

# Twitter プロフィールを用いた POI 公式アカウント抽出

落合 桂<sup>1,2,a)</sup> 松尾 豊<sup>2</sup>

**概要:** Twitter では一般ユーザだけでなく有名人、観光スポットや商業施設などの POI 公式アカウントなど様々なユーザが投稿している。POI 公式アカウントではイベントやセール情報など、ユーザにとって有益な内容を投稿しているため POI 公式アカウントを特定することは有用である。POI データベースを使って公式アカウントを判定する方法では、元となる POI データベースにない POI は抽出されないという課題がある。そこで本研究では、Twitter のプロフィール情報から生成した素性に基づき機械学習により POI 公式アカウントを判定する方法を提案する。

## 1. はじめに

Twitter や Facebook, LINE などのソーシャルネットワークワーキングサービス（以下、SNS）が普及している。SNS の 1 つである Twitter では一般ユーザだけでなく有名人や企業、観光スポットや商業施設の公式アカウントなど様々なユーザが投稿している。観光スポットや商業施設など Point-of-Interest（以下、POI）の公式アカウント（以下、POI 公式アカウント）では、イベントの告知やセール情報などを投稿することがあり、そのような情報を求めているユーザにとって有益な投稿となる。そのため、POI 公式アカウントの投稿を簡易に検索することができれば有用である。例えば、ユーザの現在地付近の POI 公式アカウントの投稿を検索できれば、近くで行われるイベントを見つけることができる。図 1 に横浜赤レンガ倉庫の公式アカウントのツイートの例を示す。このツイートではイベントの告知が行われており、横浜付近でツイートを検索した場合にこのツイートが検索されればイベントを見つけることができる。このような検索を行うためには、様々なユーザが存在する Twitter の中で、1) POI 公式アカウントの特定、2) POI の場所の特定を行う必要がある。

また、POI の公式アカウントと場所を特定することができれば、前述のような POI 公式アカウントの投稿検索だけでなく、一般ユーザがフォローやメンションを行っている POI 公式アカウントから、一般ユーザの趣味嗜好や行動範囲のプロファイリングを行ったり、POI 間の関係分析など様々な応用が考えられる。



図 1 POI 公式アカウントのツイートの例

POI 公式アカウントを特定するためにタウンページ<sup>\*1</sup>や Open Street Map<sup>\*2</sup>などの POI データベースを使う方法が考えられる。しかしながら、POI データベースを利用する方法では、元となる POI データベースにない POI は抽出されないという課題がある。Twitter のプロフィール情報や投稿内容に基づいて POI 公式アカウントを判定することができれば、POI データベースに依存せずにアカウントを特定することができる。

Twitter のプロフィール情報や投稿内容を分析する先行研究として、Twitter ユーザの属性推定の研究がある。従来研究では、一般ユーザを対象として性別、年齢、職業、興味、居住地、支持政党などを推定する研

<sup>1</sup> 株式会社 NTT ドコモ, NTT DOCOMO, INC.

<sup>2</sup> 東京大学, The University of Tokyo

a) ochiaike@nttdocomo.com

<sup>\*1</sup> <http://tpdb.jp/townpage/order>

<sup>\*2</sup> <https://www.openstreetmap.org>

究 [1], [2], [3], [4], [5], [6], [7], [8] が行われていた。しかしながら、企業、観光スポットや商業施設の公式アカウントを対象とした研究は行われていない。

そこで本研究では、Twitter のプロフィール情報から生成した素性に基づき機械学習により POI 公式アカウントを判定する方法を提案する。本研究では POI 公式アカウントを一般消費者が実世界で利用できる場所や施設の公式アカウントと定義する。例えば、観光スポットや商業施設の公式アカウントは POI 公式アカウントとする。複数の実店舗を持つ企業のアカウントも公式アカウントに含まれる。一方、同じ企業アカウントでも、一般消費者向けに実店舗を持たないニュースサイトや情報配信アカウントは POI 公式アカウントに含めない。このようなアカウントを抽出するため、POI に限らず公式アカウント全般の特徴であるフォロワー数が多いという特徴や、POI 独自の特徴として実世界の場所に関する情報や営業時間などを機械学習の素性として利用する。さらに、企業のアカウントではプロフィール画像に企業のロゴを利用していることが多いという仮定のもと、プロフィール画像の画像特徴量も機械学習の素性として利用する。機械学習に利用する素性については、3 章で詳しく説明する。また、Twitter API<sup>\*3</sup> はアクセス回数に制限 (rate limit) があるため、提案手法ではできるだけ API のアクセス回数を少なくするため、投稿内容ではなくプロフィール情報のみから POI 公式アカウントの判定を行う。なお、本稿では 1) POI 公式アカウントの特定、2) POI の場所の特定のうち POI 公式アカウントの特定までを行い、POI の場所の特定については今後の課題とする。

本研究の貢献は以下の通りである。

- POI の特徴を考慮した素性を検討し、Twitter プロフィールのテキストおよび画像から生成した特徴量により POI 公式アカウントを判定する手法を提案した。Twitter API の rate limit を考慮しプロフィール情報のみで判定できる手法を提案した。
- Twitter ユーザを対象として提案手法の定量的な評価を行った。利用する特徴量としてテキストのみを利用した場合、F 値 0.816 の性能で分類できることを確認した。さらに、テキスト特徴量と画像特徴量を組み合わせることで性能を向上できることを確認した。

以降、2 章で Twitter のデータの内容について説明する。次に 3 章で提案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行った評価実験について説明する。5 章で関連研究について述べ、最後に、6 章で本研究のまとめと今後の課題を述べる。



図 2 Twitter プロフィールの構造

表 1 Twitter プロフィールの各項目の説明

| 項目名             | 説明                                                             |
|-----------------|----------------------------------------------------------------|
| profile image   | ユーザが設定したプロフィール画像                                               |
| name            | ユーザの名前で日本語も利用できる項目                                             |
| screen_name     | name と異なり英数字と記号のみのユーザ名。<br>@screen_name とすることで相手に投稿と知らせることができる |
| description     | 160 文字以内で自己紹介文を自由記述できる項目                                       |
| location        | ユーザの居住地や所在地を自由記述できる項目                                          |
| url             | ユーザに関連する Web の URL を自由記述できる項目                                  |
| friends_count   | ユーザがフォローしているユーザ数                                               |
| followers_count | ユーザがフォローされているユーザ数                                              |
| verified        | Twitter 社が該当のアカウントが有名人やブランド本人であることを保証する項目                      |
| listed_count    | ユーザが他のユーザが作成したリストに追加されている数                                     |

## 2. Twitter データ

本章では本研究で利用する Twitter のデータについて述べる。図 2 に Twitter プロフィールの構造を示す。各項目の説明を表 1 に示す。ここでは本研究で利用する項目のみ説明している。listed\_count についてはプロフィールの画面上には明示されないが Twitter API を利用することで取得できる。なお、プロフィール画像に関しては、Twitter API ではプロフィール画像の URL のみが取得できるため、API で URL を取得した後に実際の画像データを URL にアクセスして取得する必要がある。

\*3 <https://dev.twitter.com/overview/api>

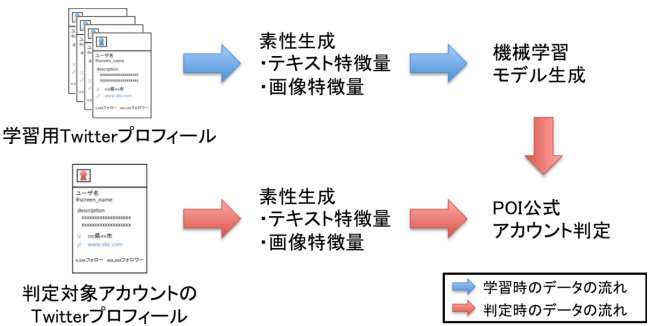


図 3 提案手法の概要

### 3. 提案手法

提案手法の概要を図 3 に示す。提案手法では Twitter のプロフィール情報からテキストとして得られるテキスト特徴量とプロフィール画像を解析して得られる画像特徴量をもとに教師あり機械学習により POI 公式アカウント判定を行う。3.1 節でテキスト特徴量の詳細を説明し、3.2 節で画像特徴量の詳細を述べる。

#### 3.1 テキスト特徴量

本研究で利用するテキスト特徴量の一覧を表 2 に示す。各特徴量について順に説明する。企業のアカウントではフォロワー数が少ない傾向があるためフォロワー数を利用する。有名人や POI 公式アカウントなど情報配信を積極的に行うアカウントではフォロワー数が多い傾向がある。フォロワー数やフォロワー数では絶対値を利用するが、知名度により絶対値は大きく影響を受けると考えられるため相対値としてフォロワー数とフォロワー数の比率を利用する。リストに登録されている数はフォロワー数と同様に有名人や情報配信系のアカウントで数が多いという想定で取り入れている。認証アカウントについては、Twitter 社が認証していれば公式アカウントであるため採用した。location フィールドおよび url フィールドの記載有無については、POI 公式アカウントではユーザーに情報提供することがアカウント開設の目的の一つであるため正確な情報が記載されていることが多いという考えで利用している。description フィールドの特徴量については、それぞれ正規表現により記載の有無を判断する。URL の記載有無については url フィールドと同様である。電話番号とメールアドレスの記載有無については、一般ユーザーや有名人など個人のアカウントではプライバシーの観点であまり記載されない内容と考えられ企業や商業施設独自の内容と考えられる。営業時間表現についても企業や観光スポット、商業施設ならではの記載内容と考えられる。提案手法で用いている特徴量は正規表現で抽出を行える特徴のみを利用しており、言語依存性がないものと考えられる。

表 2 Twitter プロフィールの各項目の説明

| 特徴量                             |
|---------------------------------|
| フォロー数                           |
| フォロワー数                          |
| フォロー数とフォロワー数の比率 (フォロワー数/フォロー数)  |
| リストに登録されている数                    |
| 認証アカウントかどうか                     |
| location フィールドの記載有無             |
| url フィールドの記載有無                  |
| description フィールドへの URL の記載有無   |
| description フィールドへの電話番号の記載有無    |
| description フィールドへのメールアドレスの記載有無 |
| description フィールドへの営業時間表現の記載有無  |

#### 3.2 画像特徴量

本研究では画像特徴量として Twitter のプロフィール画像の Bag-of-Visual Words[9] (以下、BoVW) 表現を利用する。Bag-of-Visual Words は画像認識の分野で特に一般画像認識に利用される画像特徴量 [10] である。本研究では画像上の固定間隔ごとに画像特徴量を計算する dense sampling[11] という手法を利用する。特徴量の記述には、SIFT 特徴量 [12] を利用する。SIFT アルゴリズムは特徴点検出と、検出された特徴点周辺の局所的な特徴を記述する特徴量記述の 2 つから成る。Dense sampling では SIFT アルゴリズムの特徴点検出の部分を固定の間隔の画素を対象として特徴量記述を行う。特徴量記述では、計算対象の画素の周辺との勾配を計算し、勾配方向のヒストグラムを 128 次元のベクトルとして表現する。SIFT 特徴量の詳細は文献 [12] を参照されたい。BoVW では、各画像の特徴量ベクトルを k-means クラスタリングなどでベクトル量子化し visual words と呼ばれる特徴ベクトルを生成する。この特徴ベクトルをまとめたものを codebook と呼び、各画像は codebook に含まれる特徴ベクトルの出現頻度のヒストグラムとして表現される。この表現は元々画像認識の分野で、言語処理でよく用いられる文書を単語の集合として扱う Bag-of-words モデルを画像分類に適用し、各画像を visual words の集合として表現したことから Bag-of-Visual Words と呼ばれる。各画像に対して特徴ベクトルを計算する流れを以下に示す。

- (1) 全画像から SIFT 特徴量を抽出
- (2) SIFT 特徴量を k-means クラスタリングでベクトル量子化し codebook を生成
- (3) Codebook を元に画像ごとに各特徴ベクトルの出現頻度のヒストグラムを作成し画像の特徴量とする

本研究では各ユーザーの画像を BoVW 表現した特徴ベクトルを POI 公式アカウント判定の特徴量として利用する。

#### 3.3 機械学習を利用した POI 公式アカウント判定

POI 公式アカウントの判定は 3.1 節および 3.2 節で説明した特徴量を利用し、機械学習により実施する。本研



究では教師あり機械学習の分類器として Support Vector Machine (以下, SVM) [13] を利用する. 分類器の選定基準は, 従来研究で用いられていること, 一般に広く利用されているライブラリを利用して簡易に実装できるものとした. 本研究の主目的は POI 公式アカウント判定を行うために効果のある特徴量を明らかにすることであり, 分類器そのものの選択は対象外とした. SVM への入力テキスト特徴量と画像特徴量を並列に並べた特徴量を利用する.

## 4. 評価実験

提案手法による POI 公式アカウント判定の性能を評価するため実験を行った.

### 4.1 データ

初めに機械学習の学習データの作成方法について説明する. 正例に関しては Swarm<sup>\*4</sup> というチェックインサービスを経由して投稿されたツイートを利用する. Swarm から投稿されたツイートでは Foursquare 社が POI と Twitter のアカウントと対応付けている場合, 以下のようなツイートになる.

- I'm at 横浜赤レンガ倉庫 イベント広場  
- @yokohamaredbric in 横浜市, 神奈川県  
<https://www.swarmapp.com/xxxx>
- 雨やし地下から到着 (ノ・ヱ・) ノ (@あべのハルカス (ABENO HARUKAS) - @abenoharukas in 大阪市, 大阪府) <https://www.swarmapp.com/xxxx>

そこで, 投稿元が Swarm のツイートから正規表現で @screen\_name を抽出し, Twitter API で対応するユーザーのプロフィール情報を取得する. 日本語の投稿に絞るため, ユーザーの言語設定が日本語に設定されているユーザーの投稿のみを利用した. 2015 年 7 月の 1 カ月分を対象に抽出し, API でプロフィール情報を取得できたものは 4,251 アカウントであった. そこからランダムサンプリングを行い, 海外の POI の Twitter アカウントを目視で除外し, プロフィール画像をダウンロードできたユーザーとして最終的には 1,078 アカウントとなった. 負例に関しては, 2015 年 7 月 1 日~10 日に投稿されたツイートから各日 200 ツイートずつランダムサンプリングを行い, その後, 目視でスパムユーザ, bot, POI 公式アカウントを除外し最終的に 1,260 アカウントを抽出した. プロフィール画像はユーザーがアップロードした画像を Twitter 社が 48 × 48 に加工した normal の画像<sup>\*5</sup> を利用した. 評価では前述のデータを用いて 10 分割交差検定を行った.

<sup>\*4</sup> <https://www.swarmapp.com/>

<sup>\*5</sup> <https://dev.twitter.com/overview/general/user-profile-images-and-banners>

### 4.2 実験環境

SVM の実装には Python 2.7 および scikit-learn 0.16 を利用した. SVM のカーネルは線形カーネルと RBF カーネルを利用した. 画像特徴量を計算するための画像処理ライブラリとして OpenCV 2.4.11 を用いた. 実験には Mac OS X Yosemite, CPU 3GHz Intel Core i7, メモリ 16 GB のマシンを利用した. SVM のパラメータは, グリッドサーチを行い最も性能がよいパラメータを利用した. 線形カーネルは  $C=1$ , RBF カーネルでは  $C=10$ ,  $\gamma=0.001$  とした. BoVW の特徴ベクトルの次元 (codebook サイズ) は k-means クラスタリングのクラスタ数によって決まるため, クラスタ数を  $k=10, 50, 100, 200, 500$  と変えて実験を行い, もっとも性能がよかった  $k=100$  とした.

### 4.3 結果

従来研究でプロフィール情報のみを用いて Twitter ユーザーのユーザ属性推定を行っている研究が存在しないため, テキスト特徴量のみ, 画像特徴量のみ, テキスト特徴量と画像特徴量を組み合わせた場合の性能評価の結果を表 3 に示す.

全体的な傾向としてテキスト特徴量のみを用いる方が, 画像特徴量のみを用いる場合に比べて性能がよい. SVM (RBF) の場合, テキスト特徴量のみを用いた場合の F 値は 0.816 だったものが, テキストと画像の両方の特徴量を利用した場合は 0.831 となった. この差に関して有意差検定を行ったところ有意水準 1% で有意な差があったため, テキスト特徴量と画像特徴量を組み合わせることで性能が向上したと言える. 表 3 において, \* はテキスト特徴量のみを使った場合とテキスト特徴量と画像特徴量で有意差があったことを示す. 表 4 および表 5 に線形カーネルを用いた場合の特徴量の重みを示す. テキストと画像特徴量を組み合わせたときの特徴量の重みは上位 10 件のみを記載する. 画像特徴量は k-means クラスタリングで次元削減しているため何番目のベクトルかを No で示している. テキスト特徴量のみを用いた場合は, フォロワー数をもっとも重みが大いだが, テキスト特徴量と画像特徴量を組み合わせた場合には重みが大くないという結果であった. url フィールド記載有無, description での電話番号と営業時間記載有無, location フィールド記載有無については, テキスト特徴量のみを用いた場合でもテキスト特徴量と画像特徴量を組み合わせた場合のどちらでも重みが大い. フォロワー数は POI 公式アカウントに限らず, 有名人やニュース配信などのアカウントでも値が大いことがあるため, より POI を特定できる特徴量が効果があると考えられる.

## 5. 関連研究

本研究と関連する研究として, Twitter ユーザーの属性推定の研究と Web やソーシャルメディアからの POI 抽出の

表 3 性能評価結果

| 特徴量        | SVM (線形) |       |       | SVM (RBF) |       |        |
|------------|----------|-------|-------|-----------|-------|--------|
|            | 適合率      | 再現率   | F 値   | 適合率       | 再現率   | F 値    |
| テキスト特徴量    | 0.753    | 0.879 | 0.811 | 0.754     | 0.89  | 0.816  |
| 画像特徴量      | 0.69     | 0.668 | 0.678 | 0.696     | 0.69  | 0.692  |
| テキストと画像特徴量 | 0.815    | 0.835 | 0.824 | 0.824     | 0.841 | 0.831* |

表 4 テキスト特徴量のみを用いた場合の特徴量の重み

| 特徴量                       | 重み     |
|---------------------------|--------|
| フォロワー数                    | 0.623  |
| url フィールド記載有無             | 0.479  |
| description での電話番号記載有無    | 0.326  |
| description での営業時間記載有無    | 0.273  |
| location フィールド記載有無        | 0.178  |
| 認証アカウントかどうか               | 0.055  |
| リストに登録されている数              | 0.05   |
| description での url 記載有無   | 0.027  |
| description でのメールアドレス記載有無 | -0.008 |
| フォロー数                     | -0.185 |
| フォロー数とフォロワー数の比率           | -0.196 |

表 5 テキストと画像特徴量を組み合わせたときの特徴量の重み

| 特徴量                    | 重み    |
|------------------------|-------|
| url フィールド記載有無          | 0.444 |
| description での電話番号記載有無 | 0.326 |
| description での営業時間記載有無 | 0.289 |
| location フィールド記載有無     | 0.213 |
| 画像特徴量 No.29            | 0.117 |
| 画像特徴量 No.46            | 0.112 |
| フォロワー数                 | 0.105 |
| 画像特徴量 No.20            | 0.099 |
| 画像特徴量 No.77            | 0.098 |
| 画像特徴量 No.86            | 0.085 |

研究がある。

## 5.1 Twitter ユーザ属性推定の研究

Twitter ユーザの属性推定の研究が数多く取り組まれている。属性推定の研究は (1) 投稿内容に基づく手法, (2) ソーシャルグラフに基づく手法に分けられる。

### 5.1.1 投稿内容に基づくユーザ属性推定

Rao ら [1] は顔文字や略語などの社会言語学に基づく特徴と N-gram を素性として, SVM により年齢, 性別, 地域, 政治的志向を推定する手法を提案した。池田ら [4] は属性ごとの特徴語を赤池情報量基準 (AIC) により選択し, 特徴語を素性として SVM で性別, 年代, 居住地域の推定を行った。平野ら [6] は属性の相互依存関係を考慮するため Markov Logic を用いて複数の属性を同時に考慮しながら集合的に推定する手法を提案した。例えば, ユーザが高校生と推定されたときは, そのユーザの年代が 10 代である可能性が高いなどの依存関係を考慮している。伊藤ら [5] は Twitter と Blog の共通ユーザを用いて Blog に記載さ

れているプロフィールを学習データとして Twitter のプロフィールを推定した。Blog ではプロフィール欄が自由記述ではないことを利用し教師ラベルを自動的に獲得する。榊 [7] らは投稿内容, プロフィール情報, ユーザが含まれるリスト名から SVM を利用して対象ユーザの職業を推定した。その他, 居住地 [3], [14] や発言位置の推定 [15], [16] を行う研究がある。従来研究では, 投稿内容から属性を推定する研究が多く, 本研究のようにプロフィール情報のみで推定を行う研究, テキストと画像を組み合わせた研究はない。

### 5.1.2 ソーシャルグラフに基づくユーザ属性推定

ソーシャルグラフに基づくユーザ属性推定の研究では, つながりのあるユーザは互いに似た属性を持つと仮定してユーザ属性を推定する Pennacchiotti ら [2] は Twitter のプロフィール文書, 返信や投稿方法などのツイートの傾向, ツイートに典型的に現れる単語, 友人の数などソーシャルネットワークの特徴を元に政治的所属, 民族, 特定のビジネスへの興味を推定した。またソーシャルグラフをもとにラベル更新を行う手法を提案した。上里ら [8] は属性推定対象ユーザとフォロワーやメンションの関係がある周辺ユーザに対しても属性推定を予め適用することで推定対象ユーザの推定精度を向上させる手法を提案した。

投稿内容に基づく手法もソーシャルグラフに基づく手法も, 従来は主に一般ユーザを対象としたユーザ属性推定の研究が行われており, 本研究で対象とする企業, 観光スポットや商業施設の公式アカウントを対象とした研究は行われていない。

## 5.2 Web やソーシャルメディアからの POI 抽出の研究

倉島ら [17] は Flickr におけるジオタグ付き写真から Mean-Shift クラスタリングにより人気の観光スポットを抽出した。スポット名称には Flickr で投稿されたタグを利用した。荒川ら [18] は倉島ら [17] と同様に Mean-Shift クラスタリングにより人気の観光スポットを抽出し, Foursquare などのチェックインサービスと統合することにより正確な POI 名称を得られることを明らかにした。Rae ら [19] は Web 検索のスニペットから POI を抽出する研究を行った。この研究では Web 検索のスニペットから CRF を使って POI 名称を抽出し, 場所の特定は Flickr のジオタグ付き F 写真のタグを使い, 1km 四方でタグの確率を計算した。これらの研究では, Web や Flickr のデータから POI を抽出

しており、Twitter から抽出する研究はない。

## 6. おわりに

本稿では Twitter 上でプロフィール情報のテキストおよび画像を用いて POI 公式アカウントを判定する手法を提案した。テキスト特徴量には公式性や POI らしさを判定するための特徴量を利用した。画像特徴量としては POI の公式アカウントでは企業のロゴが用いられることが多く、一般のユーザと分類が行えると考え BoVW を利用した。提案手法に対して定量的な評価実験を行い有効性を確認した。

今後の課題は、テキスト特徴量、画像特徴量の拡充および特徴量の選択や二つの特徴量の組み合わせの拡充、投稿内容を利用した分類との比較が挙げられる。特徴量の拡充に関しては、例えば、テキスト特徴量として description の文体などを利用したり画像特徴量として色情報を利用することが考えられる。

## 参考文献

- [1] Rao, D., Yarowsky, D., Shreevats, A. and Gupta, M.: Classifying Latent User Attributes in Twitter, *Proceedings of the 2Nd International Workshop on Search and Mining User-generated Contents*, SMUC '10, pp. 37–44 (2010).
- [2] Pennacchiotti, M. and Popescu, A.-M.: Democrats, Republicans and Starbucks Afficionados: User Classification in Twitter, *KDD '11*, pp. 430–438 (2011).
- [3] Hecht, B., Hong, L., Suh, B. and Chi, E. H.: Tweets from Justin Bieber's Heart: The Dynamics of the Location Field in User Profiles, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 237–246 (2011).
- [4] 池田和史, 服部 元, 松本一則, 小野智弘, 東野輝夫: マーケット分析のための Twitter 投稿者プロフィール推定手法, *情報処理学会論文誌*, Vol. 2, No. 1, pp. 82–93 (2012).
- [5] 伊藤 淳, 西田京介, 星出高秀, 戸田浩之, 内山 匡: Twitter と Blog の共通ユーザおよび会話ユーザの同類性に着目した Twitter ユーザ属性推定, *日本データベース学会論文誌*, Vol. 12, No. 1, pp. 31–36 (2013).
- [6] 平野 徹, 牧野俊朗, 松尾義博: Markov Logic を用いたテキストからのユーザ属性推定, Vol. 27, *人工知能学会*, pp. 1–4 (2013).
- [7] 柿 剛史, 松尾 豊: ソーシャルメディアユーザの職業推定手法の提案, *知能と情報*, Vol. 26, No. 4, pp. 773–780 (2014).
- [8] 上里和也, 浅井洋樹, 奥野峻弥, 山名早人: Twitter ユーザを対象とした属性推定の精度向上—周辺ユーザの属性補完を利用して—, 第 7 回データ工学と情報マネジメントに関するフォーラム (DEIM 2015), pp. D8–5 (2015).
- [9] Csurka, G., Dance, C., Fan, L., Willamowski, J. and Bray, C.: Visual categorization with bags of keypoints, *Workshop on statistical learning in computer vision, ECCV*, Vol. 1, No. 1-22, Prague, pp. 1–2 (2004).
- [10] 八木康史, 斎藤英雄 (編): CVIM チュートリアルシリーズ コンピュータビジョン最先端ガイド 3, アドコム・メディア株式会社 (2010).
- [11] Jurie, F. and Triggs, B.: Creating efficient codebooks for visual recognition, *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, Vol. 1, IEEE, pp. 604–610 (2005).
- [12] Lowe, D. G.: Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, Vol. 60, No. 2, pp. 91–110 (2004).
- [13] Cortes, C. and Vapnik, V.: Support-Vector Networks, *Machine Learning*, Vol. 20, No. 3, pp. 273–297 (1995).
- [14] Cheng, Z., Caverlee, J. and Lee, K.: You Are Where You Tweet: A Content-based Approach to Geo-locating Twitter Users, *ACM CIKM '10*, pp. 759–768 (2010).
- [15] 山口祐人, 伊川洋平, 天笠俊之, 博之北川: ソーシャルメディアにおけるローカルイベントを用いたユーザ位置推定手法, *情報処理学会論文誌データベース (TOD)*, Vol. 6, No. 5, pp. 23–37 (2013).
- [16] 伊川洋平, 榎 美紀, 立堀道昭: マイクロブログのメッセージを用いた発信場所推定, 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM 2012), pp. F7–2 (2012).
- [17] Kurashima, T., Iwata, T., Irie, G. and Fujimura, K.: Travel Route Recommendation Using Geotags in Photo Sharing Sites, *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pp. 579–588 (2010).
- [18] 荒川 豊, タチアーナシェフラー, ステファンバウマン, アンドレアスデンゲル: ソーシャル観光マッパー—ソーシャルデータからの観光スポット抽出, *情報処理学会論文誌コンシューマ・デバイス&システム (CDS)*, Vol. 4, No. 1, pp. 1–11 (2014).
- [19] Rae, A., Murdock, V., Popescu, A. and Bouchard, H.: Mining the Web for Points of Interest, *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pp. 711–720 (2012).