

イベント推薦のためのユーザの行動範囲の自動推定

早坂 絵央^b 木村 優平^b 戎 淳^b 片倉 多智^b 吉原 圭祐^b 西村 百之輔^b 延澤 志保^{*}
^b東京都市大学知識工学部 ^{*}東京都市大学情報工学部

1 研究背景

イベント情報には、さまざまなイベントのジャンルのほか、名称、開催日、開催場所、期間などの基本情報が存在する。イベント情報を対象とした研究としては、福馬らの提案した地域イベントの分類を目的とした教師ありトピックモデル推定 [1] や、山田らの提案したマイクロブログからの開催中イベント情報の機械学習を用いた抽出 [2] などが挙げられる。

Twitter¹などの SNS を対象としたユーザの行動や居住地域などの推定では、ブログ中の地名の出現頻度とその文脈に着目して地名の出現箇所が居住地域かどうか判定する研究 [3] のほか、定型文を用いて実際のユーザの位置可否かを機械学習で推定する伊川らの研究などがある [4]。Twitter を対象とした行動推定では、杉谷らの手法 [5] のようにユーザの位置情報 (ジオタグ) を推定に利用する研究が多い。しかし、ジオタグを用いるユーザは少なく、Twitter では 2019 年に正確な位置情報のタグ付けを廃止した²ため、今後の研究ではジオタグの利用は難しい。

2 ユーザの行動範囲推定

2.1 研究目的

ユーザの行動範囲や位置の推定はテキストベースでは困難とされている。本研究では、ジオタグに頼らない自動推定の実現のため、SNS 等の投稿に含まれる地名に着目したユーザの行動範囲の推定を目的とする。

2.2 ユーザの行動範囲

本研究では、SNS 等のユーザの投稿での行動パターンを、通勤通学などの日常的行動範囲と、旅行などで訪れる一時的行動範囲の 2 種類に分けて検討する。本研究では一時的行動範囲もイベント推定の候補とするが、推薦優先度は日常的行動範囲に比べて低くする。

行動範囲の記述の種類は、過去に居た、現在居る、将来居る予定である、居るかどうかわからないの 4 種類に分ける。このうち、本研究では一時的可否かの推定のため「現在居る」状況のみ「居る」と認めることとし、それ以外は「居ると認められない」とみなす。

本研究では、ユーザが実際に居た場所を判定し行動範囲を推定するための学習データとして、地名を含むツイートを利用する。表 1 では対象地名の場所に居ると認められるツイートを存在ラベル○、居ると認められないツイートを存在ラベル×で示している。

表 1: ツイート中のユーザ位置情報の分類例

ラベル	ツイート	状況
○	久しぶりに東京!	現在居る
○	私東京にいるんだけど。。	現在居る
○	これから東京へ帰ります。#岡山空港	現在居る
×	昨日東京行った!	過去居た
×	これから東京へ帰ります。#岡山空港	将来居る予定
×	東京は変わった!	居るかどうかわからない

ある町田市在住の Twitter ユーザの 2011 年から 2015 年のツイート中の地名に対して存在ラベルを付与したところ、居住地である「東京」の月ごとの出現頻度が他の地名に比べ多く、ツイートとユーザの行動範囲に密接な関係があることを確認した (図 1)。各月について図 1 の

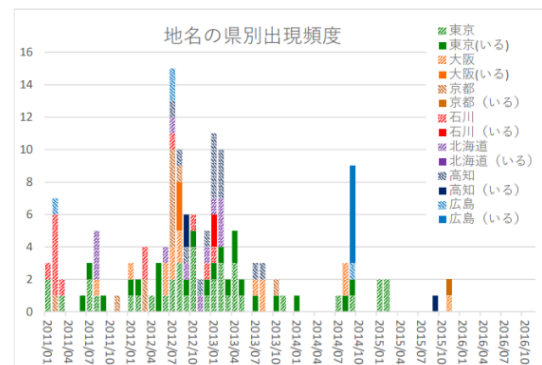


図 1: 地名の出現頻度と存在ラベルの付与結果

実線が存在ラベル○と判定したツイートの数、斜線が×と判定したツイートの数である。

3 行動範囲自動推定手法

3.1 処理の流れ

本研究では、地名を含むツイートに対しユーザが実際にその地名に居るか否かの判定を行うため、その周辺語句の特徴を学習させた二値分類器による推定を行う。

システムの全体の流れを図 2 に示す。学習フェーズでは、ユーザの投稿から地名を含むツイートを抽出し、存在ラベルと周辺文脈を SVM [6] の学習データとして二値分類器の作成を行う。

行動範囲推定フェーズでは、ユーザごとに地名を含むツイートを抽出し、これに対して SVM で分類を行って行動範囲となる地名の推定を行う。その上で、行動範囲地名とツイート日時をグループ化して、その出現頻度か

Automatic Estimation of User Activity Area for Event Recommendation.

Kaio Hayasaka (Caio Hayassaka)^b, Yuhei Kimura^b, Bo Rong (戎 ボツ)^b, Taichi Katakura^b, Keisuke Yoshihara^b, Momonosuke Nishimura^b, and Shiho Hoshi Nobesawa^{*}.

^b Faculty of Knowledge Engineering, Tokyo City University

^{*} Faculty of Information Technology, Tokyo City University

¹ Twitter, <https://twitter.com/>.

² Twitter 正確な位置情報タグ付けの廃止, <https://twitter.com/TwitterSupport/status/1141039841993355264>.

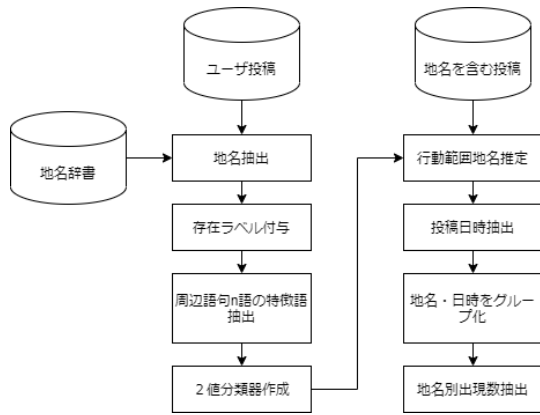


図 2: 行動範囲推定の処理の流れ

ら日常的行動範囲と一時的行動範囲の判定を行う。

3.2 学習フェーズ

形態素解析器 MeCab¹を用いてツイートの地名抽出を行う。ツイートに含まれる地名は店の名前やランドマークを含むなど多種多様であり、行動範囲を推定する上で扱い辛い。地名辞書²を用いて都道府県もしくは市町村レベルの地名に置き換える。

本研究では、存在の有無を自動推定する際、ツイートの文脈を用いる。本研究では、文脈を示す語として、対象地名から前後それぞれ最大 n 個の形態素を学習に用いる (図 3, $n = 5$)。文脈はツイート単位とし、前後の形態

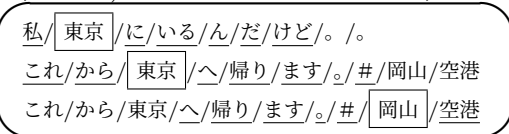


図 3: 地名の周辺語句

素どちらかが n 個に満たない場合でも前後のツイートに跨ることはしない。図 3 のように 1 ツイート中に複数の地名が出現する場合であっても、各地名の前後 n 形態素とすることで区別して学習することが可能である。

3.3 行動範囲推定フェーズ

ユーザごとに地名を含むツイートを抽出し、これに対して SVM を用いて居ると認められる地名の抽出を行う。

行動範囲として抽出された地名のうち出現する月の数が最も多いものを、このユーザの日常的行動範囲と推定する。それ以外の地名は一時的行動範囲として扱う。

4 実験結果と考察

2.2 節で行動範囲推定のための予備実験を行ったユーザのツイート 8,146 件のうち、地名を含むツイート 1,019 件に対して実験を行った結果を表 2 に示す。ここでは文脈として学習する形態素数を前後それぞれ 5 個とした。負例の正解率は 94.6% と十分高いが、正例の正解率は 32.0% に留まる (表 2)。適合率は 72% を得たが、再現率は 32% と低く、F 値は 0.44 となった。本研究の目的はイベント推薦であり、再現率の低さは推薦対象地名の減少に繋が

表 2: 行動範囲推定結果

	正	負	計
正例	98(32.0%)	208(68.0%)	306
負例	38(5.3%)	675(94.6%)	713
計	136(13.3%)	883(86.7%)	1,019

るため、精度向上のため検討が必要である。存在有無の判定では「へ帰る」など動詞句部分が重要な意味を持つが、ツイートでは動詞句部分の省略が多く、また表記揺れの問題などからも、再現率が低く留まった理由は学習データ数の影響が大きいものと考えられる。

学習で抽出できた行動範囲地名の出現頻度抽出結果を図 4 に示す。実験に用いたユーザの居住地である東京が

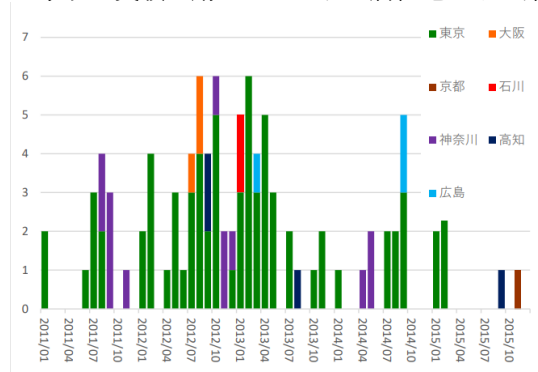


図 4: 地名別出現頻度抽出結果

全体の 79.6% と最も多く、他の地名との出現頻度の差からこのユーザの日常的行動範囲は東京と推定することに成功した (図 4)。

5 まとめ

本論文では、イベント推薦の補助として、SNS での投稿からユーザの行動を読み行動範囲を推定する手法を提案した。本研究では、投稿内に含まれる複数の地名からユーザの実際の位置を示す地名を推定することを行い正解率 75% を得た。また、その出現頻度から日常的か一時的かの行動範囲の推定に成功した。

参考文献

- [1] 福馬智生, 鳥海不二夫, “地域型イベントの収集とその分類手法の検証,” ARG 第 10 回 Web インテリジェンスとインタラクション研究会, pp.25-30, 2017.
- [2] 山田渉, 菊地悠, 落合圭一, 鳥居大祐, 稲村浩, 太田賢, “マイクロブログを用いたイベント情報抽出技術,” 情報処理学会論文誌, vol.57, no.1, pp.123-132, 2016.
- [3] 安田宜仁, 平尾努, 鈴木潤, 磯崎秀樹, “ブログ著者の居住域の推定,” 自然言語処理学会第 12 回年次大会, no.C3-5, pp.512-515, 2006.
- [4] 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 10 回日本データベース学会年次大会 (DEIM Forum 2012), no.F7-2, pp.1-4, 2012.
- [5] 杉谷卓哉, 白川真澄, 原隆浩, 西尾章治郎, “教師あり機械学習を用いたツイート投稿時のユーザ位置推定手法,” 情報研報, vol.2013-DBS-158, no.26, pp.1-8, 2013.
- [6] 栗田多喜夫, “サポートベクターマシン入門,” 産業技術総合研究所 脳神経情報研究部門, pp.1-21, 2002.

¹形態素解析器 MeCab, <https://taku910.github.io/mecab/>.
²国内観光情報サイト, <https://gojapan.jp/>.