

# マイクロブログから場所依存の様相記録を抽出する： “100ninmap”プロジェクトによる街歩きイベントの実施と応用

宮部 真衣 北 雄介 久保 圭 荒牧 英治

京都大学 学際融合教育研究推進センター

mai.miyabe@gmail.com

## 1 はじめに

都市に関する情報は、様々な場面で重要となる。例えば、旅行の場合、観光地や飲食店など、個々人の嗜好に合った情報は重要な役割を果たす。引っ越しなどで新しい住居を探す際は、予算・間取りといった情報だけでなく、「静か」「涼しい」といった感覚的な情報も役立つ。また、街並み景観の問題や、被災地復興に際して以前の「雰囲気」をいかに取り戻すかといった難しい問題においても、都市の全体性に関する考察が不可欠である。しかし、このような情報の提供には、現地での多角的な情報収集が必要であり、未だ十分に顕在化されていないことも多い。

本研究では、これらの漠然とした感覚を様相<sup>1</sup>と呼ぶ。様相とは未分化の全体性であり、また印象、感覚、記憶、空想といった諸概念を広く包含するため、そこから情報を取り出すことが難しい。都市の様相の断片を抽出する方法としては、人間が街を歩き、そこで把握した様相を発話や用紙への書き込みによって自由に表現するという方法がある [1, 2, 3]。様相を把握する当人によってなされた記録を、本稿では様相記録と呼ぶ。この手法は、「様相を記録する」ことを前提として街を歩き、記録を行うため、様相記録だけを確実に収集可能だが、多大な労力のかかる作業である。

マイクロブログなどの普及に伴い、日々蓄積される巨大なデータの中には、意図せず人々が都市の様相について記述したものが含まれると考えられる。大量のマイクロブログデータの中から、このような偶発的な様相記録を抽出することができれば、わざわざ様相の記録をしてもらわずとも、低コストで様相記録の収集が可能となる。前述したように、これまでに街歩きによる様相の記録は行われているが [3]、紙媒体での記録という形を取っており、大規模な収集は容易でない。

<sup>1</sup> 建築・都市領域では、建物や樹木のような個別の要素を越えた、我々の把握していることがらの全体性を指す概念として「様相」の語が用いられている。

また、情報端末を用いた街歩きイベントも行われているが [4, 5]、情報提供を目的として情報端末を用いており、様相記録の電子的な収集を試みた事例はない。

そこで本研究では、「新しい地図をつくる」という趣旨の、スマートフォンを用いた街歩きイベントを開催し、その参加者が記述したデータを様相記録として収集する。市民イベントの形態をとることで、多くの人員を動員できる。さらに、収集した様相記録を教師データとし、地図と併せた利用において特に有用となる、場所と密接に関連した様相記録（場所依存記録）を、マイクロブログから抽出するシステムを構築する。

## 2 街歩きイベント“100人でつくる 京都地図”による様相記録の収集

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、街歩きイベント「100人でつくる京都地図 第1回三条・四条編」<sup>2</sup>を実施した。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い、街の感じ方を地図上に表すことをイベントの目的とし、参加者を募った。参加者は学生・社会人を中心とした54名である。街歩きの際は、参加者を5~6名からなる10チームに分けた。

本イベントでは、我々の構築したつぶやきシステム“100ninmappin”を使用した。図1に、100ninmappinの画面例を示す。本稿では、街歩きで参加者が発信する自由記述データを「つぶやき」と呼ぶ。参加者らが、感じたことをつぶやき入力エリアに入力して発信すると、つぶやきが位置情報と併せてサーバに送信される。イベントの流れを以下に示す。

1. イベント概要, 100ninmappin の説明
2. 街歩き (14:00~17:00, 計3時間)
3. 地図の閲覧, 総括

<sup>2</sup><http://www.100ninmap.com/>

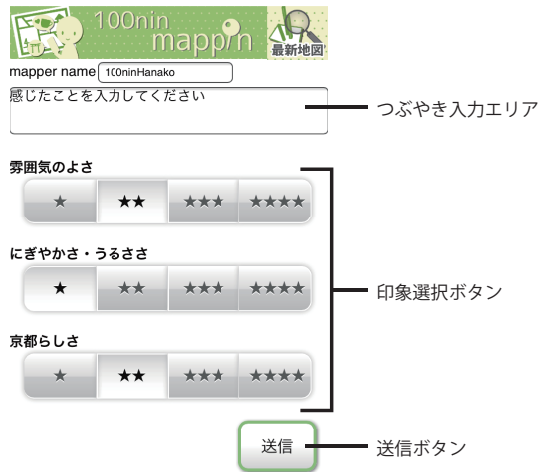


図 1: つぶやきシステム“100ninmappin”の画面例

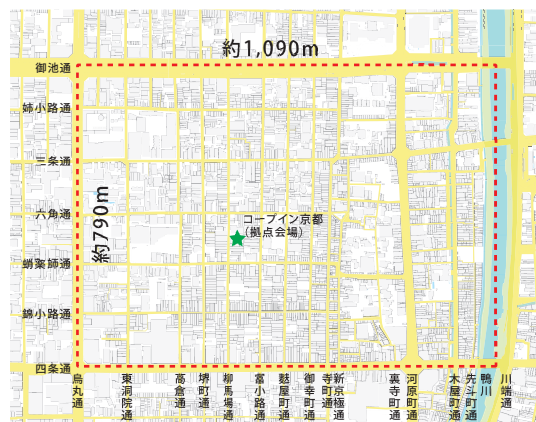


図 2: 街歩きの範囲

今回街歩きの範囲としたのは、図 2 における点線の範囲内<sup>3</sup>である。まず拠点会場（図 2 の 印の地点）でイベントの概要と 100ninmappin の使用方法を説明した後、参加者らに街歩きを行ってもらった。街歩きの様子を図 3 に示す。街歩きでは、指定した範囲内であれば自由に歩くことができ、買い物や飲食店での休憩なども自由に行ってよいこととした。3 時間の街歩きの後、拠点会場に集合し、収集した様相記録をマッピングした地図を参加者全員で閲覧した。

### 3 場所依存記録分類器

本章では、まず場所依存記録の定義について述べた後、街歩きイベントにより収集した様相記録および構築した場所依存記録分類器について述べる。

#### 3.1 様相記録と場所依存記録

都市の様相記録とは、ある場所にいる当人が、その場で感じたことを記述したものであり、その内容には

<sup>3</sup>この範囲は四条通や河原町通、寺町通、新京極通を中心にさまざまな店舗が立地しており、京都の中心部と呼ぶにふさわしい賑わいがある場所である。



図 3: 街歩きの様子

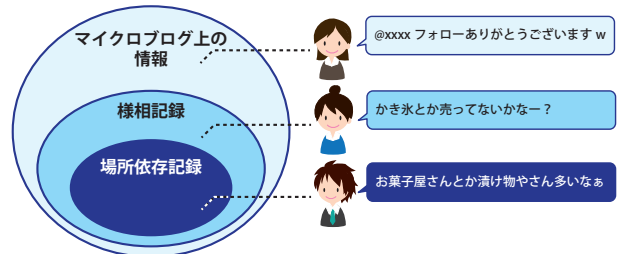


図 4: マイクロブログ上の情報と様相記録との関係

場所と密接に関連したもの（以降、場所依存記録と呼ぶ）だけでなく、そうでないものも含まれる。図 4 に、マイクロブログ上で発信される情報と、様相記録・場所依存記録との関係を示す。場所依存記録は様相記録の一部という位置付けである。マイクロブログ上の情報には、都市の様相とは無関係の情報（例えば「フォローありがとうございます」のような、都市と無関係な対話）も含まれ、部分的に様相記録、さらには場所依存記録が含まれていると考えられる。なお、街歩きイベントにおいて発信される情報は、すべて様相記録とみなしている。

地図などと組み合わせて利用するという観点では、場所依存記録は、様相記録の中でも特に有用となる可能性がある。そこで、本研究では収集した様相記録を用いて場所依存記録分類器を構築する。

本研究では、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する。

- 場所を指す言葉（地名、店名など）や商品名<sup>4</sup>
- その場所で体験できる行為<sup>5</sup>（「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<sup>6</sup>（「にぎやか」「うるさい」など）

#### 3.2 コーパス

2 章で述べた街歩きイベント（54 名、3 時間の街歩き）により、2609 件のつぶやき（様相記録）が収集で

<sup>4</sup>なお、店名、商品名などは、必ずしも具体的でなくともよいこととした。例えば、正式な店名でなく「キム子屋」のような抽象的な表現であっても、場所を指す言葉とみなした。また、全体的・複合的な場所表現（「お寺が多い」など）でもよいこととした。

<sup>5</sup>におい、香りなどの体感も体験できる行為とみなす。

<sup>6</sup>「車にひかれそう」といった印象も雰囲気とみなす。

表 1: 場所依存記録コーパス

	正例 (+1) / 負例 (-1)	つぶやき内容
(1)	+1	Member A とビール！新卒極のスタンド。いい!! ビールがうまい！
(2)	+1	TINTIN shop。行ってみる。
(3)	+1	スタンド！昭和なかんじ
(4)	-1	3時間歩くのはふだんあんまりないなと、実感中
(5)	-1	しまった！イノダコーヒーで休憩すれば良かった。
(6)	-1	ビールいいですねー ( - )

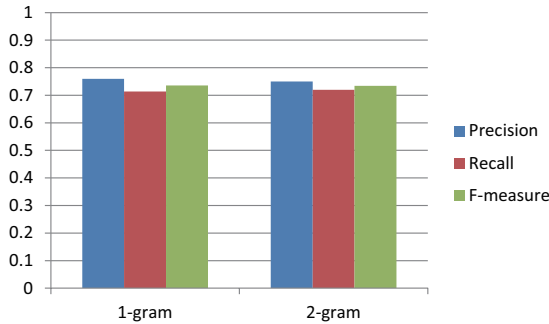


図 5: 10 分割交差検定結果

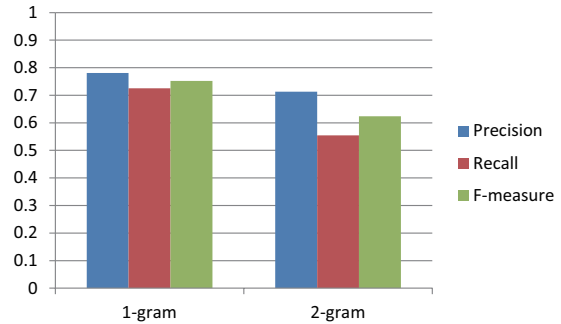


図 6: Twitter データへの適用結果

きた。これらのつぶやきに対し、3.1 節に示した基準に基づいて、場所依存記録かどうかを手で分類し、場所依存記録分類器のコーパスとした。コーパスの一部を表 1 に示す。つぶやきが場所依存記録である場合は正例、そうでない場合は負例とした。コーパスにおける正例の数は 2609 件中 1428 件 (55%) である。

表 1(1) ~ (3) のように、発信者の所在地と関連するつぶやきは正例 (場所依存記録) と分類し、表 1(4) ~ (6) のような現在地に関わらない感想は負例とした。

### 3.3 場所依存記録分類器

3.2 節で述べたコーパスを用いて、場所依存記録分類器を構築した。今回は、素性としてつぶやきに含まれる形態素 (1-gram, 2-gram) を用いた。SVM<sup>7</sup>による学習には、多項カーネル (d=2) を使い、パラメータはデフォルト値を用いた。

## 4 実験

構築した場所依存記録分類器の精度を検証するため、3.2 節で述べたデータを用いて 2 種類の実験を行う。それぞれの実験において、以下の内容を検証する。

- 実験 1: 場所依存記録分類器によって、街歩きで収集した様相記録から場所依存記録を判定できるか?
- 実験 2: 場所依存記録分類器は、一般のマイクロブログデータに含まれる場所依存記録を判定できるか?

1 章で述べたように、マイクロブログデータには、発信者が意図せず様相を記述し、発信したものが含ま

<sup>7</sup>TinySVM を利用した。

<http://chasen.org/taku/software/TinySVM/>

れるが、そうでないものも多い。一方、街歩きでは様相記録のみを確実に収集することができる。そこで、マイクロブログデータに存在する、偶発的な場所依存記録に対して、街歩きイベントでの様相記録を教師データとした分類器での判定が可能かどうかを検証する (実験 2)。実験 2 では、Twitter のデータをテストデータとする。位置情報付のツイート<sup>8</sup>のうち、京都府で発信されたツイートを無作為に 1000 件抽出したものを適用対象のデータとした。また、抽出した 1000 件のツイートについても、3.1 節の基準に基づき、場所依存記録かどうかを手で分類した。

## 5 実験結果と考察

### 5.1 分類器による判定精度

場所依存記録分類器によって、街歩きで収集した様相記録から場所依存記録を判定できるかどうか (実験 1) を検証するために、10 分割交差検定を行った。結果を図 5 に示す。F 値は、1-gram では 0.736、2-gram では 0.735 となり、概ね良好な結果が得られた。

次に、場所依存記録分類器は、マイクロブログデータにも適用できるかどうか (実験 2) を検証するために、Twitter のデータをテストデータとした場合の判定精度を確認した。結果を図 6 に示す。1-gram での F 値は 0.751、2-gram での F 値は 0.624 であった。1-gram では、実験 1 の結果と同等の精度で判定可能であり、Twitter などのマイクロブログデータに含まれる偶発的な場所依存記録についても、ある程度高精度に抽出可能であることが示唆される。

<sup>8</sup>2011 年 7 月 15 日から約 1 年間収集されたものである。

表 2: 判定失敗例

	理由	つぶやき内容
未検出	コーパス内には存在しない地名などが含まれる	(1) 法務局に来ていますが、いつもより静かなように思います。 (2) 広くて会場に辿り着けない... (@ 建仁寺) < 画像の URL > (3) 京都御所の公開、今日 4 月 8 日まで。開門前でもう長蛇の列。荷物検査は「皇宮警察」のお姉さん ( ´ ` ) でした。 < 画像の URL >
誤検出	他者のツイート内容の引用を含む	(4) うひょ~”@****: 赤レンガなう。メキシコフェスティバルのイメージガール審査員です。 (5) まあまあ近い! 車なら 10 分かかるへん“ @****: 近いんですか? RT @****: その後は、goya で晩御飯!“ @****: リベンジ in 京都市動物園 ( ´ ` ) ””
	つぶやきのみでは場所依存であると断定できない	(6) こんなにも売ってます。 @**** < 画像の URL >

## 5.2 Twitter データにおける判定失敗例

5.1 節では、Twitter のようなマイクロブログデータからでも、高精度に場所依存記録を抽出できることを示した。では、判定に失敗しているツイートはどのようなものであろうか？特徴的な失敗例を表 2 に示す。

未検出（正例を負例と判定）は、主としてコーパス内には存在しない地名や行動表現などが含まれることに起因する。例えば、表 2 における各事例には、現在地を示す表現（「法務局」「建仁寺」など）や、状況・雰囲気を示す表現（「静か」「長蛇の列」など）があるが、構築した分類器のコーパスには表現が存在せず、正しく判定できなかった可能性がある。

誤検出（負例を正例と判定）の特徴的な理由の一つとして、他者のツイート内容の引用が挙げられる。例えば、表 2 における事例 (4) では、Twitter で頻繁に用いられる「なう」という現在の状況を表す表現が引用部分に含まれている。このような引用部分の影響については、事前に引用部分を除外した上で判定を行うか、引用内容も場所依存記録とみなすことで対処できると考えられる。また、場所依存の可能性があるが、つぶやきのみでは断定できないという失敗事例もある。「こんなにも売ってます。 @\*\*\*\* < 画像の URL >」というツイートの場合、人手での分類においては負例と分類されているが、分類器では正例と判定された。このツイートでは、販売されているものは画像で示されており、一種の場所依存の情報と考えることもできるが、テキストのみでは何を売っているのかわからないため、誤検出とみなされている。

今後、より高精度な判定を実現するために、コーパスに含まれない表現への対応や、Twitter などのマイクロブログ特有の事象への対応を検討する必要がある。

## 6 おわりに

本研究では、観光や都市デザインなど、多様な場面で有用となり得る、場所依存記録の分類器を構築した。

街歩きイベント「100 人でつくる京都地図」を実施し、人間が意図的に発信した様相記録をもとに場所依存記録判定用コーパスを作成し、分類器の精度を検証した。検証の結果、約 74% の精度で判定できることを示した。また、Twitter のツイートに分類器を適用した結果、様相記録と同等の精度での判定が可能であり、Twitter のようなマイクロブログデータに含まれる偶発的な場所依存記録を抽出できることを示した。

今後は、判定に影響を与えるマイクロブログ特有の事象などに対応し、より高精度な判定を実現する。

### 謝辞

イベントの遂行にあたり、京都大学の島本裕美子氏、仲村哲明氏、臼田泰如氏、安田溪氏に協力をいただいた。東京大学の橋本康弘氏には、膨大な位置情報付ツイートデータを提供いただいた。ここに深く感謝の意を表す。なお本研究の一部は、博士課程教育リーディングプログラム「京都大学デザイン学大学院連携プログラム」および JST さきがけ「自然言語処理による診断支援技術の開発」プロジェクトの助成を受けた。

### 参考文献

- [1] Lynch, K. and Rivkin, M.: A WALK AROUND THE BLOCK, LANDSCAPE, vol.8, no.3, pp.24-34 (1959).
- [2] Nold, C.: Greenwich Emotion Map, <http://www.emotionmap.net/>
- [3] 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, 京都大学学位論文 (2012).
- [4] 服部哲, 復本寅之介, 横井茂樹, 速水治夫: まち歩きイベント向けの位置情報 SNS 応用システムの開発と評価, 情報処理学会研究報告, Vol.2011-IS-118, No.3, pp.1-6 (2011).
- [5] 鈴木昭二, 橋本真一, 布村重樹: 観光の楽しみを広げる拡張現実感用コンテンツ制作の試み, 情報処理学会 デジタルプラクティス, Vol.3, No.4, pp.313-322 (2012).