

位置情報 SNS 上の画像–テキスト間対応を利用した ユーザ嗜好抽出と推薦スポット候補拡張

大東 祐太^{1,a)} 有山 俊一郎^{1,b)} 延原 肇^{1,c)}

受付日 2017年3月15日, 採録日 2017年9月5日

概要: スポット情報推薦サービスにおいて、スポットに紐づくテキスト情報が不足することによって推薦候補として選択されない推薦被覆率の低下問題を解決するために、画像–テキスト間対応を利用したスポット情報拡張手法を提案する。提案手法では、スポットに投稿されるテキストが少なく、その一方で画像が豊富に投稿されている点に着目し、それらの画像から抽出する SIFT 特徴量に基づく BoVW (Bag of Visual Words) を当該スポットの中核特徴として採用する。また、この BoVW と当該スポットに投稿されているテキストを対応させることで、画像–テキスト間対応を構成する。これにより、スポットにテキストが投稿されていない場合においても、投稿されている画像さえあれば、画像–テキスト間対応を用いることにより、テキストを紐付けることができる。さらに、ユーザの訪問したスポット履歴を BoVW により特徴付けることが可能となり、これによるスポット情報推薦の高精度化、また画像–テキスト間対応から獲得できるテキストをユーザ側に明示的にフィードバック可能となる。提案手法の推薦精度の向上、およびユーザへの明示的なフィードバックの機能が有効であることを示すために、3 種類の主観評価実験を行う。

キーワード: 情報推薦, 画像–テキスト間対応, Bag of Visual Words, 情報拡張

An Extension of Location Information in Social Service Based on Relation between Images and Text and their Application to Coverage Improvement of Recommendation

YUTA OOHIGASHI^{1,a)} SHUNICHIRO ARIYAMA^{1,b)} HAJIME NOBUHARA^{1,c)}

Received: March 15, 2017, Accepted: September 5, 2017

Abstract: To achieve local location recommendation based on user's current position obtained by smart phone GPS and various locations (such as stores, tourist site, and venues) on social service, an extension of location information is proposed by using images submitted to the locations. In the location recommendation, it is quite difficult to characterize the location and users profile, because the posted texts in the social services are small in general. This paper presents the translation of images to Bag of visual words to use them instead of text information of each location. The number of images submitted to location is enough for characterize the feature of location and profile of users. Through three evaluation experiments, it is confirmed the effectiveness of the proposed method, especially, it increases the coverage of recommendation and accuracy of obtaining feature words represent user from histories.

Keywords: recommendation, relation between images and text, Bag of Visual Words, information expansion

1. はじめに

現在、スマートフォンの GPS 機能、位置情報を利用し

たサービスが数多く普及している [1]. これらのサービスでは、ユーザの現在位置情報に基づき、周辺の店舗や、観光スポットなどを検索することができる。よって、従来の住所入力や駅をヒントに検索する方法と比べて現在地からの検索効率は向上する。一方で、都心などの地域では推薦候補となるスポットが多すぎるため、ユーザの嗜好に合わせて候補を絞り込む推薦技術が必要である。現在、著者ら

¹ 筑波大学

University of Tsukuba, Tsukuba, Ibaraki 305–8573, Japan

a) oohigashi@cmu.iit.tsukuba.ac.jp

b) ariyama@cmu.iit.tsukuba.ac.jp

c) nobuhara@iit.tsukuba.ac.jp



図 1 FourDiary [2] のタイムライン
 Fig. 1 An example image of FourDiary.

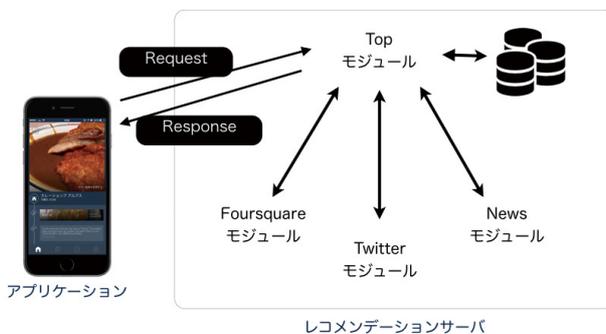


図 2 FourDiary のシステム構成
 Fig. 2 System overview of FourDiary.

はユーザに発見的なスポットの推薦を行うために、日々の行動を自動的に記録し、それらに基づいて推薦が行われるアプリケーション、FourDiary [2] をリリース・継続開発している。このアプリケーションは、ユーザが位置情報を取得可能な端末を日々の生活の中で持ち歩くことで、その中に訪れた場所の情報を自動的に記録する (図 1)。

位置情報の他にも端末で撮影した位置情報付きの画像に基づき、ユーザが訪れた場所の記録を行うことができる。このライフログアプリケーションでは、端末が位置情報を自動的に取得することで、ユーザが意識することなく、訪問先の履歴を獲得・記録できることがあげられる。また、ユーザの位置情報に合わせて、位置情報を利用したソーシャルサービス上のスポット、地域関連ニュース、マイクロブログなどの多様なコンテンツを推薦できる点もあげられる。本研究では、FourDiary の複数の推薦機能のうち、端末の位置情報を利用したスポット情報推薦に着目する (図 2)。ここで、スポット情報推薦は、Foursquare [3] などのソーシャルサービス上から取得したスポット群 (店舗・観光地・公共施設など) を、ユーザの行動履歴に合わせて推

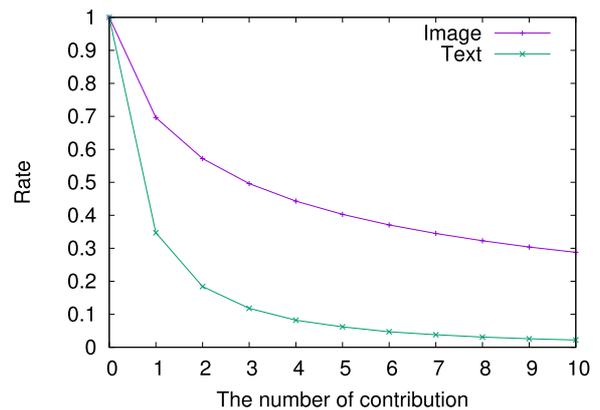


図 3 スポットに対する画像とテキストの投稿割合
 Fig. 3 Contribution ratio of image and text submission.

薦を行う機能に対応する。スポット情報推薦を有効に機能させるための重要な要素として、スポットを説明するコンテンツが豊富に紐付けられていることがあげられる。しかし、Foursquare に登録されているスポット情報のうち都内のスポット約 130,000 件に対して、ユーザ投稿のコンテンツを調査したところ、スポット 1 件あたりの画像投稿数は平均 23.6 件に対し、テキストは平均 1.1 件、画像が一枚でも存在するスポットは 7 割弱存在するのに対し、テキスト (レビュー) が 1 件でも存在するスポットは 3 割強という結果が得られた (図 3)。つまり、ここで問題となるのは、それらのスポット関連情報に投稿されているテキストが少ないため、スポット自体を特徴付けることが難しいこと、同様に、ユーザの嗜好を抽出する履歴に関して、スポットの関連情報に依存するため、テキストの不足および精度の高いプロファイリングが難しいことが明らかとなった。

本研究では、この問題を解決するために、テキストに比べて画像の投稿が圧倒的に多いという、位置情報ソーシャルサービスの特性を利用したスポット情報推薦手法を提案する。この手法では、まずスポットに投稿されている豊富な画像群に対し、Bag of Visual Words (BoVW) [4] としての画像特徴量を抽出する。これらをテキストの代わりに用いることでスポット情報拡張を行い、精度の高いプロファイリングおよびスポット推薦を行う。ユーザ側の嗜好のプロファイリングに関して、ユーザの移動履歴としてスポット情報が利用できるため、これらのスポットの画像の BoVW を集約することで、当該ユーザの嗜好を表現する BoVW を構成できる。さらに、Foursquare のスポットのうち、同一スポットに対してテキストと画像の両方が投稿されている場合に、その特徴語 (テキスト情報) と BoVW が対応すると仮定し、双方向に変換可能な対応を構成する。ユーザの履歴として BoVW が蓄積された際に、構成した対応を利用して特徴語に変換し、その出現頻度などから、当該ユーザの特徴語を抽出することができる。これにより画像ベースの推薦における、ユーザへの明示的なタグ付与

やそれに基づくフィードバックを得られない問題を解決する。提案手法の有効性を確認するため、Foursquareの飲食店カテゴリに属するスポットの情報を収集し、各スポットの関連情報として、テキストのTipsとサムネイルとして利用されている画像を用いたシステムによる評価実験を行う。2章では、位置情報に基づく推薦に関連する研究および、それらの課題について述べる。3章では、そのうちのスポット情報が不足している問題を取り上げ、それを解決するために提案する画像特徴による情報拡張を用いたスポット情報推薦の手法について述べる。4章では提案手法の有効性を確認するためにスポットの特徴語と画像特徴量の関係性があるか否かを検証し、5章で結論を述べる。

2. 関連研究

2.1 協調フィルタリング

篠田らの位置情報を用いた行動履歴に基づく行動ナビゲーションの研究 [5] では協調フィルタリングを用いた推薦を行っている。ここで用いている協調フィルタリングでは潜在的な興味を予測して推薦するため、ユーザが持っている嗜好とは異なるジャンルのスポットも検索対象に含まれ、意外性のある推薦が期待できる。その反面ユーザから評価されていないスポットは推薦の対象にできないため、被覆率を高くできない。また、ユーザがどのような嗜好を持つか考慮せずに推薦を決定するため、選択理由が不明瞭になってしまう。

2.2 トピックモデル [6]

トピックモデルではコンテンツの特徴を潜在トピックとして学習し、同時にユーザがどのような嗜好を持つか潜在トピックの選択確率として学習する。ユーザがどのような嗜好を持っているためにどのようなコンテンツを選択するかが明確になり、ユーザに提示することでフィードバックを得ることができるといったメリットがある。倉島らの研究 [6] ではトピックモデルを応用し実空間の位置関係も考慮したジオトピックモデルを採用し情報推薦を行っている。ジオトピックモデルでは現在位置を入力として、ユーザの嗜好に合った推薦を行うことができる。この手法の評価ではあらかじめ位置情報付き写真にタグ付けがなされていることが前提で、スポット情報サービスで扱われる、スポットの情報にユーザからのテキストの投稿がない限り、その情報を用いることができないため、被覆率が下がってしまう。

2.3 GPSによる軌跡のマイニング [7]

一定間隔ごとに記録した位置情報の記録からユーザの次の行動を予測する。この手法では旅行先の王道ルートを提案したり、ある行動に起因する短期的な行動予測に基づく推薦を行ったりする。この手法は本提案手法に類似しているが、周辺スポットの特性を考慮し、推薦の選択肢を広く

することで多様性を確保し、さらにデータが疎になることで被覆率 (=推薦するコンテンツとして評価可能なアイテムの割合) が下がってしまう問題も緩和している。常時GPSログをとることをユーザは嫌うため、FourDiaryで扱うような、ユーザが訪問した先の履歴のみ利用する場合、軌跡が利用できず、行動の予測が立てられない。

2.4 画像の自動タグ付け

画像中の物体を認識して、自動でタグ付けを行う研究がされている [8], [9]。これらの手法では、画像中の物体とその名称を学習したうえで、対象画像にタグ付けを行う。一方、本研究では扱うコンテンツは様々な地域で撮影された位置情報と紐付く画像である。ユーザが好みのスポットを探す場合、たとえば、ラーメン屋を求めているとき、ただラーメンを提示するより、どのようなラーメンを推薦するかが必要になる。そのため、画像に対してタグ付けをした語ではユーザの嗜好に合わせた粒度で推薦ができない。そこで本研究では、スポットの特徴としてユーザが投稿した画像と、テキストを用いる。テキストがない場合でも推薦ができるように画像特徴量によってユーザプロフィールとの類似性の評価を行う。また、スポットの画像と、投稿されたテキストの特徴語群とを関連付けておき、ユーザの行動履歴が蓄積された段階で、テキスト情報のないスポットからもユーザを表す特徴語を得ることを目的としている。

2.5 本研究着手の動機

上述で示したような従来の様々な推薦手法ではスポットを特徴付けるためのタグかテキスト、あるいは、膨大なユーザの行動履歴が必要となる。本研究ではFoursquareなどの位置ソーシャルサービス上から取得したスポットと、ユーザの行動履歴から嗜好を抽出する。しかし、スポット関連情報にテキスト情報が少ないため、スポットを特徴付けることが難しい点を解決しなければ、精度の高い推薦を行うことができない。そこで、本研究ではテキストの代わりに、膨大に存在する画像を用いて、スポットの情報拡張を行い、特徴付けを行うことで精度の高い推薦を可能にする。また、テキストと画像の両方を持つスポットを利用し、画像特徴量と特徴語の対応を構成する。これにより、ユーザの履歴から、明示的な嗜好ラベルの抽出を可能にする。これは、従来のテキスト情報がないゆえに問題となっている、ユーザへの明示的な嗜好ラベルのフィードバックが得られない問題を本質的に解決することにつながる。

3. 提案手法

提案する画像特徴およびユーザ特徴抽出を図4に示す。画像特徴による情報拡張を用いたスポット情報推薦の流れを図5に示す。

提案手法では、まず、事前にスポットに投稿された画像

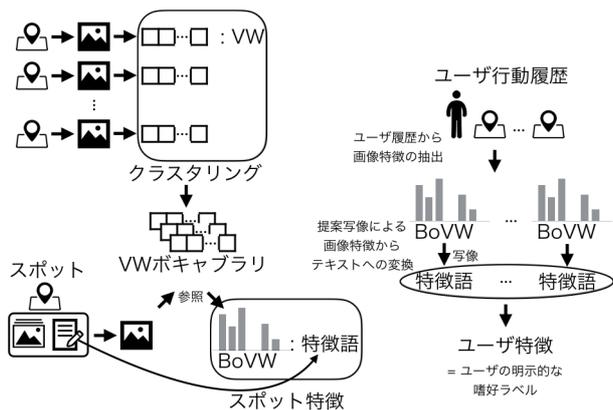


図 4 提案手法の概要 (スポットとユーザーの特徴付け)
Fig. 4 Overview of proposed method.

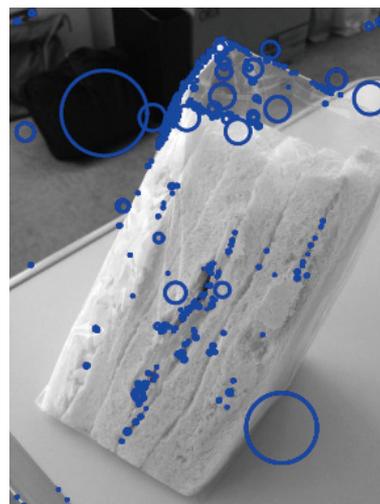


図 6 SIFT 特徴量抽出の一例
Fig. 6 An example of SIFT feature extraction.

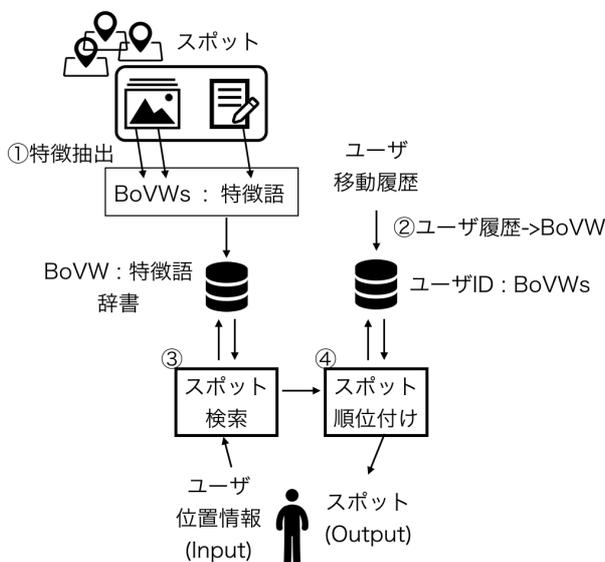


図 5 提案手法の流れ
Fig. 5 Flow of proposed method.

から BoVW を抽出、テキストから特徴語を抽出し、それぞれ紐付けた状態でデータベースに保存しておく。ユーザーのプロファイリングは訪れたスポットの画像から抽出した BoVW の集合として保存する。提案する推薦システムの入力としてユーザーの位置座標が与えられ、その位置情報に基づき、データベースから、付近のスポットを検索する。得られた複数のスポットの中で、ユーザーのプロファイルの BoVW とスポットが持つ BoVW 間で類似度を算出し順位付けを行う。順位の高いものほど、当該ユーザーの嗜好に相当であると判断し、出力のスポットとして出力する。

3.1 画像特徴による情報拡張を用いたスポット情報推薦

ニュース記事のようなテキストデータを扱う推薦モデルではテキストマイニングによりニュース記事を特徴ベクトル化し、ユーザーが好みを示したのからプロファイリング、推薦するコンテンツの評価が行われる。ベクトル化する手法として、テキスト文書中に存在する単語の出現頻度の

ヒストグラムのベクトルで表現する Bag-of-Words (BoW) モデルが情報推薦において多く用いられている。このテキストマイニング手法を画像の視覚的内容の解析に応用した手法が Bag-of-Visual-Words (BoVW) [4] モデルである。BoVW モデルでは単語に代わる視覚的要素として、局所記述子を用い、記述子の典型例を作成する。画像中の記述子を典型例で当てはめていくことで画像が表現されていると考える。この 1 つの典型例をビジュアルワードと呼び、ある 1 つのシステムのモデルで使用する典型例の集合をビジュアルボキャブラリと呼ぶ。また、ユーザーによって投稿される画像はサイズ、解像度などのフォーマットが統一されていないため、画像中に存在する記述子のヒストグラムでモデル化できる BoVW を採用する。提案手法で用いる BoVW の構成手順は、まず、各スポットに対してユーザーにより投稿された全画像から SIFT (Scale-Invariant Feature Transform) 特徴量を算出し、画像特徴量とする (図 6)。

1 枚の画像から平均 1,500 個程度の SIFT 特徴量が得られるため、たとえば数千枚の画像の全 SIFT 特徴量からそのまま BoVW を構成すると膨大な次元になってしまう。そこで、文献 [12], [13] で示される K-means クラスタリングを用いた VW の量を削減および最終的な BoVW の次元数を削減する枠組みを利用する。具体的には、SIFT 特徴量すべてを対象に K-means クラスタリングを行い、各クラスタのセントロイドを VW として定義し、VW のボキャブラリを構成する (図 7)。クラスタ数 k の推薦精度への影響について、我々が現実的に取り扱い可能な 500 から 2,000 次元で調査したところ、それほど大きな変動はなく、高次元になると若干精度が高くなるという結果が得られている。この結果に基づき、本論文では特に明示しない限り、一般的なコンピュータ (16 GB メモリ程度) で取り扱うことのできる最大の次元数 $k = 1000$ を採用することとする。文献 [12], [13] では、BoVW を画像のシーンの分類および

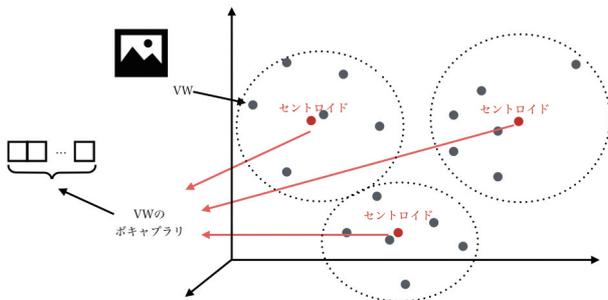


図 7 クラスタリングによる VW のボキャブラリの構成
 Fig. 7 Construction of VW vocabularies by clustering.

情報検索へ応用しているが、本研究では、さらにこの枠組みを発展させ、BoVW を利用して画像と対応するテキスト特徴語の対応を構築する。これによって、テキスト情報の付与されていないスポットに対して、新たに特徴的なテキスト情報を付与することができたり、また、ユーザの訪問したスポット画像から、そのユーザの嗜好を明示的なテキスト情報として示すことができたりするようになる。本研究で SIFT 特徴量を採用するのは、スケール、回転、明度、ノイズに対してロバスト性があり [4]、画像に写っている物体の特徴を抽出するための局所特徴量として適しているためである。

各スポットには、複数の画像が投稿されており、これらの SIFT 特徴量を算出、VW のボキャブラリの中で最も類似する VW で置き換えて BoVW を構成し、当該スポットを特徴付ける情報とする。

[スポット情報拡張]

以上の VW のボキャブラリ構成の定式化を含め、提案システムにおけるスポット情報拡張について説明する。まず推薦対象となる N 個の i 番目のスポット s_i の集合を $S = \{s_i | i = 1, \dots, N\}$ と表現でき、 i 番目のスポットは Tips から抽出された特徴語の集合と、画像から抽出された BoVW の集合を持ち、 $s_i = \{F_i, V_i\}$ と表現できる。ここで特徴語の集合は $F_i = \{f_1^{(i)}, \dots, f_{K(i)}^{(i)}\}$ で、BoVW の集合は $V_i = \{v_1^{(i)}, \dots, v_{L(i)}^{(i)}\}$ 、ここで $v_l^{(i)} \in \mathbf{Z}_+^{1000}$ と表すことができる (\mathbf{Z}_+ は正の整数を表す)。1 章の約 130,000 件のスポットの調査で明らかになったように、多くのスポットでは、画像情報は比較的投稿されている一方、テキスト情報はほとんど投稿されていない。すなわち、従来のテキストのみの情報推薦の場合 $s_i = \{F_i\}$ でスポットは特徴付けられるが、 $|F_i|$ ($||$ は集合の濃度を表す) が小さいため、推薦が機能しない。本研究では、 $s_i = \{F_i, V_i\}$ という形で新たに画像から得られる V_i を追加することで情報拡張を行い、さらに $|V_i|$ が $|F_i|$ に比べて大きくなるため、推薦候補を獲得しやすくなり、推薦被覆率を向上させることができる。

[画像-特徴語対応の構成およびスポットへのテキスト情報付与]

Foursquare のスポットのうち、同一スポットに対して

テキストと画像の両方が存在する場合、そのテキスト情報の特徴語が BoVW と対応すると仮定し、それぞれから相互に変換可能な対応を構成する。つまり、 $s_i = \{F_i, V_i\}$ と表現されるスポットにおいて、 F_i と V_i は相互に参照できるようにする。BoVW から参照する場合は、その BoVW を持つスポットの複数の特徴語が引用され、特徴語から参照する場合は、その特徴語を持つスポットの複数の画像の BoVW が引用されることになる。この対応を利用することで、たとえば、あるスポット $s_i = \{F_i, V_i\}$ に関して、 $F_i = \emptyset$ の場合においても、テキストとしての特徴語を付与することができる。具体的な手順を以下に示す。

まず、任意の 2 つの BoVW のベクトル $v_m, v_n \in \mathbf{Z}_+^{1000}$ の非類似度を、それらのユークリッド距離

$$d(v_m, v_n) = \left(\sum_{k=1}^{1000} (v_m[k] - v_n[k])^2 \right)^{1/2} \quad (1)$$

で定義する。ここで $v_m[k]$ は v_m の k 番目の要素を表す。

次に、テキスト情報が投稿されていない $F_i = \emptyset$ なるスポット s_i の BoVW ベクトル $v_m \in V_i$ ($V_i \neq \emptyset$) と、 s_i 以外の BoVW ベクトル $v_n \in (V_1 \cup V_2 \cup \dots \cup V_N) \setminus V_i$ との非類似度 (式 (1)) を計算し、最小となる (v_m, v_n) のペアを見つける。このときの v_n が属する V を持つスポットの F をテキスト情報として付与する。ここで、当該スポットのテキスト情報 F が空の場合には、次に最小となる (v_m, v_n) のペアを見つけテキストの付与を行い、テキストが割り当てられるまでこれを繰り返す。

[ユーザの嗜好のプロファイリング]

移動履歴として蓄積されたスポット情報を利用し、ユーザの嗜好のプロファイリングも行う。具体的には、当該スポットの履歴画像の BoVW を集約し、ユーザの嗜好を表現する BoVW の集合を構成することで実現する。全 M 人の m 番目のユーザ u_m の集合を $U = \{u_m | m = 1, \dots, M\}$ と表現する。各ユーザ u_m の行動履歴は $h(u_m)$ で表し、ユーザのこれまでに訪れたスポットの集合を

$$h(u_m) \in S^{|h(u_m)|}, \quad (2)$$

$$|h(u_m)| \in \mathbf{Z}_+, \quad (3)$$

で表す。ここで $S^{|h(u_m)|}$ は、推薦対象のスポット集合 S の直積集合、 $|h(u_m)|$ は $h(u_m)$ の濃度、 \mathbf{Z}_+ は正の整数を表す。ユーザの嗜好として、履歴の系列 $h(u_m)$ そのものを使ってもよいし、これらの系列に含まれるスポットに対応する BoVW の平均を算出するなどしてもよい。また、画像-特徴語対応を用いることにより、 $h(u_m)$ の各スポットにはテキスト情報が付与されているので、それらを集約した

$$\cup F_k \quad (F_k \in s_k, \forall s_k \in h(u_m)) \quad (4)$$

が、当該ユーザを表すテキスト情報となる。これらを利用することにより、ユーザへの明示的なタグ付与や、フィー

ドバックを提供することができる。ユーザの履歴によりプロファイリングする際にはコールドスタート問題が指摘される。新規ユーザの場合は行動履歴が蓄積されておらず、画像によるユーザのプロファイル定義が行えないため、適切な推薦を提示することができない。そこで、近年主流になってきているソーシャルアカウントによるログインから得られる、各サービスで投稿されているテキストデータからユーザの特徴語を抽出し、画像-特徴語間対応に基づき特徴語から BoVW に変換することで初期プロファイリングを行うことが可能になる。また、ソーシャルアカウントによるログインを行わない場合においても、本手法の画像をベースの BoVW、さらに画像-特徴語対応を利用することで、各スポットに豊富に投稿されている画像情報から、比較的少ない訪問履歴だけで、ユーザの嗜好に合ったスポット推薦を開始することができ、コールドスタート問題を解決することができる。また、ユーザのプロファイルがうまくなされているかを判断するために、明示的なフィードバックを必要とすることがある。その際に、画像-特徴語間対応を用いてユーザの BoVW から特徴語に逆変換することでユーザに提示することが可能になる。特徴語を持つスポットはそのまま対応付けができるが、テキストデータを持たないスポットも多く存在する。そこで、特徴語を持たないスポットは BoVW に基づき、最も近いスポットの特徴語を引用することで、対応する特徴語を決定する。得られた特徴語のうち最も頻度の高いものから優先的に、ユーザプロファイルを表す語として提示する。

3.2 BoVW に基づくユーザプロファイリングおよびスポット推薦

提案手法に基づいた、ユーザの嗜好に類似するスポットの推薦手法について述べる。この方法では、まずユーザがあるスポットに訪問し他のスポットへの推薦を必要とするタイミングで、現在地周辺に存在する画像群の BoVW とユーザのプロファイルの BoVW 間でユークリッド距離を計算する。これらを逆数でスコアリングし、その合計値を各スポットの類似度として定義しランキングする。推薦はランキングに基づき上位のスポットから優先的にユーザに対して提示する。以下、これらを定式化したものを示す。ユーザ u_m に対する、スポット s_i の 1 つの BoVW である $v_i^{(i)}$ の評価値は

$$Score_{v_i^{(i)}} = \sum_{k=1}^{|h(u_m)|} \sqrt{|v_i^{(i)} - RepV(h(u_m)[k])|^2}, \quad (5)$$

$$RepV(s_i) = v_1^{(i)} \quad (6)$$

で求める。ここで、 $RepV$ はスポットのサムネイルに使用されている、代表画像の BoVW を出力する関数である。この評価値を対象スポットの全画像の BoVW に対して算出し、逆数の平均値

$$RankScore_{s_i} = \frac{1}{L(s_i)} \sum_{i=1}^N \frac{1}{Score_{v_i^{(i)}}}, \quad (7)$$

でランキングする。ここで、 $L(s_i)$ は現在地周辺に存在するスポット s_i に投稿されている画像数を表す。このランキングに基づきサービスの推薦枠に合わせて優先的にユーザに対してスポットの提示をする。

4. 評価実験

提案手法の有用性を確認するために、3つの主観評価実験を行う。主観評価実験1では、提案システムが出力するコンテンツとユーザの嗜好の対応の調査を行う。主観評価実験2では、投稿された画像はあるがテキストがないスポットに対し、提案する情報拡張によってどの程度妥当なキーワードが付与されるのかを調査する。また、k-meansによる次元削減についての検討も行う。主観評価実験3では、ユーザの履歴から提案する画像-テキスト間対応を用いて明示的なラベル(テキスト)を当該ユーザにフィードバックし、それが適切であるかの調査を行う。

4.1 主観評価実験1

提案手法に基づくレコメンデーションエンジンを実装し、テスト用システム(図8)を構成する。被験者には以下の手順でシステムを利用、評価してもらう。

- (1) マップから自分の興味がある場所を、訪れたと仮定してチェックイン
- (2) (1)を繰り返す(履歴の作成)

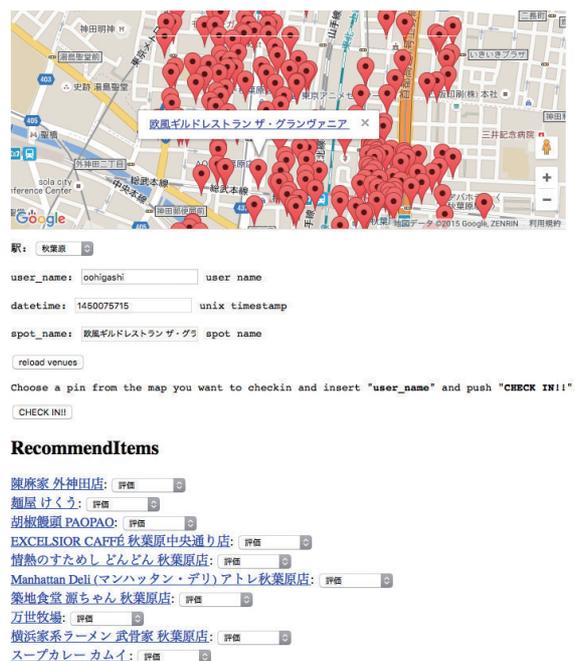


図8 主観評価実験1用のインタフェース

Fig. 8 An example of interface used in subjective experiment 1.

表 1 各ユーザの履歴数と評価値平均

Table 1 History and evaluation average of each user.

ユーザ	履歴数	評価値
1	20	2.25
2	9	2.22
3	10	1.50
4	50	1.84
5	30	1.83
6	60	1.97
7	10	2.10
8	60	1.78
9	90	1.39
10	20	2.40
11	10	2.20

(3) ある場所にチェックインしたときに提示されたレコメンドコンテンツを4段階評価

レコメンドコンテンツとして、各チェックインごとに10件を提示する。各コンテンツに対する評価は良い、まあまあ良い、まあまあ悪い、悪いの4段階で、自分の興味に合っているかどうかを評価してもらった。評価値はそれぞれ3~0点に対応させる。

この実験をユーザ11人に実施した際の履歴数および評価値の平均を表1に示す。ユーザごとの評価値の平均の総平均は、約1.953であり、中間点を上回る良好な結果が得られている。あるカフェ中心に履歴が構成されたユーザに着目してみると、レコメンドされたコンテンツの6割は同様にジャンルがカフェのコンテンツであり、画像特徴に基づいた推薦でも、ユーザの好みを反映した推薦ができていたことが確認できた。今回の実験でのレコメンド総数は369件であり、そのうち、テキストのないスポット件数は37件含まれており、テキストの平均件数は約5.5件であった。つまり、テキストのみで特徴付けをし、評価をする推薦手法では推薦できなかったコンテンツも特徴付けができ、推薦の被覆率が向上できるといえる。また、テキストのないスポットだけの評価値を見ても、平均約1.979と、全体の総平均と同等のスコアであった。そのため、テキストでは評価できないにもかかわらず、画像を用いた評価では、他のスポットと対等に評価できるようになったといえる。位置情報サービスにおいては、ユーザが物理的に移動した結果、訪問履歴が蓄積されるという性質上、できるだけ少ない訪問履歴でユーザの嗜好に合ったスポット推薦機能が駆動できるようになることが、コールドスタート問題の解決につながる。主観評価実験1では、レコメンド総数が369件であることから、被験者数が11人、1チェックインあたり10件のレコメンドがなされるので、被験者1人あたりの平均の履歴数が約3.4ということになる。この履歴数から、前述の評価値の総平均が得られたことは、提案手法が非常に少ない履歴数でも機能し、コールドスタート問



図 9 主観評価実験2用のインタフェース

Fig. 9 An example of interface used in subjective experiment 2.

表 2 各スポットに付与されたキーワードが適切であると判定された割合

Table 2 Evaluation results of validness of keywords assigned to each spot.

スポット番号	1	2	3	4	5
$k = 500$	0.00	0.07	0.00	0.54	0.25
$k = 1000$	0.00	0.07	0.14	0.54	0.11
$k = 2000$	0.29	0.82	0.74	0.24	1.00
スポット番号	6	7	8	9	10
$k = 500$	0.54	0.82	0.43	0.86	0.71
$k = 1000$	0.68	0.46	0.18	0.86	0.82
$k = 2000$	0.00	0.04	0.57	0.14	0.32

題を解決することができることを示している。

4.2 主観評価実験2

ここでは提案スポット情報拡張によるキーワード付与の有効性を検討するための主観評価実験を行う。この実験では、投稿画像はあるが、特徴語を持たないスポットに対し、提案手法によって特徴語を付与する。ユーザは、当該スポット名および画像と、付与されたキーワードを閲覧し、適切な特徴語が付与されているかを評価する。また、この実験においてk-means法のクラスタ数kの影響についても検討する。この主観評価実験で用いるスポットは10件とし、画像の投稿はあるが、レビューテキストのないスポット群からランダムに抽出した。1つのスポットにつき5~6個の特徴語を付与する。付与した特徴語はk=500, 1000, 2000の3種類で算出したものをランダムな順番で表示し、被験者はどの特徴語がどのkを用いて算出されたか分からないものとした。スポット名、画像およびそれに付与されたキーワードを評価する画面の1例を図9に示す。

この実験を被験者14人に実施してもらった結果を表2に示す。各スポットに対する評価値は、提案手法(クラスタ

表 3 各ユーザの履歴数と特徴語の評価値

Table 3 History and evaluation average of each user.

ユーザ	履歴数	評価値
A	27	2.25
B	9	2.00
C	8	1.60
D	6	1.60
E	5	0.70
F	1	1.60

数 $k = 500, 1000, 2000$) の出力したキーワードのうち適切であると選択された数の割合を、全ユーザについて平均をとったものである。これらの全スポットにおける評価の平均値をクラス数ごとにまとめると、 $k = 500, 1000, 2000$ においてそれぞれ、0.42, 0.39, 0.42 となり、 k の値によって特徴語の精度に差がないことが確認できる。3章において言及したように、我々が一般的に利用するコンピュータで計算可能な範囲のうち、 $k = 500$ と $k = 1000$ は 16 GB メモリで計算可能、 $k = 2000$ の場合は 32 GB が必要となっていることから、16 GB の範囲で最大次元数である $k = 1000$ を採用する理由にもなっている。提案手法によるキーワード付与の精度については、クラス数がいずれの場合においても約 4 割程度となっており、ある程度の精度が得られていると考える。特に提案手法が興味深いのは、一般物体認識とは異なり、画像に写っている物体そのものを表すキーワードを抽出するのではなく、画像とスポットから連想されるキーワードを付与することができる点である。具体的には、図 9 のような画像の場合、一般物体認識では「魚」、「金魚」、「水槽」といったキーワードが付与されることが予想されるが、提案手法の場合は「日本」や「夏」といった画像からは直接的に抽出することのできないキーワードが付与できている。また、これらのキーワードに対する評価も高く、それぞれ、「日本」は 0.857、「夏」は 0.786 という割合で適切であると評価されている。このことから、提案手法の有効性を確認することができる。

4.3 主観評価実験 3

提案する画像-テキスト間対応に基づき、ユーザの履歴から推測された明示的なユーザ嗜好ラベルを被験者に提示し、その有効性を検証した。ユーザ嗜好ラベルはユーザ特徴の BoVW の特徴語の頻度の高いもの最大上位 10 件を抽出した。被験者は各特徴語を良い、まあまあ良い、まあまあ悪い、悪いの 4 段階で、自分の嗜好に合っているかどうかを評価してもらった。評価値はそれぞれ 3~0 点に対応させる。

この評価を被験者 6 人に実施してもらった。各ユーザ 6 人の履歴の数、特徴語評価値を表 3 に、2 人分の特徴語サンプルを表 4 に示す。ここで、ユーザごとの評価値の平

表 4 ユーザに提示された特徴語例

Table 4 Examples of keyword assigned to each spot.

ユーザ B		ユーザ E	
特徴語	評価値	特徴語	評価値
メイドさん	3	フルーツ	2
100 円	1	ジャズ	0
土曜日	1	ジュース	1
トッピング	2	ちょっと	0
チキン	2	仕切り	2
コーヒー	3	いない	0
メニュー	2	ルーム	2
スープカレー	3	スペース	0
スパイス	2	タバコ	0
パノラマ	1	そちら	0

均の総平均を計算すると、約 1.608 であり、中間点を上回る良好な結果が得られている。また、履歴数が多いほど、評価値が高くなる傾向があり、履歴数が 5 よりも多くなると、ユーザの嗜好に適したスポット推薦が有効に機能していることが分かる。表 2 に示す特徴語例からは、「スープカレー」や「コーヒー」など、嗜好を表しているような特徴語であると評価が高かった一方で、「スペース」や「パノラマ」などその語だけでは何を表すか分からないものが含まれており評価が低かった。また、「その他」や「そちら」のような特徴語として明らかに不適切な語も含まれていた。蓄積された履歴数が少ないことも影響していると考えられるが、今回は単純に形態素解析で得られた品詞のみで構成しているため、最適な語を得るための特徴語抽出の手法を検討しなければならない。

5. まとめ

本研究では画像特徴による情報拡張を用いたスポット情報推薦の手法を提案した。提案手法は、各スポット画像およびユーザの訪問した履歴スポットから画像特徴量を抽出し、それらを BoVW として利用することで、ユーザの嗜好に合わせた画像ベースのスポット情報推薦を実現している。さらに、画像情報とテキスト情報が投稿されているスポットから、画像-特徴語の対応を構成する手法も提案した。これにより、スポットの関連情報としてテキスト投稿情報が少なく、代わりに画像が多く投稿されているようなサービスにおいて、スポットの情報拡張やユーザプロファイリングが困難になる問題を解決できる。また、提案する画像-特徴語対応を用いれば、テキスト情報が存在しないために推薦対象として扱うことのできなかつたスポットでも、画像が存在しさえすればその画像に基づいて特徴付与ができるため、推薦の被覆率を向上させることができる。提案手法を実装したテストシステムを用いて主観評価実験を行い、テキストを持たないコンテンツを推薦することができ、推薦の精度を担保しつつ、推薦の被覆率が向上

できることを確認することができた。また、提案手法の特徴である、画像をベースの BoVW、さらに画像-特徴語対応を利用することで、各スポットに豊富に投稿されている画像情報から、比較的少ない訪問履歴だけで、ユーザの嗜好に合ったスポット推薦を開始することができ、コールドスタート問題を解決することができることも示した。さらに、比較的少数の履歴でも、当該ユーザの嗜好を明示的なテキスト情報として示すことができることも確認した。また、この実験では色情報を扱わない SIFT での特徴抽出を用いたが、実際に類似度が高く評価された画像を確認した結果、色情報の重要性が確認できた。今後は画像中の色情報もともに扱うことでさらなる精度向上が期待できる。今後の評価では、FourDiary のサービス上で多数のユーザによるコンテンツに対する、コンテンツのタップや、行動履歴への影響などのアクションに基づいた評価も行いさらなる有効性を検証する予定である。

参考文献

- [1] 齊藤 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, pp.234–239 (2011).
- [2] FourDiary, available from <fourdiary.com>.
- [3] Foursquare, available from <foursquare.com>.
- [4] Lowe, D.G.: Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, Vol.60, No.2, pp.91–110 (2004).
- [5] 篠田裕之, 竹内 亨, 寺西裕一, 春本 要, 下條真司: 行動履歴に基づく協調フィルタリングによる行動ナビゲーション手法, 情報処理学会研究報告グループウェアとネットワークサービス, pp.87–92 (2007).
- [6] 倉島 健, 岩田具治, 星出高秀, 高屋典子, 藤村 孝: 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 (Mar. 2013).
- [7] Ashbrook, D. and Starner, T.: Using GPS to Learn Significant Locations and Predict Movement Across Multiple User, *Personal and Ubiquitous Computing*, Vol.7, No.5, pp.275–286 (2003).
- [8] Jeon, J., Lavrenko, V. and Manmatha, R.: Automatic Image Annotation and Retrieval using Cross-Media Relevance Models, *Proc. 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, SIGIR '03*, pp.119–126, ACM (2003).
- [9] Barnard, K. and Forsyth, D.: Learning the Semantics of Words and Pictures, *ICCV* (2001).
- [10] Solem, J.E.: *Programming Computer Vision with Python*, O'Reilly (2012).
- [11] Bao, J., Zheng, Y., Wikie, D. and Mokbel, M.: Recommendations in Location-based Social Networks: A Survey, *GeoInformatica*, Vol.19, No.3, pp.525–565 (2015).
- [12] Yang, J., Jiang, Y.-G., Hauptmann, A.G. and Ngo, C.-W.: Evaluating bag-of-visual-words representations in scene classification, *Proc. International Workshop on Multimedia Information Retrieval*, pp.197–206 (2007).
- [13] Chen, X., Hu, X. and Shen, X.: Spatial weighting for bag-of-visual-words and its application in content-based image retrieval, *Proc. Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp.867–874 (2009).



大東 祐太

2016 年筑波大学大学院システム情報工学研究科知能機能システム専攻修士課程修了。現在、IT 企業において Web 関連事業に従事。



有山 俊一郎

2016 年筑波大学理工学群工学システム学類卒業。現在、同大学大学院システム情報工学研究科知能機能システム専攻修士課程在学中。



延原 肇 (正会員)

2002 年東京工業大学大学院総合理工学研究科修了 (博士 (工学))。2002 年カナダ国アルバータ大学博士研究員。2002 年東京工業大学大学院総合理工学研究科助手。2006 年筑波大学大学院システム情報工学研究科講師。2013 年筑波大学システム情報系准教授 (現職)。現在、計算知能、ウェブ・インテリジェンス、離散数理に従事。