる閾値をユーザインタフェースで調節している。しかし、本手法が想定しているエンドユーザに閾値を設定させるのは実用的ではない。そこで、この閾値を適切に自動設定する手法が必要である。また Koala はもともと 10 次元程度の変数を有する入力データを想定した手法であり、本手法のキーワード選出結果から得られるような高次元ベクトルを扱うことを想定した実装になっていない。そこで次元削減手法の適用が望ましいと考えられる。

別の課題として、既存の旅行推薦手法および旅行推薦サービスとの比較実験があげられる。5.3 節にて我々自身による旅行推薦サービスとの比較について論じたが、これに関する被験者実験が今後の課題の一つである。また、旅行推薦に関する既存研究を我々の環境で実行して比較することも今後の課題の一つである。さらに、各被験者が各自で所有する写真を用いて評価実験を実施することにより、より望ましい結果が得られる可能性もある。これについても検討を進めたい。

現時点で我々は、入力する個人写真を旅行写真に限定することで、1 章で示した「観光地情報 C」のみを観光地情報推薦対象としている。しかし、旅行写真とそれ以外の写真の違いを明確に定義することは難しい。言い換えれば、ユーザが旅行写真と呼ぶにふさわしい写真だけを適切に選んで本手法に適用できるとは限らない。この定義に関するガイドラインを設けることが実用上の重要な課題になると考えられる。一方で、旅行写真に限定せずに全ての個人写真を対象とすれば、「観光地情報 C」だけでなく「観光地情報 B」も対象として研究を進めることも可能である。しかし 1 章でも論じたように、全ての個人写真を対象としてキーワードを抽出することで、旅行に全く関係ないキーワードが多数抽出される可能性が

^{*8} <https://www.tripadvisor.jp>

^{*9} <https://www.jtb.co.jp/>

ある。そこで全ての個人写真を対象として本手法を適用するとしたら、なおさらキーワードの重みづけが重要となる。さらに現時点の本手法では、積極的に写真を撮影できない訪問先（例えば商店、温泉、コンサート会場など）に関するキーワードを検索ワードにすることはできない。このような嗜好を写真以外の手段で抽出して検索ワードに追加することも本研究の重要な課題である。

参考文献

- [1] T. Itoh, K. Klein.: *Key-node-Separated Graph Clustering and Layout for Human Relationship Graph Visualization*, IEEE Computer Graphics and Applications, 35(6), 30-40, 2015.
- [2] T. Kurashima, T. Iwata, G. Irie and K. Fujimura.: *Travel route recommendation using geotags in photo sharing sites*, the 19th ACM international conference on information and knowledge management, 579-588, 2010.
- [3] A.-J. Cheng, Y.-Y. Chen, Y.-T. Huang, W. H. Hsu, H.-Y. M. Liao.: *Personalized travel recommendation by mining people attributes from community-contributed photos*, the 19th ACM international conference on Multimedia, 83-92, 2011.
- [4] Y. Gao, J. Tang, R. Hong, Q. Dai, T.-S. Chua, R. Jain.: *W2Go: a travel guidance system by automatic landmark ranking*, the 18th ACM international conference on Multimedia, 123-132, 2010.
- [5] L. Cao, J. Luo and A. Gallagher, X. Jin, J. Han, T. S. Huang.: *A worldwide tourism recommendation system based on geotagged web photos*, IEEE International Conference on Acoustics Speech and Signal Processing, 2274-2277, 2010.
- [6] H. Yin, X. Lu, C. Wang, N. Yu, L. Zhang.: *Photo2Trip: An Interactive Trip Planning System Based on Geo-Tagged Photos*, In Proceedings of the 18th ACM international conference on Multimedia, 1579-1582, 2010.
- [7] X. Lu, C. Wang, J. M. Yang, Y. Pang, L. Zhang.: *Photo2trip: Generating Travel Routes from Geo-tagged Photos for Trip Planning*, the 18th ACM international conference on Multimedia, 143-152, 2010.
- [8] B. B. Bederson: *PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps*, the 14th annual ACM symposium on User interface software and technology, 71-80, 2001.
- [9] A. Gomi, R. Miyazaki, T. Itoh, J. Li.: *CAT: A Hierarchical Image Browser Using a Rectangle Packing Technique*, the 12th International Conference on In-

formation Visualization, 82-87, 2008.

- [10] J. Yang, J. Fan, D. Hubball, Y. Gao, H. Luo, W. Ribarsky, M. Ward.: *Semantic Image Browser: Bridging Information Visualization with Automated Intelligent Image Analysis*, IEEE Visual Analytics in Science and Technology, 191-198, 2006.
- [11] Y. Zheng, A. Gomi, T. Itoh.: *ImageCube: A Browser for Image Collections Associated with Multi-Dimensional Datasets*, the 15th International Conference on Information Visualisation, 12-17, 2011.

北村理紗



2017年津田塾大学学芸学部情報科学科卒業。2019年お茶の水女子大学大学院人間文化創成科学研究科理学専攻博士前期課程修了。

伊藤貴之



1990年早稲田大学理工学部電子通信学科卒業。1992年早稲田大学大学院理工学研究科電気工学専攻修士課程修了。同年日本アイ・ビー・エム(株)入社。1997年博士(工学)。2000年米国カーネギーメロン大学客員研究員。2003年から2005年まで京都大学大学院情報学研究科COE研究員(客員助教授相当)兼任。2005年お茶の水女子大学理学部情報科学科助教授。2011年同大学教授。2011年から2017年まで同大学シミュレーション科学教育研究センター長。2019年から同大学文理融合AI・データサイエンスセンター長。ACM, IEEE Computer Society, 芸術科学会, 他会員。