

京都の街歩きイベントに基づいた他地域の様相記録自動分類手法： “100ninmap”プロジェクトの応用

西澤 元希[†][†]和歌山大学北 雄介[‡][‡]京都大学荒牧 英治^{††}^{††}奈良先端科学技術大学院大学宮部 真衣[†]

1 はじめに

我々は普段生活する都市の中で、街の雰囲気や環境音、できごとなど様々な情報を取得している。このような都市において把握することがらに関する情報(本研究では「様相」と呼ぶ)は、まちづくりや観光、住まい探しなどの場面で価値のあるものである。その中でも、場所と密接に関連したものは特に有用である可能性がある。例えば従来のナビゲーションシステムに代わり提案されているランドマークによるナビゲーションシステム [1] は、人が記憶しやすい地物を経路案内に利用しており、都市の様相を活用した事例の一つと言える。

都市の様相を抽出する方法としては被験者に実際に街を歩いてもらい、把握した様相について自由に記述してもらう方法が提案されている [2]。様相を把握する本人によってなされた記録を本研究では様相記録と呼び、特に場所と密接に関連したものを場所依存記録と呼ぶ。この手法では様相記録を確実に収集できるが、人手による作業が多く、実施における負担が大きい。

近年、マイクロブログの普及に伴い、個人での情報発信が活発に行われている。都市の様相記録は、マイクロブログで発信されるデータの中にも多く含まれていると考えられる。マイクロブログデータの中から様相記録を抽出できれば、人手による負担を抑え、より多くの様相記録の収集ができる可能性がある。

これまでに我々は、街歩きイベント「100人で作る京都地図」を実施し、そこで収集した様相記録をもとにマイクロブログから場所依存記録を抽出する分類器を構築した [3]。しかし、対象地域を京都に限定しており、他地域への適用については検討していない。そこで本稿では、適用地域を限定しない場所依存記録の抽出手法を検討する。

2 様相記録の自動分類

2.1 街歩きイベントに基づいた様相記録の自動分類

我々は様相記録の収集を目指したプロジェクトとして“100ninmap”を立ち上げ、2013年7月27日(土)に、

Automatic Modal Records Classification Method for Areas beyond Kyoto: Based on the Walking Event, “100ninmap” Project
Motoki Nishizawa[†] Yusuke Kita[‡] Eiji Aramaki^{††} Mai Miyabe[†]
[†]Wakayama University
[‡]Kyoto University
^{††}Nara Institute of Science and Technology

街歩きイベント「100人で作る京都地図 第1回三条・四条編」を実施した。本イベントでは、我々の構築したつぶやきシステム“100ninmappin”を使用し、街を歩き、感じたことを参加者にスマートフォンで発信してもらうことで様相記録を収集した [3]。

我々は収集した様相記録のうち、以下のいずれかの表現を含み、かつその場にいる(回想などではない)と判断されるものを場所依存記録と分類することとし、場所依存記録分類器のコーパスを作成した。

- 場所を指す言葉(地名、店名など)や商品名
- その場所で体験できる行為(「足を浸せる」など)
- その場所で行われている行為(「工事している」など)
- その場所で感じられる雰囲気(「にぎやか」など)

コーパスをもとに分類器を構築し、Twitterを対象とした分類実験を行った [3]。京都府で発信されたツイートに適用した結果、一定精度で場所依存記録を抽出できることを示した。

2.2 他地域への適用方法

京都府で発信されたツイートの抽出失敗例として、分類器の構築に用いたコーパスに存在しない地名や行動表現がツイート内に含まれているものがある。京都府以外の地域で発信されたツイートには、分類器のコーパスに存在しない表現(特に場所を示す表現)が多く含まれていると考えられる。もし、コーパスに含まれる地名が判定精度に影響を与えているとすると、京都府で収集した様相記録を元に構築した分類器では他地域の場所依存記録を適切に分類できない可能性がある。

そこで、他地域への適用方法を検討する。今回はテキストに含まれる、場所を示す表現に着目し、それらを特定の表現に置き換えることとした。具体的には、以下に示すように形態素の細分類が「地名」となっている形態素を[地名]へと置き換えるようにした。なお形態素解析にはJUMAN[4]を用いた。

変換前：こんな時間なのに人がいっぱいだ! (@ 横浜駅 (Yokohama Sta.) w/ 18 others) (URL)

変換後：こんな時間なのに人がいっぱいだ! (@ [地名] 駅 (Yokohama Sta.) w/ 18 others) (URL)

表 1: 京都府ツイートへの適用結果

	適合率	再現率	F 値
従来手法	0.567	0.538	0.552
提案手法	0.583	0.495	0.535

表 2: 京都府外ツイートへの適用結果

	適合率	再現率	F 値
従来手法	0.278	0.700	0.398
提案手法	0.231	0.567	0.328

表 3: 京都府外ツイートの分類失敗例

ツイート	変換後のツイート	従来手法による分類の成否	提案手法による分類の成否
この時間の名古屋は歩くには気持ちいい 池袋セントラルホテルのテレビがデカくなってるのだ～、見ないけど。	この時間の [地名] は歩くには気持ちいい [地名] セントラルホテルのテレビがデカくなってるのだ～、見ないけど。	×	×
二回目。落ち着いたカウンターで超デカイジョッキ。(@ 銀座酒場) <URL>	二回目。落ち着いたカウンターで超デカイジョッキ。(@[地名] 酒場)<URL>	○	×
終電です。(@ JR 大阪駅 (Osaka Sta.) w/ @****) <URL>	終電です。(@JR[地名] 駅 (OsakaSta.)w/@****)<URL>	○	×

3 実験

場所を示す表現を置き換えることで、京都府以外の地域で発信されたツイートに分類器が適用できるか検証する。検証には、街歩きで得られたつぶやき (以下、「街歩きデータ」)、京都府で発信されたツイート (以下、「京都府ツイート」)、京都府以外の地域で発信されたツイート (以下、「京都府外ツイート」) を用いる。

実験は「従来手法 (変換なし)」および「提案手法 (地名の変換あり)」の2つの手法で行い、各手法ごとに街歩きツイート 2553 件 (その内正例は 1428 件) から場所依存記録分類器を構築し、京都府ツイートと京都府外ツイートに適用する。

検証用のツイートとしては、京都府ツイート 1000 件、京都府外ツイート 1000 件を用いた。それぞれ位置情報付きツイート¹ から無作為に 1000 件抽出し、2.1 節で示した基準に沿って場所依存記録かどうかを人手 (1 名) で分類した。京都府ツイートの正例数は 604 件、京都府外ツイートの正例数は 217 件である。

4 実験結果と考察

4.1 各手法における分類精度の比較

京都府ツイートの分類精度を表 1 に示す。F 値は従来手法で 0.552、提案手法で 0.535 となり、大きな違いはみられなかった。また、京都府外ツイートの分類精度を表 2 に示す。F 値は従来手法で 0.398、提案手法で 0.328 となり、大きな違いはみられなかった。

京都府外ツイートの判定失敗例を表 3 に示す。表 3 より、地名の置き換えが行われているにもかかわらず、検出できていない事例がみられた。また、置き換えが行われることで、正しく判定できていたものが検出できなくなる事例もみられた。今回、地名による影響があると仮定し、地名の置き換えによる手法を試みたが、JUMAN の形態素の細分類を利用して取得できる地名

に関しては、場所依存記録の判定に与える影響が少なく、地名置き換えの効果が低い可能性がある。

4.2 各ツイートにおける分類精度の比較

表 1、表 2 を比較すると従来手法、提案手法のいずれも、京都府外ツイートの判定精度が低かった。京都府ツイートと京都府外ツイートの違いを確認した結果、京都府外ツイートには京都府ツイートに比べ、リプライと他者のツイートの引用が多くみられた。これらのツイートは誤検出の事例にも多くみられ、場所依存記録の判定に影響を与えている可能性が考えられる。

5 おわりに

本研究では、街歩きイベント「100 人でつくる京都地図」で得られた様相記録をもとに構築した分類器を用いて、適用地域を限定せずにマイクロブログから場所依存記録を抽出する手法を検討した。街歩きテキストに含まれる場所を示す表現を特定の表現に置き換え、分類器を適用した結果、以下のことが分かった。

- (1) JUMAN の形態素の細分類を用いた地名の置き換えでは、判定精度の改善効果が低い可能性がある。
- (2) リプライと他者のツイートの引用が場所依存記録の判定に影響を与えている可能性がある。

今後はより詳細に場所を表す表現やその他の要因に着目し、分類精度の向上を目指す。

参考文献

- [1] 米倉梨菜, 森永寛紀, 若宮翔子ほか: 点と線と面のランドマークによる道路地図に頼らないナビゲーション・システム, インタラクシオン 2015 論文集, pp1-10(2015).
- [2] 北雄介: 経路歩行実験に基づく都市の様相の分析とモデル化に関する研究, 京大大学学位論文 (2012).
- [3] 宮部真衣, 北雄介, 久保圭, 荒牧英治: マイクロブログから場所依存の様相記録を抽出する: “100inmap” プロジェクトによる街歩きイベントの応用と実施, 言語処理学会 第 20 回年次大会 発表論文集, pp.420-423(2014).
- [4] Daisuke, K. and Kurohashi, S.: A Fully-Lexicalized Probabilistic Model for Japanese Syntactic and Case Structure Analysis, In Proceedings of HLT-NAACL2006, pp.176-183 (2006).

¹2011 年 7 月 15 日から約 1 年間収集されたものである。