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

大西駿太朗† 矢田竣太郎† 若宮 翔子† 荒牧 英治†

† 奈良先端科学技術大学院大学 ソーシャル・コンピューティング研究室
〒 630-0192 奈良県生駒市高山町 8916 番地-5
E-mail: †{onishi.shuntaro.or3,s-yada,wakamiya,aramaki}@is.naist.jp

あらまし 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研究は多くあるが，基本となる前提は，投稿を観測することにより，観測対象の規模（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，実際に投稿数が観測対象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，本研究では，ある地域の投稿数から，その地域に滞在する人数（人口動態）を推定するタスクをベースに，相関検定による仮説の検証を行う．結果，仮説は常に成り立つわけではないこと．また，仮説の成立条件としては，人口動態に関しては,「観光魅力度」と「居住者数」が大きな要因となることを明らかにした．

キーワード 地理データ活用， ソーシャルメディア， ユーザモデル, 行動分析

1 はじめに

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，人々の情報発信や情報収集が活発になった．同時に，これらの情報をもとにした，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 Twitter（現，X）を用いるものが多い．Twiiter 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．1 つ目は，ユーザ数が膨大であり，大規模な観察が可能であること．2 つ目は，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．3 つ目は，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．

観測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 での投稿数をもとに，観測対象の規模，例えば，人数や社会的影響の大きさ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観測対象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．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，投稿数, コンテンツ, ユーザプロファイル情報に基づいて, コミュニティの性別や人種構成を推定している [7]．どの研究も投稿数をもとに，台風, インフルエンザ, 交通渋滞など観測対象の規模を推定している．

本研究では，ソーシャルセンサ研究が前提としていた投稿数が観測対象の規模に対応するという仮説を検証する．実験 1 ではピアソンの相関検定の結果をもとに仮説の検証を行う．実験 2 では同時に得られる相関係数を「ソーシャルセンサパフォーマンス」としてその値の変動要因を分析する，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，本研究では，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（人口動態）を推定するタスクをベースに検証を行う．

2 実験とその目的

2.1 実験 1： 仮説の検証

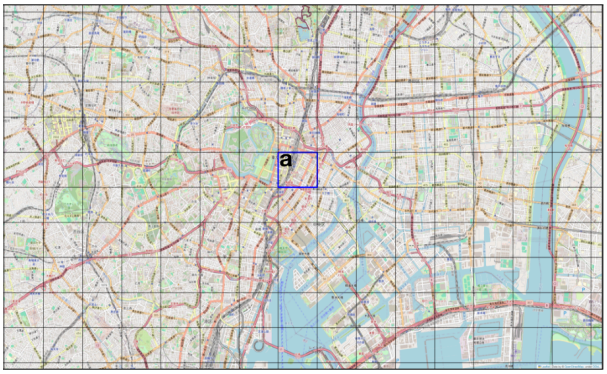
実験 1 の目的は．投稿数が観測対象の規模に対応しているという仮説の検証である．本章では，観測対象としてその地域に滞在する人数 (人口動態) を採用した．コンサートへの参加人数や被災者数などとは異なり，属性条件を持たず，同時にそれらの基礎となる変数であるため採用した．仮説は，人数と投稿数のピアソン相関検定によって検証する．帰無仮説は「人数と投稿数は無相関である」とする．帰無仮説が棄却されたすべての地域, 時間帯で相関係数が大きい正の値であるとき仮説は成立する．逆に，帰無仮説が棄却されたすべての地域で大きい正の相関がなかったり，値が条件によって大きく変動するとき仮説は成立しない．本稿では以後この相関係数を「ソーシャルセンサパフォーマンス」と呼称する．

2.2 実験 2：ソーシャルセンサパフォーマンスの値の変動要因

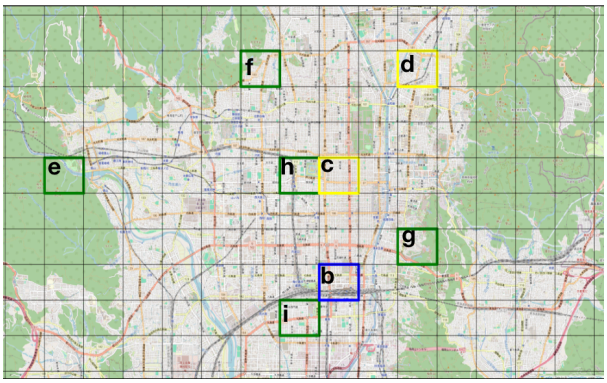
実験 2 の目的は，ソーシャルセンサパフォーマンスの値の変動要因を明らかにすることである．ソーシャルセンサパフォーマンスは理想的には 1 に近い値であることが望ましい．しかし，現実的にソーシャルセンサパフォーマンスが 1 である場面は殆どなく，値に変動があると考える．一般的には，人数が多くかつ投稿の多い都市部のほうがソーシャルセンシングが盛んに行われている．こうした状況を踏まえ，変動要因として地域ごとの人数と投稿数の多寡や，その地域が都市部であるか観光地であるかなどを候補として分析する．

表 1: 調査対象地域

名称	属性	地域メッシュID	アンケート内での名称
(a) 東京駅	駅	5339-46-11	都心部
(b) 京都駅	駅	5235-36-80	京都駅周辺
(c) 中京区	居住地域	5235-46-10	京都御所周辺
(d) 京都大学学生街	居住地域	5235-46-42	銀閣寺・哲学の道・百万遍周辺
(e) 嵐山	観光地	5235-45-13	嵯峨嵐山
(f) 金閣寺	観光地	5235-45-48	きぬかけの路
(g) 清水寺	観光地	5235-36-92	清水・祇園周辺
(h) 二条城	観光地	5235-45-19	二条城・壬生周辺
(i) 東寺	観光地	5235-35-79	京都駅周辺



(a) 東京



(b) 京都

図 1: 調査対象地域: 青枠は駅，黄枠は居住地域，緑枠は観光地を表す

3 実験 1： 仮説の検証

3.1 データ収集

3.1.1 調査対象地域

表 1 および図 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．さらに，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, 東寺を選んだ．この選定は以下の予想に基づくものである．まず，駅は人数が多いものの，人々の往来が激しく投稿が少ないと考えた．したがって，ソーシャルセンサパフォーマンスの値については予想がつかない．次に，観光地は人数も投稿数も多いと考えられ，ソーシャルセンサパフォーマンスが大きい正の値になると予想した．居住地域は人数も投稿も少ないと考えられ，ソーシャルセンサパフォーマンスが弱い正の値になると予想した．

地域の分割方法として，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する．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 [1]．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．分割された地域メッシュにはそれぞれ固有の ID が割り当てられ，分割が細かくなるに連れてその桁が増えていく．選択した地域と地域メッシュID の対応を表 1 に示す．

3.1.2 人数と投稿数の取得

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 ドコモが提供するモバイル空間統計 [2] から入手した．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 1 時間毎に推計され，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）[3]を利用し，位置情報が付与された投稿を取得した．この API では指定されたクエリに一致する過去の投稿を取得可能である．クエリは，言語情報と位置情報を含むように構築した．具体的には `lang:ja bounding_box:[w s e n]` とクエリを構築した．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:[w s e n]` は `w s e n` に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．取得したデータには，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．

3.2 ソーシャルセンサパフォーマンスの定義

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 次元データであり，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 SSP_{area} と定義する．同時に，時間を考慮した SSP_{hour} を定義する．くわえて，平日と休日では人々の行動が変容する可能性が高いと考え，分析のために全日, 平日, 休日で人数と投稿数を分割した．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．

SSP_{area} は各地域ごとに人数と投稿数を日毎で集計し，その後，相関係数を求めたものである．任意の場所 k における地域

1：https://www.stat.go.jp/data/mesh/m_tuite.html
2：https://mobaku.jp/
3：https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query

ごとのソーシャルセンサパフォーマンス SSP_{area} は以下のように定義される．

$$SSP_{area} = R_k = \frac{\mathrm{cov}_j(S_{jk}, T_{jk})}{\mathrm{std}_j(S_{jk})\ \mathrm{std}_j(T_{jk})} \tag{1}$$

(1) 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あり，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:

$$S_{jk} = \sum_i x_{ijk}, T_{jk} = \sum_i y_{ijk}$$

(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:

$$\mathrm{cov}_a(u_{ab\ldots}, v_{ab\ldots}) = \frac{1}{N_a}\sum_a (u_{ab\ldots} - \mathrm{mean}_a(u_{ab\ldots}))(v_{ab\ldots} - \mathrm{mean}_a(v_{ab\ldots})) \tag{2}$$

$$\mathrm{std}_a(u_{ab\ldots}) = \sqrt{\frac{1}{N_a}\sum_a (u_{ab\ldots} - \mathrm{mean}_a(u_{ab..}))^2} \tag{3}$$

$$\mathrm{mean}_a(u_{ab\ldots}) = \frac{1}{N_a}\sum_a u_{ab\ldots} \tag{4}$$

(2), (3), (4) において，$u_{ab\ldots}$ の次元数に関わらず，添字 a に関して計算する．このとき，N_a は全日の場合は 365 (日) である，平日と休日で分割したとき，2021 年はそれぞれ 240 (日) と 125 (日)，2022 年はそれぞれ 239 (日) と休日 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．

時間ごとのソーシャルセンサパフォーマンス SSP_{hour} は人数と投稿数の日毎の集計をおこなわず，相関係数を求めたものである．任意の場所 k における任意の時刻 i のソーシャルセンサパフォーマンス SSP_{hour} は以下のように定義される．

$$SSP_{hour} = R_{ik} = \frac{\mathrm{cov}_j(x_{ijk}, y_{ijk})}{\mathrm{std}_j(x_{ijk})\ \mathrm{std}_j(y_{ijk})}$$

このとき，平日と休日に関する分割は地域ごとのソーシャルセンサパフォーマンスと同じように行う．

3.3 結　　果

3.3.1 地域ごとのソーシャルセンサパフォーマンス

表 2 に示すように，全日における地域ごとの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 と広い範囲をとる．大きい正の値を示した地域は京都駅, 嵐山, 清水寺, 二条城と約半数にとどまり，逆に東京駅や中京区のように，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 2 に示すように，地域ごとのソーシャルセンサパフォーマンスは全日で-0.38 と負の値である．しかし，平日 0.62 と休日 0.79 はどちらも大きい正の値を示している．図 2 (a) に示すように，母集団での相関と，母集団を分割した集団での相関が，異なっている．図 2 (b) を見ると，東京駅と異なり，京都駅では全ての条件で正の値となっている．表 2 が示すように，地域ごと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．そこで， 4.2 節では地域ごとのソーシャルセンサパフォーマンスの値の変動要因について分析する．

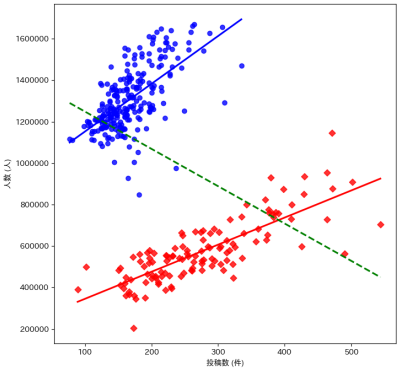
表 2: 地域ごとのソーシャルセンサパフォーマンス（太字は p 値 < 0.05）

	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)**

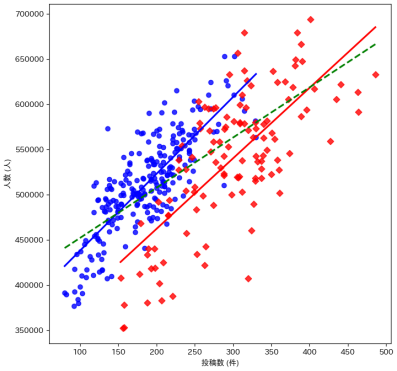
3.3.2 時間ごとのソーシャルセンサパフォーマンス

図 3 は，p 値 < 0.05 であった時間ごとのソーシャルセンサパフォーマンスを折れ線グラフで可視化したものである．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．

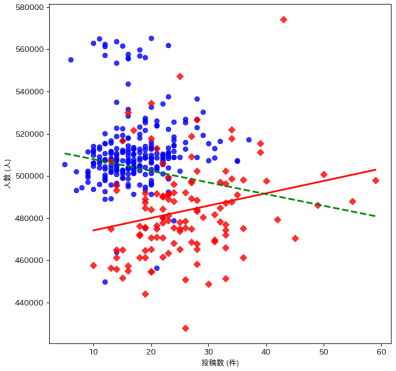
これらの結果から，本稿の結論を示す. 表 2 は場所や平日, 休日などの条件によるソーシャルセンサパフォーマンスが負の値から正の値まで大きく変動することを示している. さらに図 3 は，時間の違いによるソーシャルセンサパフォーマンスの変動を示している．すなわち，人数と投稿数の相互関係の強弱は条件によってさまざまである．したがって，投稿数が観測対象の規模に対応するという仮説は常に成り立つわけではない．



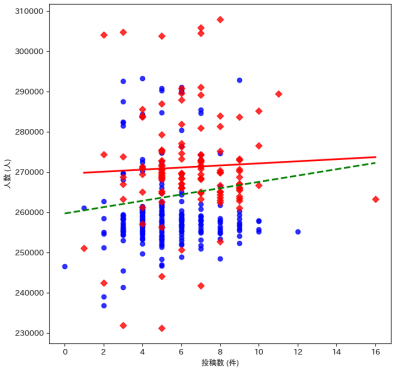
(a) 東京駅



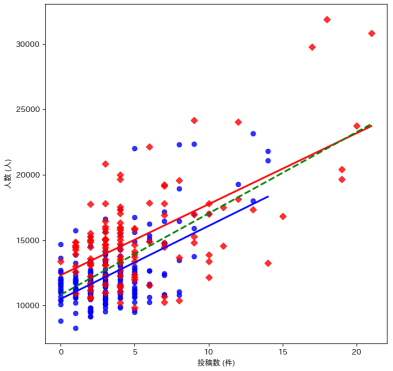
(b) 京都駅



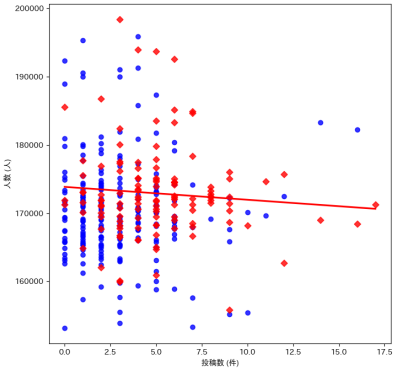
(c) 中京区



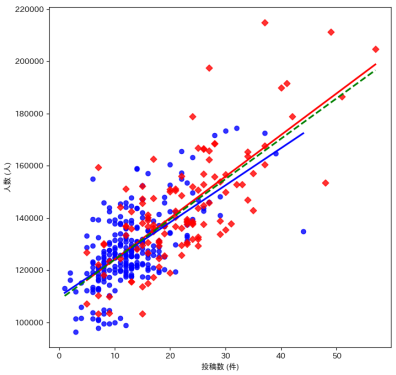
(d) 京都大学



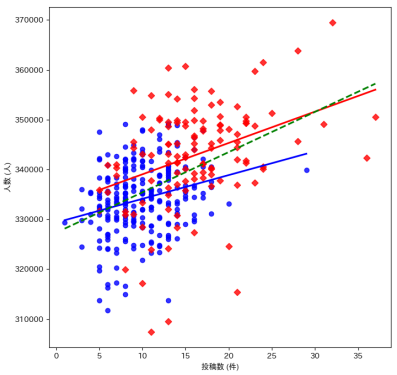
(e) 嵐山



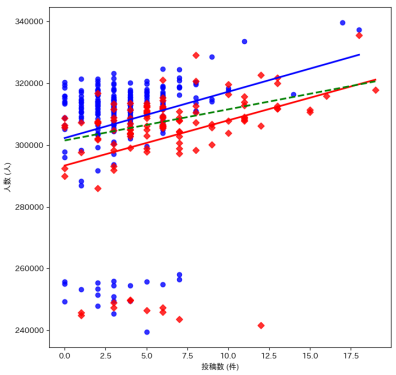
(f) 金閣寺



(g) 清水寺

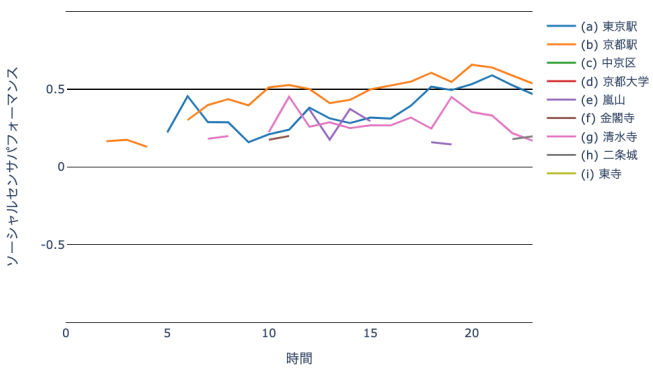


(h) 二条城

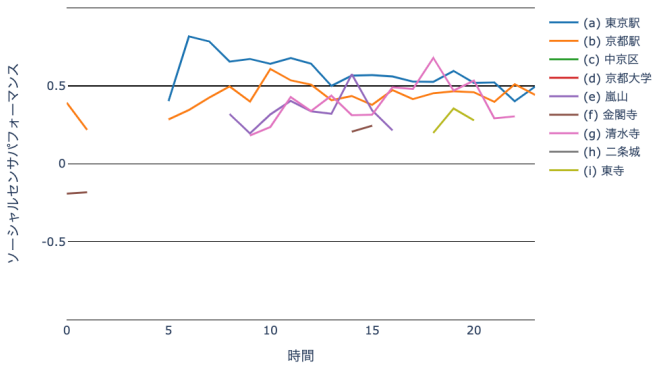


(i) 東寺

図 2: 地域ごとのソーシャルセンサパフォーマンス： 青丸は平日，赤ひし形は休日のデータ点を示す．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示す



(a) 平日



(b) 休日

図 3: 時間ごとのソーシャルセンサパフォーマンス

4 実験 2： ソーシャルセンサパフォーマンスの値の変動要因

4.1 地域の特性を示すデータの取得

実験 1 に追加して，人数と投稿数の他にソーシャルセンサパフォーマンスの値の変動要因を分析するため，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．

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 [4]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．

最寄り駅からの距離は，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座標し，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 年人口動態調査の結果から取得した [5]. この調査は 5 年ごとに実施されており，2024 年現在，2020 年版が最新である．

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: 都民の観光に対する意識調査 [6]」と「京都府市産業観光局: 観光客の動向等に係る調査 [7]」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)」をあげた人の割合とした．京都府では，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．ただし，回答エリアはそれぞれの地域メッシュと完全に一致しないことに注意が必要である．表 1 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．

4.2 手法と結果

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，人数と投稿数の代表値 (平均, 中央値, 合計, 最大値, 最小値) と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) について地域ごとのソーシャルセンサパフォーマンスとの相関を求めた．人数と投稿数に関しては，データを日毎に集計した．その後，それぞれのデータを平日と休日で分割し，代表値を算出した．

それぞれの代表値も地域特性を示すデータも全てデータ長は地域数 (=9) 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:

$$R' = \frac{\mathrm{cov}_k(SSP_{area}, a_k)}{\mathrm{std}_k(SSP_{area})\ \mathrm{std}_k(a_k)} = \frac{\mathrm{cov}_k(R_k, a_k)}{\mathrm{std}_k(R_k)\ \mathrm{std}_k(a_k)}$$

表 3 に示すように，観光魅力度と正の相関が居住者数とは負の相関が確認できた．

5 考　　察

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．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，SNS 上で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，SNS 上で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，人数と投稿数の代表値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

4：Overpass turbo `https://overpass-turbo.eu/`

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

6：`https://data.tourism.metro.tokyo.lg.jp/data/ishiki/`

7：`https://www.kyokanko.or.jp/wp/wp-content/uploads/kyoto_tourism_stat_2022.pdf`

表 3: ソーシャルセンサパフォーマンスとの相関： p 値 < 0.05 を太字で示す

	平日 (p 値)	休日 (p 値)
人数平均	0.24 (0.54)	0.25 (0.52)
人数中央値	0.23 (0.55)	0.23 (0.54)
人数合計	0.24 (0.54)	0.25 (0.53)
人数最大値	0.29 (0.45)	0.43 (0.25)
人数最小値	0.12 (0.75)	-0.16 (0.69)
投稿数平均	0.65 (0.06)	0.63 (0.07)
投稿数中央値	0.66 (0.05)	0.62 (0.07)
投稿数合計	0.65 (0.06)	0.62 (0.07)
投稿数最大値	0.66 (0.05)	0.64 (0.06)
投稿数最小値	0.64 (0.05)	0.58 (0.10)
コンビニ軒数	0.32 (0.40)	0.41 (0.28)
レストラン軒数	0.52 (0.15)	0.59 (0.09)
学校数	0.45 (0.23)	0.58 (0.10)
バス停数	0.40 (0.29)	0.35 (0.36)
ホテル軒数	0.58 (0.10)	0.46 (0.21)
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)**
最寄り駅からの距離	0.06 (0.87)	-0.11 (0.78)

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．

今後の課題は 3 つある．まず，投稿数が観測対象の規模に対応する事例を報告した先行研究との詳細な比較が必要である．例えば，投稿数と歩行者数の相関が報告されている [8]．加えて，投稿数と観光客の流れの相関も報告されている [9]．さらに，いくつかの先行研究では，投稿テキストに単語抽出, case-folding, 正規化, 単語の除去などの前処理が施されている [5], [10], [11]．一方，実験 1 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．観測対象の違いや，前処理の違いによる影響を十分に議論していく必要がある．

次に，実験 2 で用いた手法には改善の余地がある．実験 2 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．さらに，手法自体も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と改良の余地がある. 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手法を用いることで，より有効な分析ができる可能性がある．

最後に，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．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，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．

6 おわりに

本研究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．人数と投稿数のピアソンの相関係数に基づく，地域ごと, 時間ごとのソーシャルセンサパフォーマンスの分析によると，相関がある条件と無相関である条件が混在し，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常に成り立つわけではないという結論を得た．同時に第 4.2 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．

謝　辞

本研究は，JSPS 科研費 JP22H03648, JP22K12041, JST SICORP JPMJSC2107 の支援を受けたものである．

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. `https://aclanthology.org/D11-1145`.

[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] 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/\\LIPIcs.GIScience.2021.I.1`.

[9] 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.

[10] 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.

[11] 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.

appendix

表 4: 平日の時間ごとのソーシャルセンサパフォーマンス： 括弧内の値は p 値であり，p 値 < 0.05 を太字で示す

時間 (時)	0	1	2	3	4	5
(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)**
(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)
(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)
(b) 京都大学	-0.01 (0.93)	No post	-0.05 (0.40)	0.03 (0.65)	-0.11 (0.09)	**0.15 (0.02)**
(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)
(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)
(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)
(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)
(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)
時間 (時)	6	7	8	9	10	11
(a) 東京駅	**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	**0.30 (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	-0.04 (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	-0.04 (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	-0.02 (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	0.08 (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	0.09 (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	0.04 (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	0.08 (0.19)	0.10 (0.14)	**0.15 (0.02)**	0.06 (0.34)	0.09 (0.16)	0.04 (0.56)
時間 (時)	12	13	14	15	16	17
(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)**
(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)**
(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)
(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)
(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)
(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)
(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)**
(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)
(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)
時間 (時)	18	19	20	21	22	23
(a) 東京駅	**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	**0.61 (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	0.02 (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	0.06 (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	**0.16 (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	0.06 (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	**0.25 (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	**0.18 (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	0.12 (0.06)	**0.21 (0.00)**	0.10 (0.11)	0.05 (0.44)	-0.00 (0.97)	0.10 (0.11)

表 5: 休日の時間ごとのソーシャルセンサパフォーマンス： 括弧内の値は p 値であり，p 値 < 0.05 を太字で示す.

時間 (時)	0	1	2	3	4	5
(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)**
(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)**
(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)
(b) 京都大学	0.03(0.71)	-0.04(0.64)	**-0.20(0.03)**	-0.04(0.67)	No post	-0.08(0.37)
(e) 嵐山	-0.05(0.59)	0.06(0.48)	0.17(0.07)	No post	-0.00(0.96)	0.09(0.34)
(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 post
(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)
(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)
(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)
時間 (時)	6	7	8	9	10	11
(a) 東京駅	**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	**0.35(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	-0.04(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	0.01(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	0.03(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	-0.02(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	0.11(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	-0.07(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	0.18(0.05)	0.08(0.38)	0.08(0.38)	**0.23(0.01)**	0.16(0.08)	0.13(0.16)
時間 (時)	12	13	14	15	16	17
(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)**
(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)**
(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)
(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)
(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)
(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)
(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)**
(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)
(i) 当時	**0.19(0.04)**	0.06(0.50)	0.01(0.91)	-0.09(0.33)	0.07(0.47)	0.13(0.16)
時間 (時)	18	19	20	21	22	23
(a) 東京駅	**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	**0.45(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	0.15(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	-0.10(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	**0.32(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	-0.17(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	**0.68(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	0.17(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	**0.20(0.03)**	**0.36(0.00)**	**0.28(0.00)**	0.04(0.65)	-0.09(0.33)	**0.26(0.00)**