

# 空間情報ビッグデータによる 地理的な感染要因の推定

水野 貴之, 上坂 大輔, 幡 容子, 南川 敦宣

携帯電話の位置情報から、人々の居住地と滞在地を結ぶ人流ネットワークを構築し、グラフ埋め込みの一種である DeepWalk を用いて、各地の生活圏を表す特徴量を抽出する。市区町村別の感染者数を、人口あたりの各種法人施設数や住宅数、人口、生活圏を表す特徴量から予測する Lasso モデルを構築する。予測に寄与する要因が、人口や娯楽・飲食施設数、外国人比率、労働力人口比率であることが見出せる。多くの感染要因の候補から、感染者数の予測に寄与する主要因を Lasso モデルで絞り込むことで、主要因を鍵にして、人々が感染に因果関係のある要因を推定することが可能になる。

キーワード：COVID-19, 複雑ネットワーク科学, 機械学習, 位置情報

## 1. はじめに

2019 年末に中国の武漢で最初に確認された新型コロナウイルス感染症は、またたく間に、世界中に広がり、現在 (2021 年 1 月 3 日) までに、全世界で 8,422 万人以上が感染し、約 183 万人の命が失われている。感染者数の予測には短期的な流行過程を決定論的に記述する SIR タイプの感染症の数理モデルが数多く使われている [1–12]。SIR タイプのモデルはシンプルであり、主に感染症に対して免疫をもたない者、感染症が潜伏期間中の者、発症者、感染症から回復し免疫を獲得した者などについての、それぞれの人数のダイナミクスを、感染症の感染率や発症率、回復率などをパラメータにして記述している。モデルのパラメータは、人口や日々の新規感染者数などを用いて推計される。推計されたパラメータを用いて、モデルから 1 人の感染者が平均何人の感染者を生むかを表す実効再生産数を算出することができ、日々の実効再生産数をモニタリングすることで、感染症対策の政策を実行するタイミングを決めたり、政策の効果を測ったりすることを科学的にサポートすることが可能になる。

SIR タイプのモデルは感染状況を知ることに役立

つが、感染要因はブラックボックス的にパラメータの中に入り込んでしまうために、ワクチンの摂取により免疫獲得者が増加するような直接的な対策以外に、どのような政策を実行するべきかを見出すことには使にくい。たとえば、飲食店の営業時間の短縮が効果的かは、仮定を交えて SIR モデルを拡張して作り込まなければわからない。

日本政府はこれまでに、全国一律の外出の自粛要請、飲食店の夜間営業の規制などを実行してきたが、これらは、全国的な感染者数の増加や、感染経路の聞き取りを反映した感染症対策であった。しかし、このような人間の気付きに頼る方法では、未知の感染要因の発見を遅らせ、対策が後手にまわる可能性がある。したがって、過去の感染状況から感染要因を検出することを科学的にサポートするシステムが必要である。このようなシステムを開発することにより、効率的な感染症対策が可能になり、過剰な感染対策による経済活動へのダメージを回避することができるであろう。

本研究では、地理的な感染要因の候補に、地域の人口や、人々の地域間の移動、飲食街などの地域の属性を考える。各地の感染状況を、これらの感染要因で予測する機械学習モデルを構築することにより、感染リスクの高い地理的な要因を絞り込む。機械学習による予測は、因果ではなく相関を利用するため、予測に寄与する要因は代理変数である可能性があるが、多くの要因から予測に寄与する主要因を絞り込み、主要因を鍵にして、感染に因果関係のある要因を、人々が効率的に推定することができるようになる。

以降の節の、2 節では、感染要因を推定するシステムを開発するために用いる空間情報ビッグデータについて解説する。3 節では、感染の空間的な広がり与人

みずの たかゆき  
国立情報学研究所  
〒 101-8430 東京都千代田区一ツ橋 2-1-2  
mizuno@nii.ac.jp  
みずの たかゆき, かみさか だいすけ, はた ようこ,  
みなみかわ あつり  
株式会社 KDDI 総合研究所  
〒 356-8502 埼玉県ふじみ野市大原 2-1-15  
みなみかわ あつり  
KDDI 株式会社  
〒 102-8460 東京都千代田区飯田橋 3-10-10

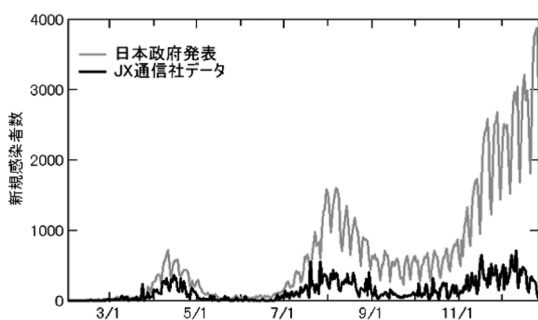


図1 日本政府公表と、感染場所の公開情報にもとづいて JX 通信社が収載した感染者数

流について述べる。人々の地域間の移動から生活圏を抽出し、感染の地域的な封じ込めについて議論する。4 節では、Lasso と LightGBM を用いた感染要因の推定システムを導入し、日本の 2020 年の春、夏、冬に発生した感染流行の要因を推計する。5 節では、空間解像度を上げて、地域に合わせた感染症対策を支援するシステムを開発するうえでの現状の課題について議論する。6 節はまとめと今後の展望である。

## 2. 空間情報ビッグデータ

2020 年 2 月 14 日から 12 月 27 日までの日々の感染者の発生場所の傾向を、JX 通信社が運営する速報通知サービス FASTALERT によって配信する新型コロナウイルス罹患事務所に関する情報を用いて把握する [13]。このデータには、日本国内における一般企業、病院、店舗、公的機関によるインターネット上への公開済み情報にもとづく新型コロナウイルス罹患患者の発生に関する情報を、施設名、住所、緯度・経度、感染者数、初回感染者報告日、施設内消毒などの対応の有無の発表状況、情報の最終更新日について収載されている。個人の住居を特定しうる情報、発生場所について特定不可能な情報については収載されていない。図 1 は、日本政府（厚生労働省）公表 [14] の新規感染者数と JX 通信社の感染者数（日付は初回感染者報告日）を示したものである。JX 通信社のデータには、経路不明や家庭内感染は含まれないため、収録されている感染者数は、日本政府が公表している感染者数の約 2 割程度となっているが、第一波、第二波、第三波は確認できる。

人々の移動については、利用許諾のある全国数百万人の匿名加工された au のスマートフォン位置情報データの 2020 年 4 月 24 日、7 月 3 日、11 月 13 日分を用いる。このデータは推定された居住地情報、および移動滞在の識別が可能な情報が付与された位置情報を、個人の履歴が追えない形に加工されたデータとなって

いる。

地域の人口（夜間人口）については、2015 年の国勢調査を用いて把握する [15]。また、昼間人口については、前述のスマートフォンの位置の 9:00 から 18:00 までの空間分布から把握する。

地域の属性については、ゼンリンが提供する 2016 年の座標付き電話帳 DB テレポイント法人版を用いて、法人施設の空間分布を把握する [16]。このデータでは、法人施設の住所とともに、各法人施設の業種番号が大分類で 39 分類付与（表 1）されている。個人の家屋や共同住宅といった、法人以外の施設の空間分布は、ゼンリンが提供する 2020 年の建物統計データを利用して把握する [17]。このデータには、住宅や商業ビル、オフィスビルなど全国約 3,800 万棟の建物について、基準地域メッシュ（500 m メッシュ）ごとに、建物種別などでの件数が集計・統計化して収録されており、この中から個人の家屋や共同住宅の件数を利用する。

## 3. 感染の空間的な広がり与人々の生活圏

流行初期において、感染リスクは全国一律ではなかった。図 2 は、47 都道府県の人口と日本政府公表の新規感染者数との相関係数を日々算出した結果である。2020 年 3 月 1 日の相関係数は 0.4 程度で、人口との相関は弱く、感染者は特定の地域（首都圏）に集中していた。しかしながら、人口と感染者数との相関は上昇を続け、年末には 0.95 に達し、各県の感染者は人口に比例した。人口あたりの感染リスクは、全国一律に近づいた。

感染の拡散は人々の移動と関係する。住宅街からの昼間の人々の流出人数を、2020 年 1 月を基準に比較したものが図 2 の Stay Home 指数である [18, 19]。ある  $t$  週の Stay Home 指数は、「(1 月の週平均での流出人数 -  $t$  週の流出人数) / 1 月の週平均での流出人数」で定義される。4 月の感染者数増とともに感染者数と人口との相関は上昇したが、Stay Home 指数の上昇とともに相関は減衰に転じた。しかし、Stay Home 指数が低下すると、再び、相関は上昇し続け、現在 (12 月) の全国的な蔓延状態になった。

人々の移動が感染を拡大させるのであれば、人々が生活や経済活動を行う行動範囲（生活圏）を知れば、感染リスクの高い地域や低い地域を見つけられる。生活圏間を分断するように地産地消で経済活動を行えば、経済損失を抑えたうえで効率的に感染拡大を防げるだろう。

利用許諾のある全国数百万人の匿名加工された au の

表 1 法人施設の 39 分類と住宅 2 分類 [16, 17]

番号	分類
1	水産・農林
2	鉱業
3	建設・土木
4	食品（食品製造など）
5	繊維（繊維工業など）
6	パルプ・紙（製紙工場など）
7	化学・薬品
8	石油・石炭製品
9	ゴム製品
10	窯業・ガラス
11	鉄鋼
12	非鉄金属
13	金属製品
14	機械
15	電気機器
16	輸送用機器
17	精密機器
18	その他製品
19	商業
20	金融保険
21	不動産
22	陸運（鉄道、地下鉄など）
23	海運（水運業など）
24	空運（航空運輸業など）
25	倉庫・運輸関連
26	通信・情報サービス
27	電気・ガス
28	技術・専門職（法律事務所、劇団など）
29	スポーツ施設
30	スポーツ用品店
31	娯楽・飲食施設（カラオケ、居酒屋など）
32	行楽地関連
33	病院・福祉
34	大型総合店舗（スーパー、百貨店など）
35	生活関連店（コンビニ、薬局など）
36	自動車関連 （ガソリンスタンド、カーディーラーなど）
37	学校・教養
38	役所・公共機関
39	その他
40	個人の家屋
41	共同住宅

スマートフォン位置情報データを用いて居住地と滞在先を結ぶ人流ネットワークを作成し、DeepWalkを用いて各地域の生活圏を表す特徴量ベクトルを算出する。

人々の1日の行動は、居住地から外出して居住地に戻ってくる。したがって、生活圏を表す人流ネットワークを得るために、各ユーザーの居住地と1時間ごとの滞在先をリンクで結ぶ、各ユーザーの居住地と滞在先のネットワークを、すべてのユーザーで合算すること

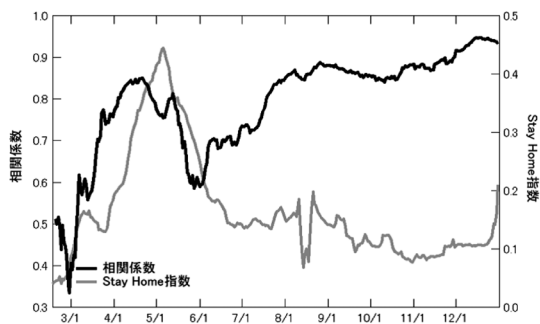


図 2 人口と新規感染者数との相関, Stay Home 指数

で、居住地と滞在地を結ぶ移動者の人数で重み付けられた二部グラフが得られる。居住地と滞在先の二分類をなくし、同じ地域を一つのノードにまとめることで、二部グラフを、地域間の移動を表す一部グラフ（人流ネットワーク）に変換する。このネットワークの各ノードは地域を表し、各地域からどの地域に流入が行われているかがネットワークで表現される。

DeepWalk とは、ランダム・ウォーカーがネットワーク上を移動して通過したノードの軌跡を、文章における単語の出現順に対応付けて、自然言語処理で行われる各単語の隣接単語の特徴量をニューラルネットワークで圧縮する Word2Vec を応用した手法である [20]。DeepWalk を用いて、人流ネットワークにおける各地域（ノード）の  $L$  リンク先までの隣接地域（隣接ノード）の情報を  $N$  次元まで集約した特徴量を得る。

DeepWalk により得られた特徴量を、直感的に理解するために可視化する。はじめに、人流ネットワークの各地域を日本の市区町村単位で設定し、各地域の  $L = 3$  リンク先までの隣接地域の情報を  $N = 32$  次元まで集約する。各地域の 32 次元ベクトルを、 $k$ -means の逐次繰り返しと BIC による分割停止基準によりクラスター数  $k$  を自動的に決定する  $x$ -means [21] を用いてクラスタリングすると、クラスター数は 23 個となり、図 3 のように各市区町村をグレースケールのポイントで表すことができる。大きなクラスターに沿って模式的に分割線を併記した。クラスターは都道府県単位で見た経済圏・生活圏に定性的に一致しており、人々がクラスターの中で、まとまって経済・社会活動を行っていることがわかる。

クラスターは階層構造をもっており、高解像度で人流ネットワークを構築することで狭域のクラスターを検出することができる。図 4 は都内を 500 m 四方（メッシュ）に区切り、それらをノードにした人流ネットワークから DeepWalk と  $x$ -means を用いてクラスターを

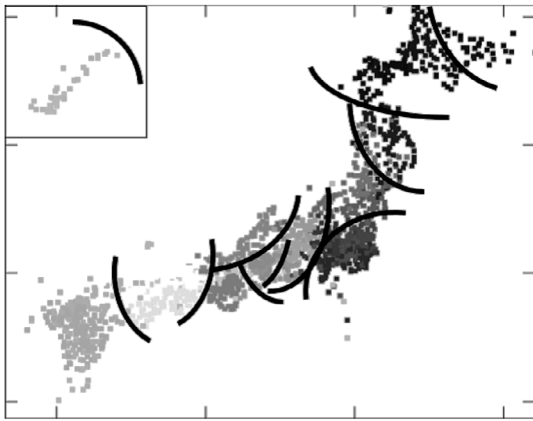


図3 11月13日の人流ネットワークのクラスター（市区町村単位）

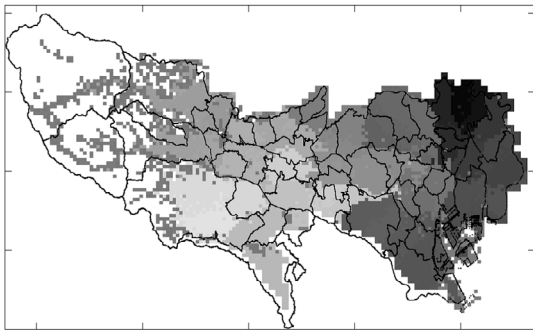


図4 11月13日の都内の人流ネットワークのクラスター（500m基準メッシュ単位）

検出し、各メッシュをグレースケールのポイントで表した結果である。クラスター数は48個となり、そのうち、100ノード以上含む大きなクラスターは28個である。都内の人々の生活圏のまとまりが可視化できていることがわかる。

市区町村を跨ぐ移動や500m四方を跨ぐ移動の内、図3や図4で確認された生活圏を跨ぐ移動は15%程度である。つまり、人々は生活圏の中で経済活動を密に行っている。したがって、同一の生活圏を分断させるような移動自粛要請は、経済損失が大きく実効性が乏しい。人々が生活圏を跨いで移動しないように気をつけることで、地域経済を守り、そのうえで、感染の封じ込めや感染の拡大が防止できるであろう。

#### 4. 地理的な感染要因の推定

感染リスクの地理的な要因としては、人口が多く感染者と接する確率が高くなる場所での滞在、マスクを外すなど感染が起きやすい状況が発生しやすい施設の利用、感染が蔓延している地域での滞在などが考えら

れる。JX通信社のデータを用いて、全国1,741市区町村について、感染施設が判明している感染者を数え、各市区町村の感染者数を地理的な要因で予測する機械学習モデルを構築することで、主要な感染要因を推定する。

各市区町村の業種分類別の物件数、人口、年齢別や性別の人口、外国人人口、3節のDeepWalkにより作成した各地の生活圏の特徴量を表す32次元の人流ベクトルを、地理的な感染要因の候補となる説明変数に採用する。人口を除き、各件数は人口1人あたりに規格化する。すべての要因候補の説明変数を標準化したうえで、1,741市区町村の各感染者数をL1正則化項を加えた線形重回帰モデルLassoおよび、決定木アルゴリズムに基づいた勾配ブースティングを行うLightGBMで予測する。ハイパーパラメータは、Lassoの正則化については0.05に、LightGBMはPython PackageのLightGBM [22]のデフォルト値を用いた。学習データとテストデータは、1,741市区町村を4:1にランダムに分割し、10セット実行した。

第一波（2020年1月-6月）、第二波（7月-9月）、第三波（10月-12月）について、テストデータで評価された感染者数の実数と予測値との相関は表2で示される。LassoおよびLightGBMともに、0.55を超える高い予測精度が実現している。LightGBMと比較して、解釈性の高いLassoでも遜色のない予測ができていたため、Lassoの係数を用いて、説明変数の感染者に関する寄与を議論する。

Lassoの変数選択と各変数の係数から、感染者数の予測に最も寄与する地理的な要因が人口であることが読み取れた。人口の係数は、第一波→第二波→第三波で、0.32→0.39→0.48と強くなっており、時間とともに感染が全国に拡大したことを表す図2の結果と整合的である。

表3は、Lassoにおける感染者数の予測に寄与する上位10変数である。39業種の中では人口あたりの娯楽施設・飲食店数が多い地域ほど感染者が多くなることがわかる。一方で、自動車関連施設（ガソリンスタンドやカーディーラー）では感染者はあまり出ていないことがわかる。この結果は、米国での先行研究と整合的である [23]。個人の家屋が感染者数にマイナスで寄与するのは、JX通信社のデータには、個人情報保護のために個人の住居を特定しうる情報、つまり、個人の家屋における感染の情報を収載していないからである。

表3の人口に関連する要因を見ると、外国人や労働



表 2 市区町村での感染者数の実数と予測値の相関係数

期間	Lasso	LightGBM
第一波	0.55 ± 0.003	0.55 ± 0.07
第二波	0.60 ± 0.004	0.69 ± 0.06
第三波	0.78 ± 0.0012	0.61 ± 0.04

表 3 市区町村での感染者数へ寄与する地理的な要因の Lasso 係数 (第二波). 人口以外の要因は人口あたりに規格化

順位	要因	係数
1	人口	0.392
2	娯楽・飲食施設	0.113
3	個人の家屋	-0.105
4	外国人人口	0.091
5	自動車関連施設	-0.036
6	人流ベクトル 25	0.025
7	20-65 歳人口	0.023
8	人流ベクトル 13	-0.022
9	15-20 歳人口	-0.020
10	人流ベクトル 21	-0.012

生産人口の割合が高い地域ほど感染者が多くなるとわかる。一方で、20 歳以下の人口の割合が高い地域は感染者数にマイナスで寄与しており、子供は感染者になりにくい可能性を示唆している。

第二波で有意に働いた人流ベクトルを Lasso の係数で重みづけして足し合わせたベクトルの成分は、北陸および東北でマイナスの値を示した。つまり、これらの地域は、人口や施設状況を加味しても感染者が少ない状況であったことを意味する。

## 5. 高い空間解像度での要因推定における課題

諸外国に比べて、日本は感染者が少ないため空間解像度を上げると個別要因によるノイズが強くなってしまい予測精度が下がる。ここでは、感染者が多い東京都を分析事例として、空間解像度を上げた場合の、感染要因の推定の困難さと今後の課題について述べる。

東京都を 500 m 四方の基準メッシュで分割し、9:00 から 18:00 までに平均して 1,000 人以上いる約 3,700 メッシュを分析対象とする。JX 通信社では個人情報保護の観点から住宅における感染の情報を収録していないため、500 m 四方まで解像度を上げると、住宅地において感染者数と人口との相関が無相関になる。これにより、都内全体での昼間人口と感染者数との相関は、第一波、第二波、第三波で、0.055、0.061、0.077 と低くなる。つまり、解像度を高めると、誰の家で感染者が出たかという個人情報保護の問題が現れるために、都市計画法に基づいて指定される用途地域での住居地域

表 4 都内 500 m 四方での感染者数の実数と予測値の相関係数

期間	Lasso	LightGBM
第一波	0.23 ± 0.04	0.14 ± 0.05
第二波	0.29 ± 0.01	0.27 ± 0.07
第三波	0.11 ± 0.008	0.13 ± 0.03

を除いたり、人口を直接的に説明変数に採用しないなどの対策が必要である。本稿では、人口を説明変数として採用しない。

500 m メッシュでの感染者数を目的変数にセットし、不動産を除いた各法人施設数、住宅数、32 次元の人流ベクトルを、人口あたりには変換せずに説明変数にセットして、目的変数を Lasso および LightGBM で予測した。ハイパーパラメータや機械学習での条件は、市区町村単位での予測モデルと同じにした。

第一波 (2020 年 1 月-6 月)、第二波 (7 月-9 月)、第三波 (10 月-12 月) について、テストデータで評価された感染者数の実数と予測値の相関は表 4 で示される。現状では、空間解像度に対して各メッシュの感染者数が少なく、Lasso と LightGBM とともに、相関係数は 0.1 から 0.3 程度であり、市区町村の解像度で感染者数を予測したモデルと比べて予測精度が低くなっている。つまり、説明変数の Lasso の係数や LightGBM の重要度による感染者数の予測に寄与する要因は、参考値程度になることに留意が必要である。

Lasso の係数から、感染者数に対して正に働く要因を探すと、娯楽・飲食施設、大型総合店舗、技術・専門職 (法律事務所、劇団など) が入ってくるのがわかった。これらは妥当な結果ではあるが、現在続いている第三波による多数の感染者の情報が活用できるようになれば、日々、予測精度は高まり、より多様な業種分類を説明変数に採用することが可能になるだろう。しかし、一方で、図 1 で見られるように、一般企業、病院、店舗、公的機関によるインターネット上への感染者発生情報の公開率が下がってきている。これは、感染者の増加にともない、各自の情報公開の業務負担が過大になっていることを反映している。感染施設と消毒状況とを手間なく迅速に機械可読が可能な状態で各法人が公開できる共通システムの構築、そのようなシステムにより集められた大量の情報による感染要因の推定での空間解像度の向上、多様な説明変数の採用は、今後の課題となる。

## 6. まとめと展望

本稿では、過去の感染状況から感染要因を推定することを科学的にサポートするシステムを開発するためのわれわれの取り組みについて紹介した。新型コロナウイルスの感染は、主に会話や咳やくしゃみによる飛沫感染であり、人々が移動してさまざまな場所で接触することで感染が広がっていく。われわれは、人々の居住地と滞在地を結ぶ人流ネットワークを構築し、グラフ埋め込みの1種である DeepWalk を用いて、各地の人々の生活圏の特徴量を抽出した。人々の生活圏は階層構造をもっており、人流ネットワークの粗視化の規模を変えることにより、各階層に合わせた生活圏を抽出することができる。市区町村間を跨ぐ移動の人流ネットワークでは、関東地方、中部地方などの広域の生活圏が取り出せる。また、空間を 500 m 四方に区切り、各 500 m メッシュ間を跨ぐ移動の人流ネットワークでは、市区町村に近いが、交流の多い市区町村は繋がるような狭域の生活圏が抽出できる。生活圏を跨ぐ移動は、生活圏内の移動に比べて少なく、人々は生活圏を単位に社会活動を行っている。したがって、同一の生活圏を分断させるような移動自粛要請は、経済損失が大きく実効性が乏しいであろう。人々が生活圏を跨いで移動しないように気をつけることで、地域経済を守り、そのうえで、感染の封じ込めや感染の拡大が防止できるであろう。われわれは、各地域の感染者数を、人口や、人口あたりの各種法人施設の数や住宅の数、生活圏を表す特徴量から予測する Lasso モデルを構築することにより、感染者増に相関する要因が、地域の人口や娯楽・飲食施設数、外国人比率、労働力人口比率であることを抽出した。この結果は先行研究で言及されている結果と整合的である。機械学習による予測は、因果ではなく相関を利用するため、予測に寄与する要因は代理変数である可能性がある。したがって、相関のある要因を鍵にして、感染に因果関係のある因果を特定することが重要である。

2020年1月8日時点で、日本における新型コロナウイルスの新規感染者は大幅な増加傾向にある。それにともない、一般企業、病院、店舗、公的機関などによる各自の感染発生と消毒状況の情報整理と公開に関する業務負担が過大になっており、手間なく迅速に情報登録と公開ができる共通システムの構築が望まれる。このような共通システムが構築されれば、今後、感染者に関するデータが増加し、感染要因を推定するモデルの精度が高まることが期待できる。そうなれば、現

在の 39 業種分類ではなく、多様な業種分類を利用することができ、対象を絞った感染症対策ができるであろう。また、コロナ後に、感染症対策を振り返るうえでも貴重なデータが失われずにすむ。本技術を活用することで、感染予防と経済活動が両立する社会が到来することを望む。

**謝辞** 立教大学大学院人工知能科学研究科の大西立顕氏には本研究について有益なコメントを頂いた。本研究の一部は、東京大学 CSIS 共同研究 (No. 674) による成果である (利用データ：座標付き電話帳 DB テレポイント法人版 (ゼンリン提供))。水野は、本研究に関して情報・システム研究機構、および JSPS 科研費 (JP19K22852) の助成を受けた。

## 参考文献

- [1] J. Dehning, J. Zierenberg, F. P. Spitzner, M. Wibral, J. P. Neto, M. Wilczek and V. Priesemann, "Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions," *Science*, **369**, eabb9789, 2020.
- [2] A. L. Bertozzi, "The challenges of modeling and forecasting the spread of COVID-19," *PNAS*, **117**, pp. 16732–16738, 2020.
- [3] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. D. Filippo, A. D. Matteo and M. Colaneri, "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy," *Nature Medicine*, **26**, pp. 855–860, 2020.
- [4] A. Aleta, D. M-Corral, A. P. y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini Jr, S. Merler, A. Pentland, A. Vespignani, E. Moro and Y. Moreno, "Modeling the impact of testing, contact tracing and household quarantine on second waves of COVID-19," *Nature Human Behaviour*, **4**, pp. 964–971, 2020.
- [5] S. B. Bastos and D. O. Cajueiro, "Modeling and forecasting the early evolution of the Covid-19 pandemic in Brazil," *Scientific Reports*, **10**, 19457, 2020.
- [6] A. Radulescu, C. Williams and K. Cavanagh, "Management strategies in a SEIR-type model of COVID 19 community spread," *Scientific Reports*, **10**, 21256, 2020.
- [7] K. B. Law, K. M. Peariasamy, B. S. Gill, S. Singh, B. M. Sundram, K. Rajendran, S. C. Dass, Y. L. Lee, P. P. Goh, H. Ibrahim and N. H. Abdullah, "Tracking the early depleting transmission dynamics of COVID-19 with a time-varying SIR model," *Scientific Reports*, **10**, 21721, 2020.
- [8] Z. Liao, P. Lan, Z. Liao, Y. Zhang and S. Liu, "TW-SIR: Time-window based SIR for COVID-19 forecasts," *Scientific Reports*, **10**, 22454, 2020.
- [9] U. Nguemdjo, F. Meno, A. Dongfack and B. Ventelou, "Simulating the progression of the COVID-19 disease in Cameroon using SIR models," *PLoS ONE*, **15**, e0237832, 2020.
- [10] G. G. Katul, A. Mrad, S. Bonetti, G. Manoli and

- A. J. Parolari, “Global convergence of COVID-19 basic reproduction number and estimation from early-time SIR dynamics,” *PLoS ONE*, **15**, e0239800, 2020.
- [11] J. Li, Y. Wang, J. Wu, J.-W. Ai, H.-C. Zhang, M. Gamber, W. Li, W.-H. Zhang and W.-J. Sun, “Do stay at home orders and cloth face coverings control COVID-19 in New York city? Results from a SIER model based on real world data,” *Open Forum Infectious Diseases*, ofaa442, 2020.
- [12] Y.-C. Chen, P.-E. Lu, C.-S. Chang and T.-H. Liu, “A time-dependent SIR model for COVID-19 with undetectable infected persons,” *IEEE Transactions on Network Science and Engineering*, **7**, pp. 3279–3294, 2020.
- [13] JX 通信社, 「FASTALERT 新型コロナウイルスリアルタイム API (罹患施設情報)」, <https://anti-covid-19.faxwire.jp/> (2020 年 12 月 27 日閲覧)
- [14] 厚生労働省, 「国内の発生状況など」, <https://www.mhlw.go.jp/stf/covid-19/kokunainohasseijoukyou.html> (2021 年 1 月 10 日閲覧)
- [15] 総務省統計局, 「平成 27 年国勢調査」, <https://www.stat.go.jp/data/kokusei/2015/> (2021 年 1 月 10 日閲覧)
- [16] ゼンリン, 「緯度経度座標付き電話帳データベース - テレポイント Pack! 」, <https://www.zenrin.co.jp/product/category/gis/contents/telpt/index.html> (2021 年 1 月 10 日閲覧)
- [17] ゼンリン, 「建物統計データ」, <https://www.zenrin.co.jp/product/category/gis/contents/building-statistics/index.html> (2021 年 1 月 10 日閲覧)
- [18] 水野貴之, 大西立顕, 渡辺努, “流動人口ビッグデータによる外出の自粛率の見える化,” *人工知能*, **35**, pp. 667–672, 2020.
- [19] T. Mizuno, T. Ohnishi and T. Watanabe, “Visualizing social and behavior change due to the outbreak of COVID-19 using mobile phone location data,” *New Generation Computing* (投稿中) .
- [20] B. Perozzi, R. Al-Rfou and S. Skiena, “DeepWalk: Online learning of social representations,” *KDD’14*, pp. 701–710, 2014.
- [21] D. Pelleg and A. Moore, “X-means: Extending k-means with efficient estimation of the number of clusters,” In *Proceedings of the 17th International Conference on Machine Learning*, pp. 727–734, 2000.
- [22] GitHub, <https://github.com/microsoft/LightGBM> (2021 年 1 月 3 日閲覧)
- [23] S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky and J. Leskovec, “Mobility network models of COVID-19 explain inequities and inform reopening,” *Nature*, **589**, pp. 82–87, 2021.