

個人の旅行写真の一般物体認識に基づく観光地推薦

北村 理紗[†] 伊藤 貴之[†]

[†] お茶の水女子大学 人間文化創成科学研究科 〒112-0012 東京都文京区大塚 2-1-1

E-mail: †{risak, itot}@itolab.is.ocha.ac.jp

あらまし 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。今や写真が生活や行動履歴を記録するものとなりつつあり、ライフログの一種といっても過言ではない。そこで、このような写真を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。本報告では過去に撮影した旅行写真に着目して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とそのユーザインタフェースを提案する。我々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、物体認識結果より得られたキーワード間の共起関係からグラフを構築し、さらに写真ノードを追加したグラフを可視化してユーザに提示する。このユーザインタフェースにより、ユーザが観光地推薦のために参照してほしい写真を選択したり、参照しなくてよい写真を除外することで、旅行の目的を絞り込めるようにする。本報告では、実際の旅行写真を使用してグラフ可視化した例とユーザインタフェースを紹介する。

キーワード 観光地推薦, 一般物体認識, ユーザインタフェース

1. はじめに

スマートフォンやタブレット端末の普及によって、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。これに伴い、旅行や誕生日、記念日などの特別なイベントだけでなく、食事の写真や購入したものなど、日々の生活の一部を撮影する人も増えている。以前は写真や動画は特別なイベントを記録するものと捉えられることが多かったが、最近では生活や行動履歴の記録媒体にもなろうとしている。そこで、このようなライフログの一種ともいえる写真を分析することで、個人の嗜好を推測できるのではないかと考えた。

我々は、ユーザが過去に撮影した旅行写真から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研究に着手している。本研究のターゲットユーザの例として、過去に旅行などで写真をたくさん撮影し、特定の都市や地域を訪れることを考え、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ユーザを想定している。写真を利用した観光地推薦やトラベルルート推薦手法は数多く研究されているが、写真に付与されている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。そこで本報告では、位置情報を使わずに、旅行写真に対して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手法を提案する。

ところで、誰もが手軽に写真を撮影できるようになった一方で、過去に撮影された大量の写真についてその全貌を必ずしもユーザが詳細に把握しているとは限らない。そこで本報告では、旅行写真の被写体情報にもとづいて大量の旅行写真を構造化し、旅行目的に応じて被写体ごとに写真を一覧表示するインタ

フェースを提案する。

このユーザインタフェースでは、ノードをクラスタリングして配置するグラフ可視化手法 Koala [1] を拡張して、一般物体認識から得られたキーワードの関係性を可視化する。本手法ではこの可視化結果に写真を追加して表示することで、旅行目的で旅行写真を整理して提示する。その画面上でユーザが特に関心のある写真を選択すると、本手法は選択された写真の被写体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の情報を抽出し、ユーザに提示する。

本報告の構成は以下の通りである。2章では関連研究について述べ、3章では提案手法を述べる。4章では本手法の実行結果について述べる。5章では本研究で得られた成果のまとめと今後の展望について述べる。

2. 関連研究

写真を利用した旅行推薦システムに関する研究は数多く発表されており、その多くは写真のジオタグ情報を用いている [2-5]。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トラベルルート推薦を提案している。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ため、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] は、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旅行ガイドシステムを提案している。また、Cao ら [5] は、撮影場所をもとに写真をクラスタリングし、各グループ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とをマッチングを行う手法を提案している。

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。そこで本報告では、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提案する。なお本研究では、旅行の目的地となる都市や地域は決まっているが、具体的にどの都市や地域の中でのどのような観光名所を訪れるか決まっていないという状況を想定して、その都市や地域における観光名所を推薦するシステムの開発を目指している。

3. 提案手法

本手法では、写真に付与されたジオタグ情報を使わずに、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る。このキーワードを検索ワードとし、Google Places API Web Service [6] を利用して特定の都市や地域における観光情報を抽出する。また、物体認識結果より得られたキーワード間の共起関係を Koala で可視化し、可視化されたネットワーク構造を JSON 形式で出力する。そして、キーワード間の共起関係を cola.js [7] を利用してグラフ描画することで、ユーザ操作によって旅行目的を選択できるユーザインタフェースを提供する。

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、我々は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。そこで本研究では位置情報が付与されていない写真に対して、一般物体認識を用いて被写体の情報を取得する。我々の実装では、Microsoft の画像認識 API である Computer Vision API [8] を用いることで一般物体認識の情報を取得する。Computer Vision API では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。

3.2 物体認識結果を利用したグラフ可視化

本手法のユーザインタフェースでは、キーワードをノードとしたグラフを構築し、これを可視化する。 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ノードのクラスタリングを実行する。

- 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。

以上のクラスタリングを施したグラフを Koala により画面配置する。

3.3 旅行目的を選択するためのユーザインタフェース

Koala より得られたグラフ配置結果を JSON 形式で出力し、cola.js を利用してユーザインタフェースを提供する。cola.js では、グラフ描画する際にノードをできるだけ重ならないようにするなどの制約を加えることができる JavaScript のライブラリである。本手法では、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することができる。

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。

エッジ数が多いノードであればあるほど、そのノードにおけるキーワードは写真に付与される確率の高いキーワードであり、極めて一般性の高い単語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ノードをユーザに選択してもらい、それに関連する、より具体的なキーワードが含まれるノードを表示することで、ユーザの旅行目的に沿った形でキーワードを絞り込める。逆に、旅行目的に沿わないノードを観光地推薦の対処から除外することもできる。

3.4 観光地の検索

我々の実装では、観光地情報の抽出に、特定の地域やキーワードを指定して場所を検索できる Google Places API Web

Serviceを使用する。これらのAPIを用いることで、指定したキーワードを含む情報が、場所の名前やタイプ、住所、ユーザーのクチコミなどと照らし合わせて検索される。例えばテキスト検索では「東京のピザ屋」などを検索できる。本手法では3.3節のユーザインタフェースを用いて、ユーザに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。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情報もみられる。そこで本手法では、有用なキーワードに改善するために、クエリとして与えられたキーワードによって得られた推薦結果からユーザが適合・不適合の評価を行い、クエリを修正する適合性フィードバックを用いることを検討している。

4. 実行結果

本章では、実際の旅行写真 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 cola.js でグラフを描画した。物体認識により得られたキーワードの総数は延べ 145,000 個で、本実験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。

まず、Koala を拡張してグラフ可視化し、ハブノードにあたるキーワードを抽出した例を図 1 に示す。ハブノードはエッジ数が多いクラスターであるため、一般性の高い単語が抽出されていることがわかる。図 1 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あたるものであり、outdoor や sky など、極めて一般性の高い単語が抽出されている。

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 2 に示す。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 2 の (a) は山に関連するキーワード、(b) は動物園や水族館に関連するキーワードである。

次に、このグラフを JSON 形式で出力し、cola.js で表示した例（初期画面）を図 3 に示す。最もエッジ数の多いルートノードが中央に配置され、ルートノードからエッジが張られているハブノードには、各ハブノードのキーワードが付与された写真を 4 枚選んでノードとして配置している。このユーザインタフェースの操作手順として、旅行目的に沿ったキーワードに対応するノード、もしくは写真ノードをクリック操作で選択してもらい、観光地推薦のキーワードとしてプレイスライブラリに入力することを想定している。

5. まとめと今後の課題

本報告では、旅行写真の被写体情報を利用した観光地推薦手法と、ユーザ操作によって旅行目的を絞り込むためのユーザインタフェースを提案した。本手法では、まず過去に撮影された旅行写真に対して物体認識を適用し、被写体に関するキーワードを集める。得られた物体認識結果をもとに、キーワードの共起関係をグラフ可視化することで、旅行目的に関連するクラスターを形成する。ユーザインタフェースでは、このグラフ可視化

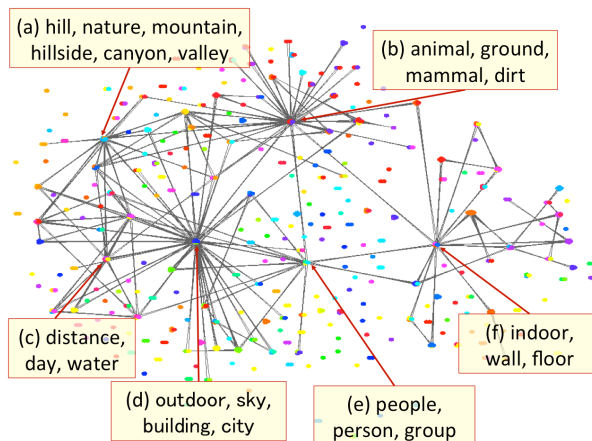


図 1 ハブノードのキーワード

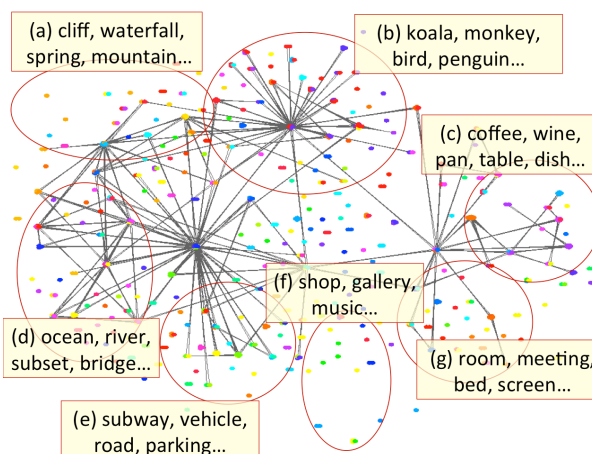


図 2 旅行カテゴリを表すキーワード



図 3 ユーザインタフェースの初期画面

結果から得られたクラスターを 1 つのノードとして表現し、各ノードのキーワードが付与された写真をノードとして表示することで、過去にどのような写真を撮影されているのかを提示し、旅行目的を絞り込めるようにする。

本研究はまだ観光地の検索機能を含めて実装が完成していない。そこで当面はユーザインタフェースの実装の完成を課題とする。また、現段階の実装では写真ノードとしてユーザに提示

する写真を、各ノードに対応するキーワードが付与された写真の中からランダムに選択しているため、写真の選出手法について検討したい。そのほかの課題としては、ユーザ操作による重み付けと可視化されたネットワークの接続構造からキーワードの重み付けを計算する手法を検討したい。

謝 辞

本研究を進めるにあたり、東京工業大学の榊惇志助教に大変有用なアドバイスをいただきました。深く感謝いたします。

文 献

- [1] T. Itoh, K. Klein, “Key-node-Separated Graph Clustering and Layout for Human Relationship Graph Visualization”, IEEE Computer Graphics and Applications, Vol. 35, No. 6, pp. 30-40, 2015.
- [2] T. Kurashima, T. Iwata, G. Irie and K. Fujimura, “Travel route recommendation using geotags in photo sharing sites”, In Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 579-588, 2010.
- [3] A.-J. Cheng, Y.-Y. Chen, Y.-T. Huang, W. H. Hsu and H.-Y. M. Liao, “Personalized travel recommendation by mining people attributes from community-contributed photos”, In Proceedings of the 19th ACM international conference on Multimedia, pp. 83-92, 2011.
- [4] Y. Gao, J. Tang, R. Hong, Q. Dai, T.-S. Chua and R. Jain, “W2Go: a travel guidance system by automatic landmark ranking”, In Proceedings of the 18th ACM international conference on Multimedia, pp. 123-132, 2010.
- [5] L. Cao, J. Luo and A. Gallagher, X. Jin, J. Han and T. S. Huang, “A worldwide tourism recommendation system based on geotagged web photos”, IEEE International Conference on Acoustics Speech and Signal Processing, pp. 2274-2277, 2010.
- [6] Google Places API Web Service, <https://developers.google.com/places/web-service/intro?hl=ja>.
- [7] cola.js, <http://marvl.infotech.monash.edu/webcola/>.
- [8] Computer Vision API, <https://azure.microsoft.com/ja-jp/services/cognitive-services/computer-vision/>.