

修士論文

ソーシャルセンサのパフォーマンスと位置情報に関する 研究

大西 駿太郎

奈良先端科学技術大学院大学

先端科学技術研究科

情報理工学プログラム

主指導教員: 荒牧 英治

ソーシャル・コンピューティング研究室 (情報科学領域)

2024 年 1 月 18 日提出

本論文は奈良先端科学技術大学院大学先端科学技術研究科に
修士（工学）授与の要件として提出した修士論文である。

大西 駿太郎

審査委員：

主査	荒牧 英治	（情報科学領域 教授）
	中村 哲	（情報科学領域 教授）
	若宮 翔子	（情報科学領域 准教授）
	矢田 竣太郎	（情報科学領域 助教）

ソーシャルセンサのパフォーマンスと位置情報に関する 研究*

大西 駿太郎

内容梗概

ソーシャルセンシングとは、ソーシャルメディア上のユーザ活動を計測することで社会的変化を探知することであり、主に、X（旧、Twitter）をプラットフォームとして用いている。ソーシャルセンシングの基本となる前提は、ツイート投稿を観測することにより、対象となるイベントの規模（参加者数や社会的影響）を推定するというものが多い。しかし、投稿数が現実世界の規模に対応しているという仮説自体の検証は、厳密にはされていない。そこで本稿では相関検定により仮説を検証した。結果、仮説は常に成り立つわけではないことを明らかにした。また、位置情報を利用することで、地理的位置を考慮したソーシャルセンシングが行われている。その際、位置情報が付与された投稿が重要であるが、このような投稿は全体の0.4%ほどであると報告されている。不足した位置情報を推定する試みが多く行われているが、未だ基盤となる方法は確立されていない。そこで本稿ではメッセージレベルの位置推定 Web API を構築し、手法とツールを一般に公開することで、位置情報推定の基盤を確立することを目指した。

キーワード

自然言語処理, ソーシャルメディア, ソーシャルセンシング, 位置情報

*奈良先端科学技術大学院大学 先端科学技術研究科 修士論文, 2024 年 1 月 18 日.

Location assignment and performance analysis in social sensing*

Shuntaro Ohnishi

Abstract

Social sensing detects social changes by measuring user activity on social media, primarily utilizing X (formerly Twitter) as a platform. Geotagged posts are crucial for spatial analysis in social sensing, but they reportedly constitute only about 0.4 percent of the total. Numerous attempts have been made to estimate missing location information, but a fundamental method has not been established. Therefore, this study aims to establish a foundation for location information estimation by developing a message-level location estimation Web API and sharing the method and tools with the general public.

Furthermore, the fundamental assumption of social sensing often revolves around estimating the scale (number of participants and social impact) of a target event by observing tweet posts. However, the hypothesis that statements become louder as the scale increases has not been rigorously tested. In this study, we tested the hypothesis through a correlation test. The results revealed that the hypothesis does not always hold.

Keywords:

Natural Language Processing, Social Media, Social Sensing, Geo Location

*Master's Thesis, Graduate School of Science and Technology, Nara Institute of Science and Technology, January 18, 2024.

目次

図目次		v
第 1 章	はじめに	1
1.1	背景	1
1.2	本稿の構成	2
第 2 章	関連研究	3
第 3 章	投稿数と現実規模の一致	5
3.1	方法	5
3.1.1	データ収集	5
	調査対象地域	5
	人数と投稿数の取得	6
	地域の特性を示すデータの取得	7
3.1.2	ソーシャルセンサパフォーマンスの定義	8
3.2	結果	11
3.2.1	時間単位のソーシャルセンサパフォーマンス	11
3.2.2	年単位のソーシャルセンサパフォーマンス	12
3.3	考察	13
第 4 章	投稿に対する位置情報の補完	18
4.1	方法	18
4.1.1	データ収集	18
4.1.2	提案手法	18
4.2	結果	19
4.3	考察	22
4.3.1	緯度経度推定モデル	22
4.3.2	メッシュ推定モデル	22
第 5 章	おわりに	26

謝辞	28
参考文献	29
業績一覧	33
付録 A 時間単位のソーシャルセンサパフォーマンス	34

図目次

3.2	時間単位のソーシャルセンサパフォーマンス	11
3.1	調査対象地域: 青枠は京都駅, 黄枠は居住地域, 緑枠は観光地を表す	15
3.3	年単位のソーシャルセンサパフォーマンス: 青丸は平日, 橙プラスは休日のデータ点を表す. 青線は平日, 橙線は休日, 緑破線は全日のソーシャルセンサパフォーマンスを表す	17
4.1	Web API の出力画面: 1 次地域メッシュ推定モデルにおける「琵琶湖なう」の推定結果	21
4.2	「奈良公園なう」の予測結果, (a) 赤色のピンは緯度経度推定モデルによる「奈良公園なう」の予測位置, 青色のピンと黄色の円は「奈良公園」を含む train データの位置を示す	25

第 1 章 はじめに

1.1 背景

現代人の多くがスマートフォンを持ち、何らかのソーシャルメディアを利用している。この結果、情報発信や情報収集が活発になるとともに、ソーシャルメディアのインフラを用いたソーシャルセンシングが注目されている。ソーシャルセンシングとは、ソーシャルメディア上のユーザ活動を計測することで社会的変化を探知することである。[1]. ソーシャルセンシング自体は、特定のソーシャルメディアに依存するものではないが、通常は X (旧, Twitter) を用いるものが多い。Twitter を用いたソーシャルセンシングには、主に 3 つの利点がある。まず、ユーザ数が膨大であり、大規模な観察が可能であること。つぎに、Twitter の即時性を利用した迅速な情報収集が可能であること。さらに、位置情報を利用することで地理的位置を考慮した分析が可能であることが挙げられる

ソーシャルセンシングに関する先行研究では、災害・病気・交通状況やコミュニティ構成などさまざまな対象を観測してきた。対象は多岐にわたるもののソーシャルセンシングでは一般的に、Twitter での投稿数を集計して、イベントの規模や影響を推定することは共通している。ただし、この大前提である投稿数が現実世界の規模に対応しているという仮説そのものを直接検証した研究は、我々の知る限り存在しない。そこで、本稿では、ソーシャルセンサが仮定していた投稿数が現実世界の規模に対応するという仮説を検証する。これまで提案されたソーシャルセンサの種類は多くあるが、本稿では、最も素朴なソーシャルセンサとして、ある地域の投稿数から、その地域に滞在する人数（いわゆる、人口動態）を推定するタスクをベースに検証を行う。

また、位置情報を利用することで、地理的位置を考慮したソーシャルセンシングが行われている。しかしながら、地理的位置を考慮する際、位置情報が付与された投稿が重要であるにも関わらず、そのような投稿は極めて少ない。そこで、本稿では入力を投稿の本文、出力を緯度経度もしくは地域メッシュとした BERT ベースの推定モデルを構築した。同時に Web API として広く使用が可能な状態で公開し、

位置推定タスクの基盤となることを目指した*。

これらの取り組みを通じ，現状を明らかにすることで，ソーシャルセンシングを通じた社会現象の理解に貢献する。

1.2 本稿の構成

本稿の構成は以下の通りである。第2章ではソーシャルセンシングの代表的な先行研究を挙げる。同時に位置情報の不足に対処した位置情報推定タスクの概況を述べる。第3章では前章で述べた概況を踏まえ，位置情報推定タスクに取り込む。第4章では投稿数が現実世界の規模に対応するという仮説を検証する。第5章では本稿のまとめと位置情報を利用した地理空間上でのソーシャルセンサに関する展望を述べる。

*<https://aoi.naist.jp/texttolocation/>

第 2 章 関連研究

ソーシャルセンシングの代表的な先行研究としては、地震や台風の検知 [2]・人々の気分の推定 [3, 4]・インフルエンザの患者数を推定 [5]・交通渋滞把握 [6] が挙げられる。さらに、最近の研究では、ツイート投稿・コンテンツ・ユーザプロフィール情報に基づいて、地位ごとのコミュニティの性別や人種構成を推定するものなど多岐にわたる [7]。対象は多岐にわたるもののソーシャル・センシングでは一般的に、Twitter での投稿数を集計して、イベントの規模や影響を推定することは共通している。しかし、この大前提である投稿数が現実世界の規模に対応しているという仮説そのものを直接検証した研究は、我々の知る限り存在しない。

また、ソーシャルセンシングでは、位置情報を利用し地理的位置を考慮することがよくある。しかしながら、位置情報の付与された投稿は全体の 0.4% ほどであると報告されている [8]。このため、多くのソーシャルセンシングサービスは、メタ情報やテキスト中のランドマーク表現など、何らかの方法を用いて位置情報を推定することでカバレッジを上げている。位置情報を補完するためのツイートの位置予測研究も、この 10 年ほど活発に行われてきた。中でも、W-NUT2016 の Twitter の位置情報予測シェアードタスク* [9] のデータは、現在でもベンチマークとしてしばしば利用されている。位置推定タスクは、入力の形式と出力の形式の 2 つの軸で整理が可能である。

入力の形式: 位置推定の入力としては、発言者・ユーザの一連の文書か、単一文書（発言メッセージ、ツイート）かの 2 つのレベルがある。本稿では、前者の発言者の居住地を予測することをユーザレベル、後者の発言メッセージが投稿された位置を予測することをメッセージレベルと呼ぶ。

出力の形式: 位置推定の出力は、緯度経度、県や都市など様々な形式がある。等間隔のグリッドに分割することや [10]、投稿の多い地域では細かいグリッド、投稿の少ない地域では大きいグリッドを生成する手法が提案されている [11]。またメッセージレベルにおいて、その推定の不確実性を表現するために、出力位置の確率分布を出力する場合もある [12]。代表的な等間隔のグリッドとして、地域メッシュが

*<https://noisy-text.github.io/2016/geo-shared-task.html>

ある。地域メッシュは緯度経度から算出され、行政区画の変更にも左右されないという利点から、統計調査によく用いられる。このように考えると、例えば、前述のシェアードタスクの距離ベースの評価によるベストシステム [13] は、メッセージレベルを入力とし、緯度経度を出力すると捉えることができる。

第 3 章 投稿数と現実規模の一致

3.1 方法

投稿数が現実世界の規模に対応しているという仮説を検証する。本章では、現実世界の規模としてその地域に滞在する人数を採用した。人数は単純でわかりやすい変数あり、また人数の変動は、イベントの開催や社会的に重要な出来事などの現実世界の現象の影響を受ける可能性が高い。

仮説は、人数の時系列データと投稿数の時系列データのピアソン相関検定によって検証する。また同時に得られるピアソン相関係数を本研究では「ソーシャルセンサパフォーマンス」と呼び分析する。この値は理想的には 1 に近い値であることが望ましい。

3.1.1 データ収集

調査対象地域

調査対象地域は表 3.1 および図 3.1 に示すように、9つの地域を調査した。具体的には都市の主要駅である東京駅と京都駅、また居住地域として京都府中京区と京都大学学生街、くわえて有名な観光地として嵐山・金閣寺・清水寺・二条城・東寺を調査した。地域の分割方法として、まずその施設の地図平面上での形状を採用することが考えられる。しかしながら、この分割方法では施設形状の取得とその形状に則した投稿数・人数の取得が難しい。したがって、前章と同じく、各種統計でよく採用され、データの取得が容易である地域メッシュを採用した。地域メッシュとは、総務省統計局位が策定した緯度経度に基づく地域の分割方法である* [14]。まず、地域は第 1 次地域区画と呼ばれる一辺がおよそ 80km の正方形で分割される。その後、第 2 次地域区画・基準地域メッシュ・2 分の 1 地域メッシュ・4 分の 1 地域メッシュに分割される。それぞれ一辺の長さがおよそ 10km・1km・500m・250m の正方形である。本章では基準地域メッシュを元に地域を選択した。

*地域メッシュ統計について

https://www.stat.go.jp/data/mesh/m_tuite.html

名称	属性	地域メッシュ ID
(a) 東京駅	駅	5339-46-11
(b) 京都駅	駅	5235-36-80
(c) 中京区	居住地域	5235-46-10
(d) 京都大学学生街	居住地域	5235-46-42
(e) 嵐山	観光地	5235-45-13
(f) 金閣寺	観光地	5235-45-48
(g) 清水寺	観光地	5235-36-92
(h) 二条城	観光地	5235-45-19
(i) 東寺	観光地	5235-35-79

表 3.1: 調査対象地域

人数と投稿数の取得

データ収集期間は一致させるのが理想であるが、東京駅のデータは 2021 年、それ以外の 8 つの京都府の地域は 2022 年のデータである。

人数のデータは NTT ドコモから入手した[†] [15]. この人数は、各地域メッシュにおける携帯端末数に基づいて推計したものである。集計は 1 時間単位で行われる。1 台の端末が 1 時間該当の地域メッシュに留まった場合 1 人とカウントされ、端末が 30 分間滞在した場合、1/2 人としてカウントされる。

投稿数は Twitter のアプリケーション・プログラミング・インターフェース (API) を利用し、位置情報が付けられた投稿を取得した[‡] [16]. この API は GET リクエストによる単一のクエリに対応し、指定されたクエリに一致する過去の投稿を返す。クエリは、言語・位置情報を含むように構築した。具体的には、"lang: ja bounding_box:[west_long south_lat east_long north_lat]"とク

[†]モバイル空間統計

<https://mobaku.jp/>

[‡]X Developer Platform

<https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query>

エリを構築した。ここで, "lang: ja "は言語を日本語に設定する部分である。また, "bounding_box:[west_long south_lat east_long north_lat]"はそれぞれの地域メッシュの西・南・東・北の境界の緯度と経度の値をそれぞれ入力することで位置情報を設定する部分である。その後, 特定の時間帯に過剰に投稿するユーザの影響を軽減するため, 1時間以内に行われた連続した投稿をまとめて1投稿として集計した。その結果, 実質的に1時間の投稿数を集計するのではなく, 1時間の間に投稿したTwitter ユーザ数を集計している。

地域の特性を示すデータの取得

人数と投稿数の他に地域の特性を反映していると考えられるデータを取得した。具体的には, 施設の数・最寄り駅からの距離・居住者数・観光魅力度などを取得した。これらのデータは考察のセクションで用いる。

まず, 施設の数について, 地域ごとにコンビニエンスストア, 飲食店, 学校, ホテルの件数を Overpass turbo [§] [17] を用いて取得した。Overpass turbo は, OpenStreetMap データのウェブベースのデータマイニングツールであり無料で利用できる。クエリは, "shop=convenience・amenity=restaurant・amenity=school・tourism=hotel・highway=bus_stop "である。データの取得は2023年10月1日に行われた。

最寄り駅までの距離は geopy (version 2.3.0) を用いて算出した。地域メッシュの代表点は地域メッシュの中心し, 地域メッシュに駅が含まれる場合は0kmとする。

居住者数は, 日本政府が実施した2020年人口動態調査から取得した[¶] [18]。この調査は5年ごとに実施されており2020年版が最新である。

観光魅力度は, 観光客数の代替案である。観光客数の適当なデータが発見できなかったため, 観光魅力度を採用した。観光魅力度は東京都と京都府で実施された別々のアンケート調査から推計する。それぞれ「東京都データカタログ: 都民の

[§]Overpass turbo

<https://overpass-turbo.eu/>

[¶]e-stat 人口および世帯

<https://www.e-stat.go.jp/gis/statmap-search?page=1&type=1&toukeiCode=00200521&toukeiYear=2020&aggregateUnit=S&serveyId=S002005112020&statsId=T001100>

観光に対する意識調査^{||} [19] と「京都市産業観光局: 観光客の動向等に係る調査^{**}」 [20] を採用した. 東京駅の観光魅力度は, 東京都の魅力のあるエリアとして「東京都心部 (千代田区, 中央区, 港区, 新宿区, 文京区)」をあげた人の割合である. また京都府においては, 回答者のうちその観光スポットに実際に訪れた人の割合に基づいている. 注意として, 回答エリアは個々の地域メッシュと完全に一致しない. 表 3.2 に, 地域メッシュと対応するアンケート内での名称を示す.

名称	属性	アンケート内での名称
(a) 東京駅	駅	都心部
(b) 京都駅	駅	京都駅周辺
(c) 中京区	居住地域	京都御所周辺
(d) 京都大学学生街	居住地域	銀閣寺・哲学の道・百万遍周辺
(e) 嵐山	観光地	嵯峨嵐山
(f) 金閣寺	観光地	きぬかけの路
(g) 清水寺	観光地	清水・祇園周辺
(h) 二条城	観光地	二条城・壬生周辺
(i) 東寺	観光地	京都駅周辺

表 3.2: アンケート調査と地域の対応

3.1.2 ソーシャルセンサパフォーマンスの定義

まず人数 X と投稿数 Y はそれぞれ時間 ($i = 24$ 時間) 日付 ($j = 365$ 日) 場所 ($k = 9$ 地点) で抽出可能な 3 次元データである. これを X_{ijk}, Y_{ijk} と表す.

ソーシャルセンサパフォーマンスは, 地域ごとに時間単位 SSP_{daily} , また年単位 SSP_{yearly} で定義する. くわえて分析のために全日・平日・休日で人数と投稿数を

^{||} 東京都データカタログ 都民の観光に対する意識調査
<https://data.tourism.metro.tokyo.lg.jp/data/ishiki/>

^{**} 京都市産業観光局 観光客の動向等に係る調査
https://www.kyokanko.or.jp/wp/wp-content/uploads/kyoto_tourism_stat_2022.pdf

分割した. ここで, 休日は, 土曜日・日曜日・祝日・年末年始 (12月29日~1月3日) を指す.

任意の時刻 i における任意の場所 k の時間単位のソーシャルセンサパフォーマンス SSP_{daily} は以下のように定義される.

$$SSP_{daily} = R_{ik} = \frac{\text{cov}_j(X_{ijk}, Y_{ijk})}{\text{std}_j(X_{ijk}) \text{std}_j(Y_{ijk})} \quad (3.1.1)$$

(3.1.1) において, cov_j は添字 j に関する共分散であり, std_j は添字 j に関する標準偏差である. それぞれ以下のように定義される:

$$\begin{aligned} \text{cov}_a(X_{ab\dots}, Y_{ab\dots}) \\ = \frac{1}{N_a} \sum_a (X_{ab\dots} - \text{mean}_a(X_{ab\dots}))(Y_{ab\dots} - \text{mean}_a(Y_{ab\dots})) \end{aligned}$$

$$\text{std}_a(X_{ab\dots}) = \sqrt{\frac{1}{N_a} \sum_a (X_{ab\dots} - \text{mean}_a(X_{ab\dots}))^2}$$

$$\text{mean}_a(X_{ab\dots}) = \frac{1}{N_a} \sum_a X_{ab\dots}$$

このとき, N_a は全日の場合は 365(日) であり, 平日と休日で分割したときはそれぞれの日数である. 具体的には, 2021 年において平日 240(日) 休日 125(日), 2022 年において平日 239(日) 休日 126(日) である. あわせて人数 X_{ijk} と投稿数 Y_{ijk} も平日のデータのみ, 休日のデータのみを用いる.

次に任意の場所 k における年単位のソーシャルセンサパフォーマンス SSP_{yearly} は以下のように定義される.

$$SSP_{yearly} = R_k = \frac{\text{cov}_j(S_{jk}, T_{jk})}{\text{std}_j(S_{jk}) \text{std}_j(T_{jk})} \quad (3.1.2)$$

(3.1.2) において, S_{jk}, T_{jk} は X_{ijk}, Y_{ijk} をそれぞれ日単位で集計したもので, 一日の合計人数と合計投稿数を表す. それぞれ以下のように定義される:

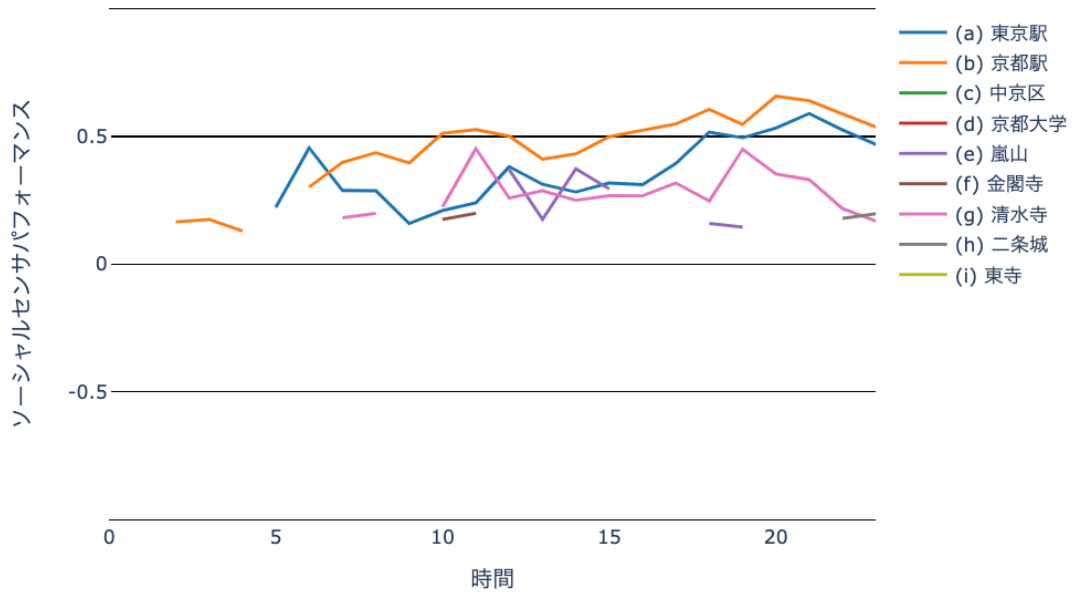
$$S_{jk} = \sum_i X_{ijk}$$

$$T_{jk} = \sum_i Y_{ijk}$$

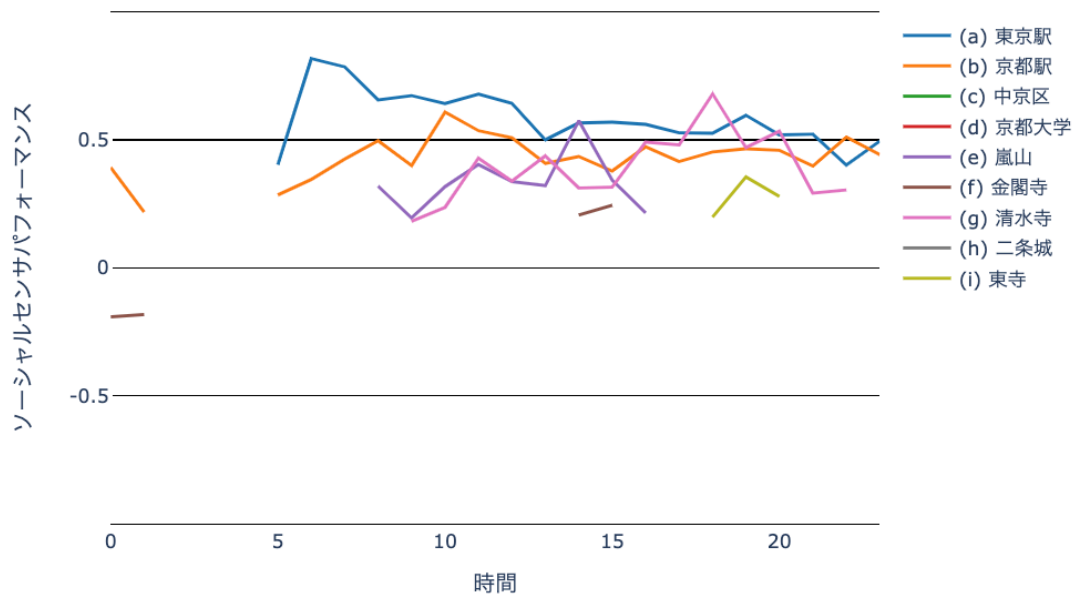
このとき平日・休日の分割において、時間単位のソーシャルセンサパフォーマンスと同様の操作を行う。

3.2 結果

3.2.1 時間単位のソーシャルセンサパフォーマンス



(a) 平日



11
(b) 休日

図 3.2: 時間単位のソーシャルセンサパフォーマンス

時間単位のソーシャルセンサパフォーマンスを図 3.2 に示す. より具体的な値は付録のセクションに示す.

0時から5時の間の p 値は多くの地域で 0.05 を超えている. また, 中京区と金閣寺の p 値は多くの時間帯で 0.05 を超えている. ソーシャルセンサパフォーマンスは, 日中の明るい時間帯は正の値である. しかし, その増減は地域によって異なる.

3.2.2 年単位のソーシャルセンサパフォーマンス

	全日 (p 値)	平日 (p 値)	休日 (p 値)
(a) 東京駅	-0.38 (0.00)	0.62 (0.00)	0.79 (0.00)
(b) 京都駅	0.69 (0.00)	0.81 (0.00)	0.72 (0.00)
(c) 中京区	-0.19 (0.00)	0.06 (0.39)	0.23 (0.01)
(d) 京都大学	0.14 (0.01)	0.01 (0.84)	0.04 (0.63)
(e) 嵐山	0.65 (0.00)	0.60 (0.00)	0.62 (0.00)
(f) 金閣寺	0.02 (0.00)	-0.02 (0.75)	-0.09 (0.35)
(g) 清水寺	0.74 (0.00)	0.63 (0.00)	0.73 (0.00)
(h) 二条城	0.48 (0.00)	0.26 (0.00)	0.36 (0.00)
(i) 東寺	0.18 (0.00)	0.22 (0.00)	0.30 (0.00)

表 3.3: ソーシャルセンサパフォーマンス: 太字は正の値でかつ p 値 ≤ 0.05

表 3.3 に示すように, 全日におけるソーシャルセンサパフォーマンスは, 最小-0.38 から最大 0.74 と広い幅をとる. 強い正の値を示した地域は約半数にとどまり, 逆に負の値を示した地域もあった. 平日と休日を比べると東京駅, 京都駅, 清水寺, 二条城ではおよそ 0.1 程度の差がある.

最も興味深い地域は東京駅である. 表 3.3 に示すように, ソーシャルセンサパフォーマンスは全日でマイナスである. しかし, 平日と休日で見ると, 平日と休日どちらも強い正の値を示している図 3.3(a) に示すように, シンプソンのパラドックスが発生している. 図 3.3(g) を見ると, 東京駅と異なり, 清水寺では全ての条件で正の値となっている. また, 東寺は清水寺よりもソーシャルセンサパフォーマンス

スが低い。全日、平日、休日で一貫して 0.4~0.5 程度の差が見られる。表 3.3 が示すように、年単位のソーシャルセンサパフォーマンスは地域によってさまざまである。考察のセクションでは、年単位のソーシャルセンサパフォーマンスの変動要因について分析をおこなう。

これらの結果から、本研究の結論を示す。表 3.3 は場所や平日・休日などの条件によるソーシャルセンサパフォーマンス (相関) の有無、またその大小を示している。さらに図 3.2 は、時間の違いによるソーシャルセンサパフォーマンスの有無とその大小を示している。これらの結果が示すように、ソーシャルセンサパフォーマンスは常に認められるわけではなく、その値も負の値から正の値まで大きく変動する。したがって、投稿数が現実世界の規模に対応するという仮説は棄却される。

3.3 考察

ソーシャルセンサパフォーマンスの変動要因を明らかにするために、まず、人数と投稿数を日毎に集計し、その後、それぞれの基本統計量 (平均・中央値・合計・最大値・最小値) を算出した。更にこれらの基本統計量と年単位のソーシャルセンサパフォーマンス、また、地域特性を示すデータ (施設の数・最寄り駅からの距離・居住者数・観光魅力度) と年単位のソーシャルセンサパフォーマンスの相関を定義した。

それぞれの基本統計量もそれぞれの地域特性を表すデータも全てデータ長は地域数 N_k ^{††}である。したがって区別せずすべて a_k と表すと、相関 R' は以下のように定義される:

$$\begin{aligned} R' &= \frac{\text{cov}_k(SSP_{yearly}, a_k)}{\text{std}_k(SSP_{yearly}) \text{std}_k(a_k)} \\ &= \frac{\text{cov}_k(R_k, a_k)}{\text{std}_k(R_k) \text{std}_k(a_k)} \end{aligned}$$

シンプソンのパラドックスが発生している東京駅の影響を考慮し、平日・休日の結果のみを示す。

^{††} $N_k = 9$

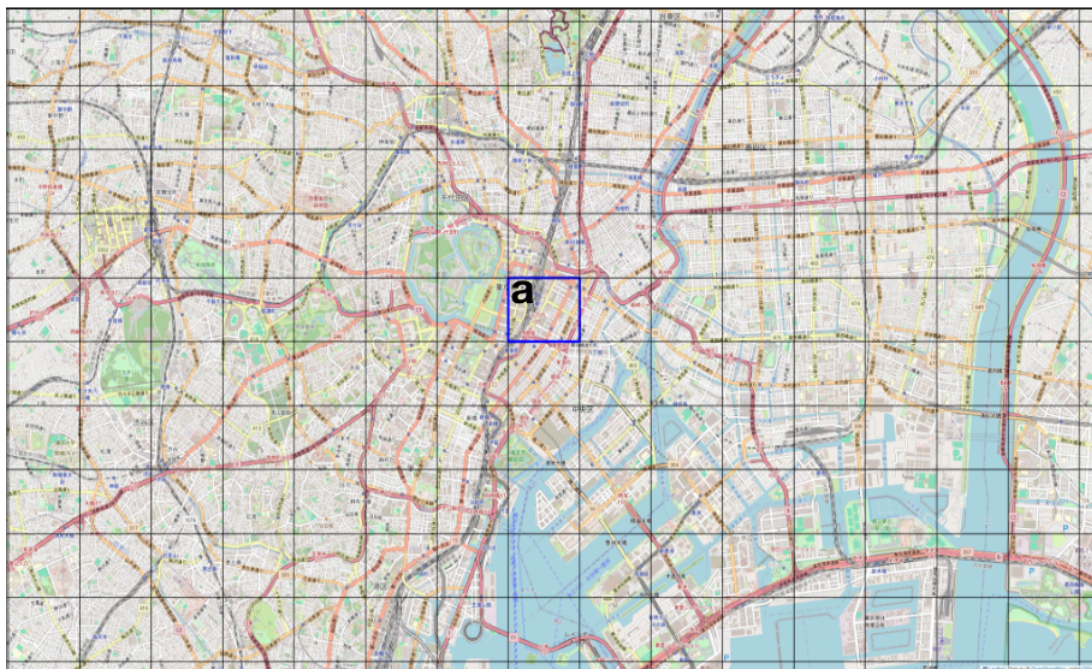
	平日 (p 値)	休日 (p 値)
観光魅力度	0.69 (0.04)	0.74 (0.02)
居住者数	-0.83 (0.01)	-0.73 (0.02)

表 3.4: ソーシャルセンサパフォーマンスとの相関: p 値 ≤ 0.05 である結果のみ示した

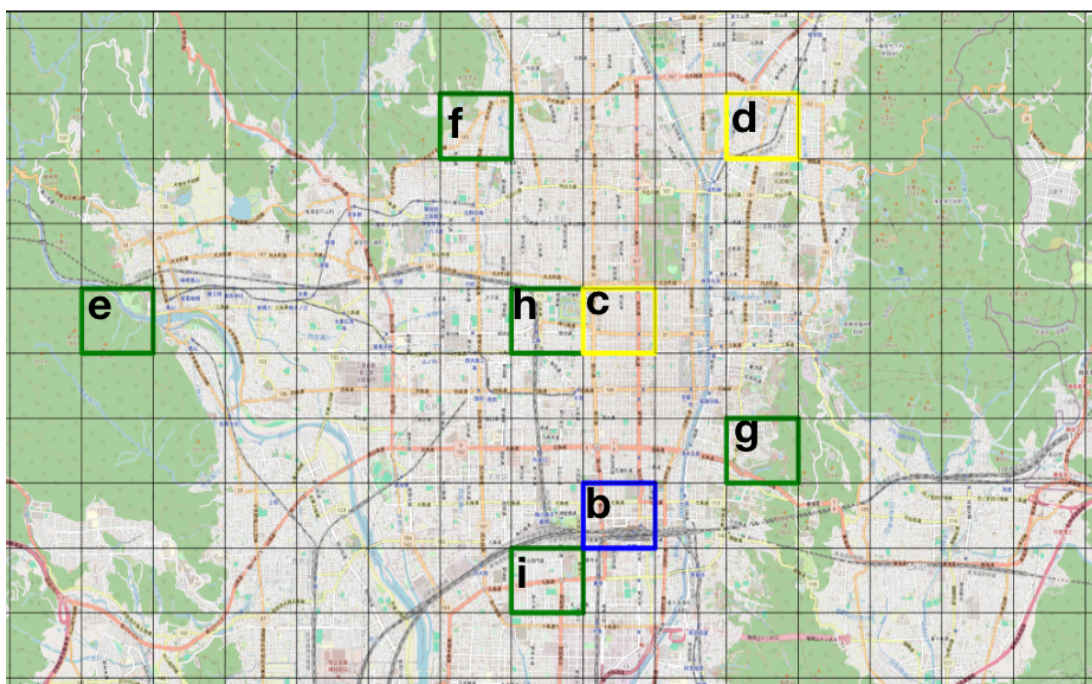
表 3.4 は, p 値が 0.05 以下の項目を示している. 観光魅力度と正の相関があり居住者数とは負の相関がある.

一般的に, 居住者は観光客よりも投稿が少ないと考えられる. また, 特に人気のある観光地では, 観光客が自分の体験を共有する意欲が高いことが多い. 言い換えれば, 居住者が非活動的な傾向があるのに対し, 観光客は活動的な傾向にある. つまり, 活動的な人が多くかつ非活動的な人が少ない条件下で, ソーシャルセンサパフォーマンスが大きくなると言える. 同じ理由で, コンサートやフェスティバルのような特別なイベント時にソーシャルセンサパフォーマンスが高くなると考えられる.

意外なことに, 人数と投稿数の基本統計量との相関は見られなかった. 人口が密集し, ソーシャルメディアの利用が盛んな都市部では, ソーシャルセンサパフォーマンスが大きくなることが期待される. それに反して今回の結果は期待に沿うものではなかった.

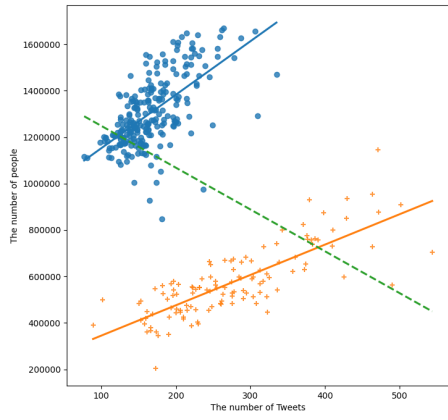


(a) 東京

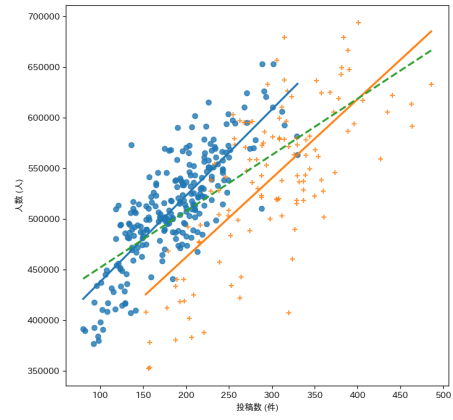


(b) 京都

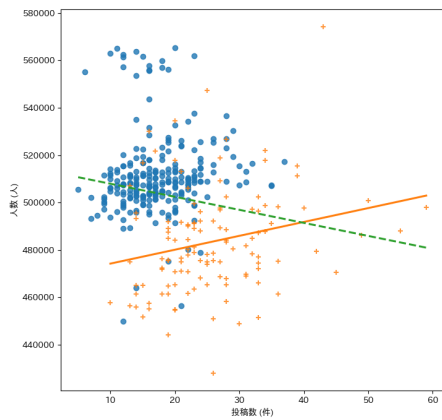
図 3.1: 調査対象地域: 青枠は京都駅, 黄枠は居住地域, 緑枠は観光地を表す



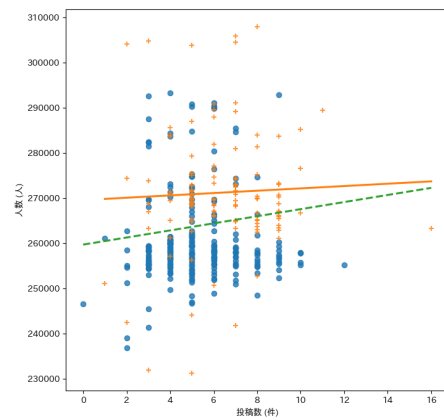
(a) 東京駅



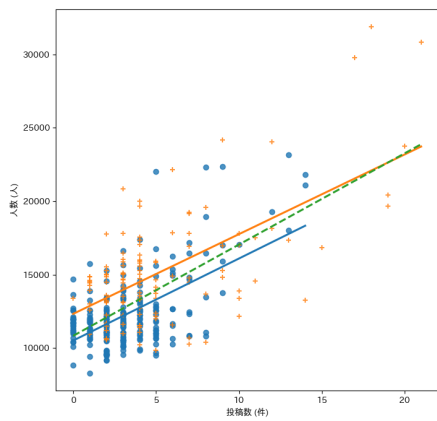
(b) 京都駅



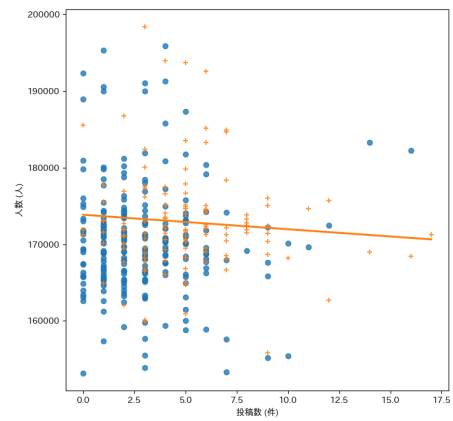
(c) 中京区



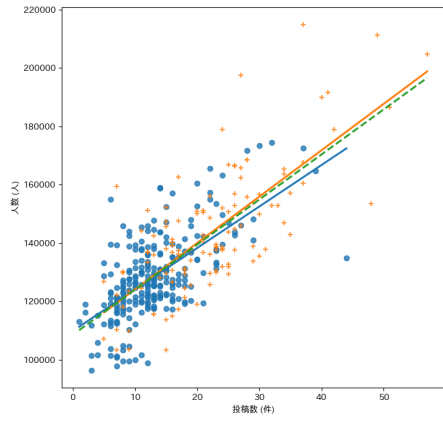
(d) 京都大学



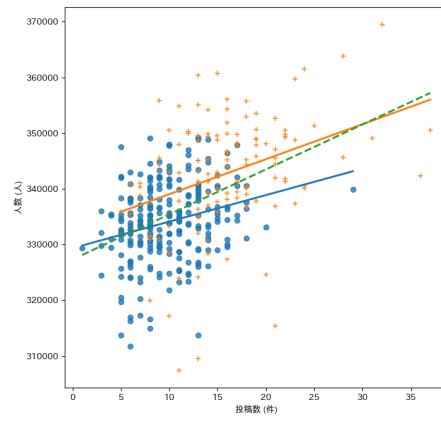
(e) 嵐山



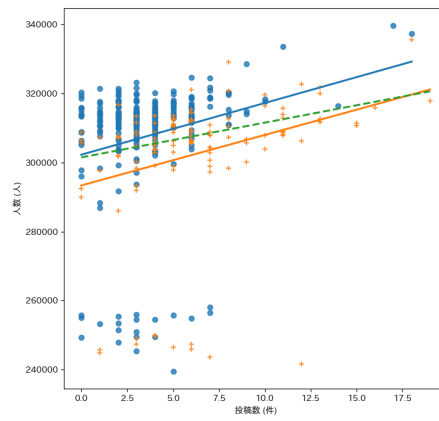
(f) 金閣寺



(g) 清水寺



(h) 二条城



(i) 東寺

図 3.3: 年単位のソーシャルセンサパフォーマンス：青丸は平日，橙プラスは休日のデータ点を表す．青線は平日，橙線は休日，緑破線は全日のソーシャルセンサパフォーマンスを表す

第 4 章 投稿に対する位置情報の補完

4.1 方法

4.1.1 データ収集

投稿の日本国内における位置を推定するモデルを構築するために、日本における位置情報が付与された投稿を利用する。Twitter のアプリケーション・プログラミング・インターフェース (API) を利用し、位置情報の付与された投稿を取得した。この API は GET リクエストによる単一のクエリに対応し、指定されたクエリに一致する過去の投稿を返す。構築したクエリは "place_country:jp" であり、日本国内の位置情報の付与された投稿をすべて指定するクエリである。2022 年 7 月 1 日から同月 31 日まで、1 時間あたりおよそ 1500 件、合計 1,111,576 件のデータを収集した。

次に、いくつかの前処理を行う。まず、前提として、各投稿には位置情報として緯度・経度の情報が付与されている。この投稿に地域メッシュの情報を付与する。本章では各投稿に一辺が約 80 km の正方形である第 1 次地域区画 (1 次メッシュ) と約 10 km である第 2 次地域区画 (2 次メッシュ) の情報を付与した。

次に特定の地域ごとのサンプル数の不均衡に対処するために、第一次地域区画においては同一メッシュ内のデータは 300,000 件以内に制限し、過剰な分は削除した。続いて、本文に含まれる URL, メンションおよび半角・全角スペースを除去した。

4.1.2 提案手法

本章では、緯度・経度を回帰問題として推定するモデル (緯度経度推定モデル) と、地域メッシュを分類問題として推定するモデル (メッシュ推定モデル) を構築する。地域区画推定モデルはさらに、1 次メッシュ推定モデルと 2 次メッシュ推定モデルとをそれぞれ作成する。入力テキストをエンコードするモデルとして Bidirectional Encoder Representations from Transformers (BERT) [21] を日本

表 4.1: 各モデルに対する test loss と test accuracy

推定モデル	Test loss	Test accuracy
緯度経度推定モデル	24	–
1次メッシュ推定モデル	3.3	0.21
2次メッシュ推定モデル	2.9	0.27

語コーパスで事前学習したモデル* を採用した。緯度経度推定モデルおよび1次メッシュ推定モデルでは日本全体を推定範囲とする。一方、2次メッシュ推定モデルでは推定範囲を東京（第1次区画: 5339）に限定し、計算負荷が高くなり過ぎない程度の予測メッシュ数に抑えた。というのも、日本全体を2次メッシュで分割すると、クラス数が約50,000とモデルが肥大化し学習が困難であった。このときの投稿数は323,234件である。1次メッシュ推定モデルで140、2次メッシュ推定モデルで61のクラス分類問題である。

緯度経度推定モデルの目的関数と評価関数は、ともに緯度経度の平均二乗誤差とする。なお、地球が完全な球でないことによって起きる実際の距離のずれは補正しない。また、メッシュ推定モデルの目的関数と評価関数には cross entropy を用いた。学習条件として最適化手法は Adam, 学習率は 1.0×10^{-5} , エポック数は5, バッチサイズは2とした。

4.2 結果

位置情報が付与された投稿を8（訓練）:1（検証）:1（テスト）の割合で分割し、構築した緯度経度推定モデルとメッシュ推定モデルの精度を評価した。表4.1に緯度経度およびメッシュ推定のテストデータ推定結果を示す。緯度経度推定モデルについて、日本周辺の緯度、経度1°あたりの距離は約90kmであること、また、 $\sqrt{\text{test loss}} = \sqrt{24} \doteq 4.9$ より平均で約441km推定位置がズレている。また、2つのメッシュ推定モデルの予測精度はどちらも2割程度であった。

*<https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

またこれらのモデルを用いた Web API を構築した。図 4.1a は Web API の入力画面であり、投稿本文の入力と推定範囲の指定が可能である。推定範囲で全国を選択すると 1 次地域メッシュ推定モデル、東京を選択すると 2 次地域メッシュ推定モデルでの推定結果を表示する。また、追加機能として京都および奈良を推定範囲として選択可能である。



(a) Web API の入力画面



(b) 休日

図 4.1: Web API の出力画面: 1 次地域メッシュ推定モデルにおける「琵琶湖なう」の推定結果

表 4.2: 地名の有無と推定精度の分析: 地名を含む投稿の方が Accuracy が高いことがわかる.

	投稿数		Accuracy	
	1次	2次	1次	2次
すべて	65,504	32,323	0.21	0.27
地名を含まない	55,933	27,319	0.13	0.21
地名を含む (1つ)	7,479	3,833	0.54	0.57
地名を含む (複数)	2,092	1,171	0.63	0.70

4.3 考察

4.3.1 緯度経度推定モデル

緯度経度推定モデルの予測が大きくズレる原因を分析するために、緯度経度推定モデルによる「奈良公園なう」の予測位置と、学習データにおいて“奈良公園”を含む投稿を地図上に可視化した (図 4.2a).

赤ピンがモデルの出力であり、黄円で示された範囲に 25 件の学習データがある、また、青ピンに 1 件学習データが存在した。実際には奈良公園は黄円の内部に存在する。青ピンの投稿本文は「奈良公園の鹿みたいに白鳥がいます」であり、この投稿においては、投稿者が実際にいる地点と言及された地点が一致していない。この地点のずれは A/B 問題 [22] と呼ばれている。この影響により推定が東側に大きく引っ張られていると考えられる。

4.3.2 メッシュ推定モデル

1 次メッシュ推定モデルにおける「奈良公園なう」の出力結果を図 4.2b に示す。緯度経度推定モデルと異なり、奈良公園を含む地域メッシュを正しく出力している。これは緯度経度推定モデルが回帰問題を扱っているのに対し、1 次メッシュ推定モデルはクラス分類問題を扱ったためであると考えられる。

続いて 1 次メッシュ推定モデルにおける本文に含まれる地名の数と推定精度の関

係を分析した。具体的には、Spacy の “ja_core_news_md”[†] の GPE ラベルに基づき、投稿本文中の地名を抽出した (表 4.2)。

テストデータのうち、地名を含まないものが大半 (約 85%) を占め、地名を 1 つ含むものが約 11%, 地名を複数含むものが約 4% であった。

表 4.3: メッシュ推定モデルにおける正解例と誤り例。 (1)-(6) は 1 次メッシュ推定モデルによる結果, (7)-(12) は 2 次メッシュ推定モデルによる結果。太字は固有表現抽出により抽出された地名を表す。

ID	正解 (おおよその位置)	出力 (おおよその位置)	本文
(1)	5235 (大阪府)	5235 (大阪府)	あんたらはなーんもせんやんけ。
(2)	4931 (大分県)	4931 (大分県)	大分に来たら待ち合わせは大分駅のあの場所に#夜のイチスタ
(3)	6542 (北海道旭川市)	6542 (北海道旭川市)	旭川であれ北海道に来たからには寿司を食わねばなりません。
(4)	5338 (山梨県)	5436 (石川県)	サントリー白州蒸溜所ツアー樽貯蔵庫に足を踏み入れた瞬間の香り…
(5)	5135 (奈良県)	5238 (静岡県)	当券があるみたいなのでそろそろ難波寄ってから現場に向かいますか
(6)	5133 (瀬戸内海)	5135 (京都府)	九州から「さくら」に乗って新大阪から大和西大寺へ ww
(7)	533965 (埼玉県さいたま市)	533965 (埼玉県さいたま市)	雨の中ですが、着々と準備が進んでいます！
(8)	533946 (東京都台東区)	533946 (東京都台東区)	上野着弾!!! [URL]
(9)	533945 (東京都新宿区)	533945 (東京都新宿区)	新宿 1352 発東京メトロ丸ノ内線池袋行き東京まで#ノア乗車録
(10)	533931 (神奈川県相模原市)	533945 (東京都新宿区)	セミリタイアしてしまったんですかね
(11)	533904 (神奈川県横浜市)	533913 (神奈川県綾瀬市)	神奈川の結果…
(12)	533944 (東京都武蔵野市)	533945 (東京都新宿区)	札幌が重要現場被りで行けないから次の大阪を見たら終了か

同様に 2 次メッシュ推定モデルにおいても地名の有無と予測精度の関係を分析した。表 4.3 の (1)(4)(7)(10) は地名を含まない例, (2)(5)(8)(11) は地名を 1 つ含む例, (3)(6)(9)(12) は地名を複数含む例である。(1) のように地名を含まない投稿の予測精度は低い。このような投稿の位置を予測することは人間にも難しいので妥当な結果と言える。地名を複数含むときは、地名を 1 つしか含まないときと比べ予測精度が高い。その要因として、(3) の旭川市と北海道のように包含関係にある地名が出現する、または同一の地名が繰り返し出現することが挙げられる。また同一の地名が繰り返し出現する投稿は、ハッシュタグにより地名が追加されていることが多かった。一方で、(12) のように位置的に全く異なる地点を挙げている場合は予測が難しい。

また、1 次メッシュ推定モデルにおける地名を含まない正解データは 7,305 件で

[†]<https://spacy.io/models/ja>

あった。地名以外の地点を特定できる要因を調べるために該当データを分析したが、有益な分析結果は得られなかった。

このように地名を含まない投稿が多いことが、予測精度が向上しない一因であると考えられる。加えて、隣接する地域メッシュにおいて、投稿内容に大きな差が見られないことが、メッシュ推定モデルの正解率が20%から30%にとどまった原因であると考えられる。



(a) 緯度経度推定モデル



(b) 1次メッシュ推定モデル

図 4.2: 「奈良公園なう」の予測結果, (a) 赤色のピンは緯度経度推定モデルによる「奈良公園なう」の予測位置, 青色のピンと黄色の円は「奈良公園」を含む train データの位置を示す

第 5 章 おわりに

本稿ではソーシャルセンシングによる地理空間分析における 2 つの課題に取り組んだ。

第 3 章では投稿数が現実世界の規模に対応しているという仮説を検証した。ピアソンの相関係数と相関検定の結果は、相関の見られる条件と無相関である条件が混在し、また相関係数の大小も様々であった。したがって、投稿数が現実世界の規模に対応するという仮説は棄却された。

しかしながら、先行研究では投稿数が現実世界の規模に対応する事例も報告されている。例えば、投稿数と歩行者数の相関が報告されている [23]。加えて、投稿数と観光客の流れの相関も報告されている [24]。本研究の貢献は、人数と投稿数の間に相関関係がなく、投稿数が現実世界の規模に対応しない事例があるという潜在的な反例を提供したことである。

また、いくつかの先行研究と異なる点として、投稿の本文を自然言語処理をもちいて分析している点が挙げられる [5]。第 3 章では投稿に対してほとんど前処理をせずそのまま利用している。しかし、先行研究では投稿本文に対して自然言語処理の技術、例えば、単語抽出・case-folding・正規化・単語の除去などがよく使われている [25] [26]。これらの手法は推定精度の向上に寄与するが、その効果や影響についてはまだ十分に議論されていない。

また第 3.3 節では、「観光魅力度」と「居住者数」がソーシャルセンサパフォーマンスの変動要因である可能性を示した。ただし、相関関係は因果関係を示すものではないという点に注意が必要である。加えて、ソーシャルセンサパフォーマンスとしてピアソン相関係数を採用し、そのソーシャルセンサパフォーマンスと相関関係にあるデータを探索した。つまり、相関係数の変動要因を相関検定により推定した。同じ方法を再帰的に利用している。このようなアプローチは、これまでの研究ではほとんど行われておらず、その有効性と解釈可能性については議論・改良の余地がある。今後の研究では、ソーシャルセンサパフォーマンスを従属変数とする線形回帰のような、別の方法をもちいることでより有効な分析ができる可能性がある。

さらに、我々はソーシャルセンサのパフォーマンスという概念を導入したが、効果的なソーシャルセンシングのための閾値を明らかにする必要がある。ソーシャル

センサパフォーマンスは1であることが理想だが、そのような条件が達成されることは現実的ではない。したがって、効果的なソーシャルセンシングに必要な閾値を明らかにすることが必要である。

第4章では位置情報の付与された投稿を用いて位置推定モデルを作成した。同時に Web API として公開した。緯度経度推定モデルでは A/B 問題により予測が大きくなること、また地域メッシュ推定モデルでは地名を含まない投稿が多いため位置推定が困難であることを示した。

第4章で提案した手法の精度は、緯度経度モデルで平均約 441km のズレ、メッシュ推定モデルでは Accuracy が 20% から 30% であり、位置情報をもとにしたサービスや研究への応用のためには、さらなる精度の向上は不可欠である。ツイート本文には地名が含まれていないものが多く、人間にも推定が難しいと考えられる。ユーザのプロフィール情報や投稿時間など他の情報を用いることにより精度が改善する可能性がある。

また出力の提示方法についても検討の余地がある。例えば「朝起きるのが辛かった」というツイートがあったとき、人間であれば予測不可能な問題として扱うことになる。しかし、提案モデルは必ず特定の位置を出力するため、位置推定における曖昧性を適切に表すことができていない。特定の地域メッシュではなく、すべての地域メッシュに対する確率を出力する方法も検討したい。

位置推定タスクは、本質的に、真の正解・位置であるという保障が得られない。推定結果はあくまで推定結果である。例えば、第4章におけるメッシュ推定モデルを用いると、確率として2割ほど正しい位置が出力される。モデルの精度を向上させても、位置は確率的にしか定まらない。したがって、本稿の目指した位置推定タスクの基盤構築について述べると、ある種のベースラインとなる手法は確立できるが、その利用については各利用方法に応じて調整する必要があると考える。

謝辭

参考文献

- [1] Dong Wang, Boleslaw K. Szymanski, Tarek Abdelzaher, Heng Ji, and Lance Kaplan. The Age of Social Sensing. *Computer*, Vol. 52, No. 1, pp. 36–45, 2019.
- [2] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, p. 851–860, New York, NY, USA, 2010. Association for Computing Machinery.
- [3] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Niels Rosenquist. Pulse of the Nation: U.S. Mood Throughout the Day inferred from Twitter, 2010. <https://www.ccs.neu.edu/home/amislove/twittermood/>.
- [4] Mark E. Larsen, Tjeerd W. Boonstra, Philip J. Batterham, Bridianne O’Dea, Cecile Paris, and Helen Christensen. We Feel: Mapping Emotion on Twitter. *IEEE Journal of Biomedical and Health Informatics*, Vol. 19, No. 4, pp. 1246–1252, 2015.
- [5] Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. Twitter Catches The Flu: Detecting Influenza Epidemics using Twitter. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 1568–1576, 2011.
- [6] Senzhang Wang, Lifang He, Leon Stenneth, Philip S. Yu, and Zhoujun Li. Citywide Traffic Congestion Estimation with Social Media. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Association for Computing Machinery, 2015. <https://doi.org/10.1145/2820783.2820829>.
- [7] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and Rosenquist. Understanding the Demographics of Twitter Users. *Proceedings of the International AAAI Conference on Web and Social Media*,

- Vol. 5, No. 1, pp. 554–557, Aug. 2021. <https://ojs.aaai.org/index.php/ICWSM/article/view/14168>.
- [8] ツイッターの空間分析. 古今書院, 2019.
- [9] Bo Han, Afshin Rahimi, Leon Derczynski, and Timothy Baldwin. Twitter Geolocation Prediction Shared Task of the 2016 Workshop on Noisy User-generated Text. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, 2016.
- [10] Benjamin Wing and Jason Baldridge. Simple Supervised Document Geolocation with Geodesic Grids. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011.
- [11] Stephen Roller, Michael Speriosu, Sarat Rallapalli, Benjamin Wing, and Jason Baldridge. Supervised Text-Based Geolocation Using Language Models on an Adaptive Grid. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL '12*, p. 1500–1510, 2012.
- [12] Hayate Iso, Shoko Wakamiya, and Eiji Aramaki. Density Estimation for Geolocation via Convolutional Mixture Density Network. *arXiv preprint arXiv:1705.02750*, 2017.
- [13] Yasuhide Miura, Motoki Taniguchi, Tomoki Taniguchi, and Tomoko Ohkuma. A Simple Scalable Neural Networks based Model for Geolocation Prediction in Twitter. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, pp. 235–239, 2016.
- [14] 総務省統計局. 地域メッシュ統計について. [\https://www.stat.go.jp/data/mesh/m_tuite.html](https://www.stat.go.jp/data/mesh/m_tuite.html).
- [15] NTT docomo. モバイル空間統計. <https://mobaku.jp/>.
- [16] X Developer Platform . [\https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query](https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query).
- [17] OpenStreetMap Wiki. Ja:overpass turbo — openstreetmap wiki,, 2022. <https://wiki.openstreetmap.org/w/index.php?\title=JA:>

- Overpass_turbo&oldid=2275722.
- [18] e-stat. 人口及び世帯. 2020. \a href="https://www.e-stat.go.jp/gis/statmap-search?%5Cpage=1&%5Ctype=1&%5CtoukeiCode=00200521&%5CtoukeiYear=%5C2020&%5CaggregateUnit=S&%5CserveyId=S002005112020&%5CstatsId=%5CT001100">https://www.e-stat.go.jp/gis/statmap-search?%5Cpage=1&%5Ctype=1&%5CtoukeiCode=00200521&%5CtoukeiYear=%5C2020&%5CaggregateUnit=S&%5CserveyId=S002005112020&%5CstatsId=%5CT001100.
- [19] 東京都観光データカタログ .
都民の観光に関する意識調査 . \a href="https://data.tourism.metro.tokyo.lg.jp/data/ishiki/">https://data.tourism.metro.tokyo.lg.jp/data/ishiki/.
- [20] 京都市産業観光局. 観光客の動向等に係る調査. p. 47. \a href="https://www.kyokanko.or.jp/wp/wp-content/uploads/%5Ckyoto_tourism_stat_2022.pdf">https://www.kyokanko.or.jp/wp/wp-content/uploads/%5Ckyoto_tourism_stat_2022.pdf.
- [21] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, author =.
- [22] Rabindra Lamsal, Aaron Harwood, and Maria Rodriguez Read. Where did you Tweet from? Inferring the Origin Locations of Tweets Based on Contextual Information. *arXiv preprint arXiv:2211.16506*, 2022.
- [23] Debjit Bhowmick, Stephan Winter, and Mark Stevenson. Using Georeferenced Twitter Data to Estimate Pedestrian Traffic in an Urban Road Network. In Krzysztof Janowicz and Judith A. Verstegen, editors, *11th International Conference on Geographic Information Science (GIScience 2021) - Part I*, Vol. 177 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 1:1–1:15, Dagstuhl, Germany, 2020. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. <https://drops.dagstuhl.de/entities/document/10.4230/%5CLIPIcs.GIScience.2021.I.1>.
- [24] Johannes Scholz and Janja Jeznik. Evaluating Geo-Tagged Twitter Data to Analyze Tourist Flows in Styria, Austria. *ISPRS International Journal of Geo-Information*, Vol. 9, No. 11, 2020. <https://www.mdpi.com/2220-9964/9/11/681>.
- [25] Rudy Arthur, Chris A. Boulton, Humphrey Shotton, and Hywel T. P. Williams. Social sensing of floods in the uk. *PLoS ONE*, Vol. 13, , 2017. <https://api.semanticscholar.org/CorpusID:5460691>.

- [26] Syifa Khairunnisa, Adiwijaya Adiwijaya, and Said Al Faraby. Pengaruh text preprocessing terhadap analisis sentimen komentar masyarakat pada media sosial twitter (studi kasus pandemi covid-19). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 2021. [\https://api.semanticscholar.org/CorpusID:235251521](https://api.semanticscholar.org/CorpusID:235251521).

業績一覧

国内会議

- [1] 大西駿太郎, 矢田竣太郎, 若宮翔子, 荒牧英治
ツイート発言の座標またはグリッドの予測基盤の開発
言語処理学会第 29 回年次大会 (NLP), 2023
- [2] 大西駿太郎, 矢田竣太郎, 若宮翔子, 荒牧英治
人口動態を基盤としたソーシャルセンサのパフォーマンス分析
データ工学と情報マネジメントに関するフォーラム (DEIM), 2024

付録 A 時間単位のソーシャルセンサパフォーマンス

表 A.1 と表 A.2 に時間単位のソーシャルセンサパフォーマンスの値を示す。「No tweets」はその時間帯にツイートがなく、分散が 0 であったため、ソーシャルセンサパフォーマンスを定義できなかった条件である。

Hour	0	1	2	3	4	5	6	7	8	9	10	11
(a) 東京駅	0.03 (0.59)	0.07 (0.29)	-0.07 (0.30)	-0.01 (0.85)	-0.05 (0.44)	0.22 (0.00)	0.46 (0.00)	0.29 (0.00)	0.29 (0.00)	0.16 (0.01)	0.21 (0.00)	0.24 (0.00)
(b) 京都駅	0.39 (0.00)	0.08 (0.20)	0.17 (0.01)	0.18 (0.01)	0.13 (0.04)	0.08 (0.19)	0.30 (0.00)	0.40 (0.00)	0.44 (0.00)	0.40 (0.00)	0.51 (0.00)	0.53 (0.00)
(c) 中京区	-0.05 (0.46)	-0.02 (0.70)	-0.00 (0.98)	-0.05 (0.43)	-0.04 (0.58)	0.02 (0.72)	-0.04 (0.49)	0.03 (0.64)	0.09 (0.17)	0.03 (0.63)	-0.04 (0.55)	0.05 (0.48)
(b) 京都大学	-0.01 (0.93)	No tweets	-0.05 (0.40)	0.03 (0.65)	-0.11 (0.09)	0.15 (0.02)	-0.04 (0.49)	0.04 (0.57)	0.06 (0.37)	-0.00 (0.96)	0.01 (0.89)	-0.01 (0.87)
(e) 嵐山	0.10 (0.11)	0.04 (0.58)	-0.05 (0.43)	0.11 (0.10)	0.00 (0.94)	0.08 (0.23)	-0.02 (0.77)	0.16 (0.01)	0.10 (0.12)	0.27 (0.00)	0.01 (0.92)	0.12 (0.06)
(f) 金閣寺	0.01 (0.90)	-0.00 (0.98)	-0.05 (0.47)	0.09 (0.14)	0.04 (0.56)	-0.01 (0.92)	0.08 (0.22)	0.09 (0.18)	0.00 (0.99)	0.09 (0.18)	0.18 (0.01)	0.20 (0.00)
(g) 清水寺	0.08 (0.19)	0.07 (0.31)	0.02 (0.77)	0.02 (0.81)	-0.13 (0.04)	0.02 (0.78)	0.09 (0.16)	0.18 (0.00)	0.20 (0.00)	-0.06 (0.35)	0.23 (0.00)	0.45 (0.00)
(h) 二条城	0.17 (0.01)	-0.05 (0.46)	-0.02 (0.71)	0.04 (0.56)	0.05 (0.48)	-0.03 (0.68)	0.04 (0.56)	0.05 (0.44)	0.04 (0.49)	0.14 (0.03)	0.02 (0.72)	0.14 (0.03)
(i) 東寺	0.09 (0.17)	0.05 (0.41)	0.04 (0.57)	0.04 (0.50)	-0.07 (0.26)	-0.07 (0.31)	0.08 (0.19)	0.10 (0.14)	0.15 (0.02)	0.06 (0.34)	0.09 (0.16)	0.04 (0.56)
Hour	12	13	14	15	16	17	18	19	20	21	22	23
(a) 東京駅	0.38 (0.00)	0.31 (0.00)	0.28 (0.00)	0.32 (0.00)	0.31 (0.00)	0.39 (0.00)	0.52 (0.00)	0.50 (0.00)	0.53 (0.00)	0.59 (0.00)	0.53 (0.00)	0.47 (0.00)
(b) 京都駅	0.50 (0.00)	0.41 (0.00)	0.43 (0.00)	0.50 (0.00)	0.53 (0.00)	0.55 (0.00)	0.61 (0.00)	0.55 (0.00)	0.66 (0.00)	0.64 (0.00)	0.59 (0.00)	0.54 (0.00)
(c) 中京区	0.08 (0.24)	-0.04 (0.56)	0.00 (0.98)	0.03 (0.68)	-0.04 (0.57)	0.05 (0.41)	0.02 (0.72)	0.04 (0.53)	0.07 (0.28)	0.13 (0.04)	0.07 (0.27)	0.06 (0.38)
(b) 京都大学	-0.03 (0.69)	0.05 (0.48)	0.12 (0.06)	-0.19 (0.00)	0.03 (0.67)	0.01 (0.91)	0.06 (0.34)	-0.02 (0.70)	0.04 (0.58)	-0.02 (0.77)	0.00 (0.97)	-0.00 (0.96)
(e) 嵐山	0.37 (0.00)	0.18 (0.01)	0.37 (0.00)	0.30 (0.00)	0.11 (0.08)	0.07 (0.29)	0.16 (0.01)	0.15 (0.02)	0.04 (0.54)	0.06 (0.34)	0.06 (0.32)	0.14 (0.03)
(f) 金閣寺	0.02 (0.74)	0.07 (0.30)	0.02 (0.78)	0.11 (0.09)	0.11 (0.09)	0.05 (0.47)	0.06 (0.36)	-0.12 (0.06)	0.10 (0.11)	-0.14 (0.03)	-0.09 (0.17)	-0.14 (0.02)
(g) 清水寺	0.26 (0.00)	0.29 (0.00)	0.25 (0.00)	0.27 (0.00)	0.27 (0.00)	0.32 (0.00)	0.25 (0.00)	0.45 (0.00)	0.35 (0.00)	0.33 (0.00)	0.22 (0.00)	0.17 (0.01)
(h) 二条城	0.02 (0.80)	0.07 (0.29)	0.09 (0.15)	0.04 (0.52)	0.09 (0.14)	0.11 (0.08)	0.18 (0.01)	0.08 (0.23)	0.18 (0.00)	0.02 (0.72)	0.18 (0.01)	0.20 (0.00)
(i) 東寺	0.08 (0.22)	0.21 (0.00)	-0.04 (0.56)	0.04 (0.51)	0.07 (0.31)	0.08 (0.21)	0.12 (0.06)	0.21 (0.00)	0.10 (0.11)	0.05 (0.44)	-0.00 (0.97)	0.10 (0.11)

表 A.1: 平日における時間単位のソーシャルセンサパフォーマンス: 括弧内は p 値を表す

Hour	0	1	2	3	4	5	6	7	8	9	10	11
(a) 東京駅	0.21 (0.02)	-0.03 (0.72)	-0.11 (0.21)	-0.19 (0.04)	-0.06 (0.52)	0.40 (0.00)	0.82 (0.00)	0.79 (0.00)	0.66 (0.00)	0.67 (0.00)	0.64 (0.00)	0.68 (0.00)
(b) 京都駅	0.39 (0.00)	0.22 (0.02)	0.10 (0.25)	0.26 (0.00)	0.13 (0.17)	0.29 (0.00)	0.35 (0.00)	0.43 (0.00)	0.50 (0.00)	0.40 (0.00)	0.61 (0.00)	0.54 (0.00)
(c) 中京区	0.10 (0.27)	0.01 (0.91)	-0.01 (0.95)	0.21 (0.02)	0.01 (0.95)	0.03 (0.72)	-0.04 (0.64)	0.12 (0.18)	0.11 (0.21)	0.04 (0.64)	0.32 (0.00)	0.15 (0.09)
(b) 京都大学	0.03 (0.71)	-0.04 (0.64)	-0.20 (0.03)	-0.04 (0.67)	No tweets	-0.08 (0.37)	0.01 (0.88)	-0.06 (0.50)	0.00 (1.00)	0.09 (0.30)	0.03 (0.74)	-0.02 (0.82)
(e) 嵐山	-0.05 (0.59)	0.06 (0.48)	0.17 (0.07)	No tweets	-0.00 (0.96)	0.09 (0.34)	0.03 (0.76)	0.01 (0.89)	0.32 (0.00)	0.20 (0.03)	0.32 (0.00)	0.40 (0.00)
(f) 金閣寺	-0.19 (0.03)	-0.18 (0.04)	0.02 (0.80)	-0.05 (0.59)	-0.01 (0.92)	No tweets	-0.02 (0.81)	-0.11 (0.21)	-0.16 (0.09)	-0.02 (0.85)	0.03 (0.72)	0.21 (0.02)
(g) 清水寺	-0.04 (0.64)	0.03 (0.77)	0.13 (0.15)	0.01 (0.93)	0.04 (0.67)	-0.03 (0.78)	0.11 (0.22)	0.28 (0.00)	0.05 (0.56)	0.18 (0.04)	0.24 (0.01)	0.43 (0.00)
(h) 二条城	-0.04 (0.65)	0.10 (0.29)	0.05 (0.55)	0.01 (0.90)	-0.00 (0.96)	-0.00 (1.00)	-0.07 (0.45)	-0.08 (0.40)	0.04 (0.70)	0.19 (0.03)	0.09 (0.34)	0.22 (0.02)
(i) 東寺	-0.06 (0.54)	0.08 (0.40)	-0.03 (0.75)	0.04 (0.66)	0.08 (0.41)	0.11 (0.21)	0.18 (0.05)	0.08 (0.38)	0.08 (0.38)	0.23 (0.01)	0.16 (0.08)	0.13 (0.16)
Hour	12	13	14	15	16	17	18	19	20	21	22	23
(a) 東京駅	0.64 (0.00)	0.50 (0.00)	0.57 (0.00)	0.57 (0.00)	0.56 (0.00)	0.53 (0.00)	0.53 (0.00)	0.60 (0.00)	0.52 (0.00)	0.52 (0.00)	0.40 (0.00)	0.50 (0.00)
(b) 京都駅	0.51 (0.00)	0.41 (0.00)	0.44 (0.00)	0.38 (0.00)	0.47 (0.00)	0.42 (0.00)	0.45 (0.00)	0.47 (0.00)	0.46 (0.00)	0.40 (0.00)	0.51 (0.00)	0.44 (0.00)
(c) 中京区	0.29 (0.00)	0.18 (0.05)	0.17 (0.05)	0.08 (0.38)	-0.02 (0.85)	0.02 (0.79)	0.15 (0.10)	0.13 (0.14)	0.02 (0.82)	0.28 (0.00)	0.05 (0.58)	0.17 (0.06)
(b) 京都大学	0.05 (0.59)	-0.01 (0.87)	0.04 (0.65)	0.03 (0.78)	0.03 (0.71)	-0.04 (0.65)	-0.10 (0.27)	0.01 (0.87)	-0.03 (0.72)	0.08 (0.38)	-0.14 (0.12)	0.01 (0.92)
(e) 嵐山	0.34 (0.00)	0.32 (0.00)	0.58 (0.00)	0.34 (0.00)	0.22 (0.02)	0.01 (0.87)	0.32 (0.00)	0.10 (0.28)	0.09 (0.30)	0.13 (0.17)	-0.11 (0.23)	0.13 (0.16)
(f) 金閣寺	0.11 (0.21)	-0.01 (0.94)	0.21 (0.02)	0.25 (0.01)	0.12 (0.17)	-0.12 (0.20)	-0.17 (0.06)	-0.01 (0.90)	-0.09 (0.34)	-0.05 (0.57)	-0.05 (0.59)	-0.07 (0.46)
(g) 清水寺	0.34 (0.00)	0.44 (0.00)	0.31 (0.00)	0.32 (0.00)	0.49 (0.00)	0.48 (0.00)	0.68 (0.00)	0.47 (0.00)	0.53 (0.00)	0.29 (0.00)	0.30 (0.00)	0.01 (0.94)
(h) 二条城	0.07 (0.46)	0.07 (0.47)	0.09 (0.31)	0.10 (0.27)	0.23 (0.01)	-0.03 (0.71)	0.17 (0.07)	0.21 (0.02)	0.17 (0.06)	0.08 (0.39)	0.14 (0.11)	0.07 (0.42)
(i) 東寺	0.19 (0.04)	0.06 (0.50)	0.01 (0.91)	-0.09 (0.33)	0.07 (0.47)	0.13 (0.16)	0.20 (0.03)	0.36 (0.00)	0.28 (0.00)	0.04 (0.65)	-0.09 (0.33)	0.26 (0.00)

表 A.2: 休日における時間単位のソーシャルセンサパフォーマンス: 括弧内は p 値を表す