

# 周辺エリアのPoI別時系列混雑情報を用いた メッシュの混雑予測

山田 理<sup>1,2</sup> 松田 裕貴<sup>1,2,3</sup> 諏訪 博彦<sup>1,2</sup> 安本 慶一<sup>1,2</sup>

**概要:** 近年流行している COVID-19 は世界中で感染者を増やし、社会的問題へと発展している。感染拡大を防ぐためには、混雑を避け、人との接触を抑えることが重要である。混雑を避けるためには、事前に混雑を把握できる必要があり、混雑度の推定においては群衆における行動の予測が不可欠である。行動予測は空間的特徴と時間的特徴の2つの要素からなり、2つの特徴を組み合わせた予測モデル構築が行われてきた。人の行動は、POI (Point of interest) 情報と密接に関連している。PoI は空間的特徴を表し、場所の特徴を表現する。既存のアプローチでは、PoI がもつ空間的特徴に着目し、さまざまな手法で予測モデルに組みこんでいる。しかし、PoI は空間的特徴だけではなく、時間的特徴も持つ。本論文においては、PoI が持つ時間的特徴に着目し、PoI 別時系列混雑度データを用いてメッシュ領域混雑度予測手法を提案する。評価の結果、PoI 別時系列混雑度情報を用いることで、用いない場合と比べて、メッシュ領域の混雑度（存在人数）の予測に関して、最大 23 % 予測誤差が減少した。また、PoI 別人数混雑度情報を用いることで、30 分先のメッシュ領域混雑度を最小誤差率 1.27% で予測可能であることを示した。

## Mesh congestion estimation through analyzing time series PoI congestion information in neighboring areas

Osamu Yamada<sup>1,2</sup> YUKI MATSUDA<sup>1,2,3</sup> HIROHIKO SUWA<sup>1,2</sup> KEIICHI YASUMOTO<sup>1,2</sup>

### 1. はじめに

近年、中国の武漢市にて発生した COVID-19 (以下、新型コロナウイルス) は、驚異的な早さと感染力で世界中の国々へ感染が拡大した。2021 年 4 月 29 日時点で世界中の感染者数は約 1.4 億人、死者数は 300 万人を超え、大きな社会問題へと浮上している [1]。新型コロナウイルスの感染拡大防止のため、各国でワクチンの研究開発、接種が進められているが、未だ収束の目処はたっていない。とりわけ日本においては、特措法に基づく蔓延拡大防止措置や緊急事態宣言で人流を抑え、感染拡大防止をはかっている。しかし、これらの施策により飲食店等では営業の自粛や、営業時間の短縮などといった対応がもとめられ、人々の暮らしにおいては移動が制限されるなど、日常生活に大きな支

障をきたしている。

この新型コロナウイルスは、感染した人の咳やくしゃみ、会話などに伴う飛沫感染が主な感染拡大の原因になっているため、人と人との間隔「ソーシャルディスタンス」をとることが感染拡大防止の手段として挙げられている。日本の厚生労働省は、ソーシャルディスタンスをとる新たな生活様式を推奨している。行動例として、「公園はすいた時間、場所を選ぶ」、「公共交通機関の利用に関して、混んでいる時間帯は避けて」、「3密(密集、密閉、密接)の回避」などを示している [2]。ソーシャルディスタンスをとりながら外出を行えるよう、人の行動を予測し、未来における混雑度を推定する仕組みが求められている。

現在、人混みを避ける仕組みとして、リアルタイムな空き・混雑がマップ上でわかる Web サービス「VACAN」 [3] や Yahoo「混雑度レーダー」 [4] などが存在する。これらのサービスは過去または現在の混雑度を把握することができる。しかし、人が行動を判断するためには、現在の混雑度ではなく未来の混雑度を把握できる必要がある。例えば、

<sup>1</sup> 奈良先端科学技術大学院大学, Nara Institute of Science and Technology

<sup>2</sup> 理化学研究所 革新知能統合研究センター (AIP), RIKEN, Center for Advanced Intelligence Project (AIP)

<sup>3</sup> 国立研究開発法人科学技術振興機構 さきがけ

歩いて 10 分かかかる店舗や施設に向かう場合、現在の混雑度よりも 10 分後という未来における混雑度を把握できる方が、行動を起こす判断をしやすい。

人の行動予測に関する研究においては、個人の移動軌跡の予測が盛んに研究されているが、このような個人を対象としたモデリング手法を都市全体に拡大し、群衆の動きを予測するのは困難である。各人の軌跡データを収集し、各人の予測モデルを構築し、これらの結果をもとに混雑度を推定するのは現実的ではなく、プライバシー侵害にもなりかねない。群衆における行動予測においては、空間的特徴を考慮した予測モデル構築が行われている。対象領域内にはさまざまな PoI (point of interest) が存在し、空間的特徴として用いられている。しかし、既存の手法においては領域内に存在する PoI の個数や PoI カテゴリー名に着目し、PoI の時系列情報 (どの PoI に、いつ、どれくらいの間、何人が滞在したか) は考慮されていない。店や施設カテゴリーによって、混む時間帯と混まない時間帯が存在するため、PoI の時系列情報考慮することが望ましい。

そこで、本研究においては人混みを避けた行動計画を支援することを目的に、過去におけるある地域の各メッシュごとの PoI 別時系列混雑情報から、未来におけるあるメッシュの混雑度を予測する手法を提案する。この手法では、ある地域をメッシュに分割し、人の行動ログから各メッシュごとに PoI 別人数を抽出する。その後、過去における予測対象メッシュとその周囲のメッシュの PoI 別人数をもとに予測対象メッシュの人数予測を行う。これにより、PoI の時系列情報が考慮され、より実社会に即した予測を行うことができる。と考える。

以降の章構成は以下の通りである。2 章では、人の行動、混雑度予測に関連した既存研究を概説すると共に、本提案の位置付けを明らかにする。3 章では、扱うデータセット、4 章ではその前処理について取り扱う。その後、5 章では、混雑度推定の提案手法を述べ、6 章では、機械学習手法を用いた混雑度推定の評価実験について述べる、7 章では、評価実験の判定結果について述べる。8 章では、7 章の結果についての議論を行い。最後に、本論文の結論と今後の展望について述べる。

## 2. 関連研究

携帯電話ユーザの移動パターンの調査によると、人の行動 (移動) の予測に関して予測可能性が 93 % であることがわかっている [5]。一般的に行動予測においては、空間的特徴と時間的特徴という 2 つの特徴を利用した予測が行われている。空間的特徴とは、予測地域における PoI 情報や道路情報といった地理的特徴であり、時間的特徴は、時間依存における人の行動の規則性である。都市における次の場所、次のタイムウィンドウにおける予測では、一般的なモデルとしてマルコフモデルが利用される。また、時系列解

析手法として機械学習や統計学的手法を用いて、潜在的な規則性を抽出する研究などが行われている。以下では、本研究と関係する既存研究として、個人における行動予測、群衆における行動予測、PoI と人の行動の関係について概説し、提案手法の意義を述べる。

### 2.1 PoI と人の行動の関係

Zeng らは、人々の動きと PoI によって特徴づけられる活動分布との関係を効果的にモデル化し、視覚化する「POI-mobility-signature」を提案した [6]。これにより、人の動きと PoI が非常に関連していることが実証された。また Yuan らは、地域間の人の移動と地域内の POI の両方を用いて、都市内の異なる機能を持つ地域を発見するフレームワーク (DRoF) を提案した [7]。この中で、PoI や人の行動軌跡のみを使用するよりも、2 つのデータを組み合わせることにより、高精度に地域の機能を特定している。

### 2.2 個人における行動予測

遠藤らは、従来のマルコフベースではなく RNN を使用した次のタイムウィンドウにおける遷移確率推定し、目的地候補におけるセ訪問確率を推定している [8]。Kong らは、空間的・時間的影響を LSTM に組み合わせた空間的・時間的長短記憶 (ST-LSTM) モデルを提案した [9]。数日から数ヶ月という長い時間をかけて訪問先を予測するのではなく、数分から数時間後という短時間での移動に焦点をあて予測を行なった。また、同様のモデル手法として Liu らは、RNN を拡張して Spatial Temporal Recurrent Neural Networks (ST-RNN) モデルを提案し、従来の RNN よりも大幅に行動予測が改善された [10]。これらの、研究においては、過去に訪れた場所の履歴を使って次の場所を予測することに着目している。一方、Sadari らは過去に訪れた場所から行動を予測するのではなく、1 日のユーザの軌跡に着目した [11]。午前中の軌跡とそれに対応する午後の軌跡の関係について分析を行い、両者の間に正の相関関係があることを発見した。この結果から、過去における午前中の行動の軌跡と現在の午前中の行動の軌跡の類似度を測ることで午後の軌跡を予測する手法を提案した。この手法において、LSTM との比較を行い、サンプル数が少ないほど効果的であることを示した。個人における行動予測においては、時系列的解析が行われており、空間的特徴と時系列的特徴を組み合わせたモデル構築が行われている。

### 2.3 群衆における行動予測

NTT 技術ジャーナルの「群集誘導のための人流予測技術」では、GPS を利用した位置情報データやイベント会場内に設置された位置センサによる人数カウントデータ等を入力とし、各時間に共通する、場の特徴を捉えたモデルを構築し、時系列パターンを得ることで、人流を予

測する取り組みを行っている [12]. この手法においては, BLE (Bluetooth Low Energy) や Wi-fi アクセスポイントログ, カメラによる人数情報等を使用しているため, ある特定の場所による行動予測となっている. Zang らは, Double-Encoder という手法を提案し, 空間的, 時間的特徴と日々の動きの相関性をモデル化した [13]. これは, 各地域が市民の規則的な生活に起因してほぼ毎日同じフローを持っており, いくつかの地域が似たようなフローパターンを共有し, 相関関係をもっていることに着目した手法である. 時空間依存性と日々のフローの相関性をそれぞれ捉えるための 2 つのエンコーダを構成し, モデル構築を行い, 対象領域における流入量・流出量の予測においてベースラインと比べ, 大きな優位性を示した. Jiang らは, CNN と LSTM を組み合わせた予測手法を提案した. これは, 人の活動と PoI 情報の関連に着目し, 人の軌跡データと都市の PoI データを入力とした [14]. 従来の方法では, 異なる 2 つのデータを入力として持つことが難しい. そこで, 予測地域をメッシュごとにくぎり, 各メッシュにおける PoI 情報を空間的特徴として, CNN の畳み込みを利用することで入力データとしてもたせた. また, この研究においては, 転移学習を用いてある都市から別の都市へと学習を移すことで, 限られたデータしか得られない対象都市に対して, 他の都市のデータを活用してより強力なモデルを構築した. 特に学習データが限られている場合に, ベースライン手法と比較し優れた性能を発揮した.

## 2.4 PoI 別人数予測サービス

PoI 別人数の予測においては, GoogleMap における「お店の混雑状況」がサービスとして存在する. これは, ロケーション履歴を有効にしているユーザから集計した匿名データをもとに混雑具合が割り出され, データが十分に集まっていれば, 店舗の混雑する時間帯が表示される. しかし, 統計的に算出したものとなっており, 異なる PoI が及ぼす影響や, 地域的な特徴が考慮されていない.

## 2.5 関連研究の課題および本研究の目的

個人における行動予測においては, マルコフモデルでの予測と, RNN 等の機械学習を用いた手法がおこなわれている. しかし, これらの研究における課題設定は, 個人における次のタイムウィンドウで, 目的候補地の中から一番遷移確率が高い場所を予測することを目的としている. また, 各個人のモデルを構築し結果を統合する必要がある, データ数の問題やプライバシーの問題, 学習モデル構築にかかる時間等を考えると現実的ではない. 群衆における行動予測においては, LSTM を用いた手法が多く行われており, CNN を用いるといった様々な工夫で空間的特徴を考慮している. しかし, 空間的特徴を PoI 情報を使用する場合, 地域特徴として PoI の個数や PoI カテゴリー名が利用

されている. しかし, PoI には空間的特徴のみではなく時間的特徴も持つ. 例えば, レストラン等の飲食店においては, 昼間や夕方の時間帯に人が集まるが, バーや飲み屋といった飲食店においては夜の時間帯に人が集まることは容易に想像がつく. このように, PoI の時間的特徴は混雑度を推定する上で重要な情報である. 本研究においては, PoI が示す空間的特徴だけではなく, 時系列的特徴を考慮することで予測精度の向上を図り, より実社会に即した予測モデル構築を目指す.

## 3. データセット

本研究では, PoI の時系列混雑情報を予測するためのデータセットとして, 軌跡データ, メッシュデータ, PoI データを組み合わせて作成した (図 1). 本章では使用した各データについて説明する.

### 3.1 軌跡データ

軌跡データは, 株式会社 Agoop の「ポイント型流動人口データ」を使用した [15]. これは, 株式会社 Agoop が提供するアプリケーション「Walkcion」「ラーメンチェッカー」と提携先のアプリケーションに Agoop の SDK を組み込み, ユーザの同意のもとスマートフォン, タブレット端末などの GPS 機能によって位置情報データを取得している. これらのデータは, 個人情報情報を隠蔽するために, 日々の位置情報を元に居住地エリアや勤務地エリアの推定を行ない, これらの推定結果を用いて居住地エリアや勤務地エリアに該当する位置情報データを削除 (秘匿化) している. また, 推定が不可能だった場合には, 株式会社 Agoop 独自に定義した「住宅エリア」に該当する位置情報を削除 (秘匿化) している. 使用するデータは, デイリー ID, GPS データ (緯度, 経度), 年月日, GPS 精度, 都道府県コード, 市町村コードを含んでいる (表 1). デイリー ID とは, ユーザ毎に付与する固有の ID となっており, 毎日 0 時にユーザ毎に新規の ID を付与している. そのため, 同日内において同一のデイリー ID をつなぐことにより, 移動経路を把握することができる. また, GPS 精度に関しては, GPS 測位の水平方向の精度を示すため正の緯度経度が取りうる範囲を円で表し, その半径をメートルで示す. 市町村コード, 都道府県コードは, 総務省が設定した全国地方公共団体コード (5 桁 JIS コード) を使用しており, 「都道府県コード 2 桁」+「市区町村コード 3 桁」で表されている.

### 3.2 メッシュデータ

メッシュデータには, 株式会社 Agoop の「メッシュ型流動人口データ」を使用し, メッシュ ID を付与した. このメッシュデータには, 各メッシュの最大緯度, 最大経度, 最小緯度, 最小経度のデータがあり, このデータをもとに, 軌跡データの緯度経度からメッシュ ID を判定する. 1km

表 1 使用するデータ例

種類	データ
デイリー ID	000021ebcf390c6e9194d42d0a6...
GPS データ	35.710390 139.88533
年月日	2020-07-01 08:00:00
GPS 精度	35.7103
都道府県コード	13 (東京)
市町村コード	131041 (新宿)

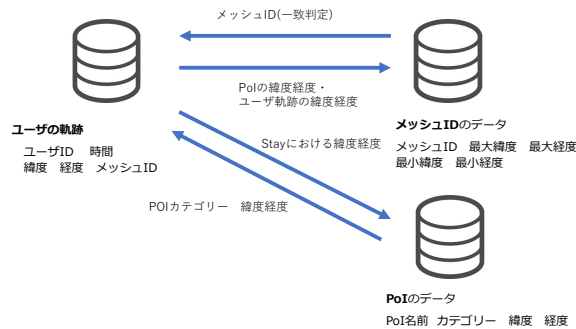


図 1 各データの使用方法

四方のメッシュまでは、標準地域メッシュによって定められたコードを用いる。標準地域メッシュとは、総務省によって出された「統計に用いる標準地域メッシュおよび標準メッシュ・コード」であり、第 1 次メッシュ (1 辺の長さが約 80km, 4 桁のコード)、第 2 次メッシュ (1 辺の長さが約 10km, 6 桁のコード)、第 3 次メッシュ (1 辺の長さが約 1km, 8 桁のコード) まで定められている。

今回実験においては、第 3 次メッシュを使用した。

### 3.3 PoI データ

株式会社 Agoop の東京における「POI データ」を利用した。東京都内の 456,450 件の PoI データである。このデータには、掲載名、業種大分類名、業種中分類名、業種小分類名、都道府県コード、緯度、経度情報などが含まれている。今回の実験においては業種中分類名と業種大分類名を使用した。この業種中分類においては、195 種類のカテゴリーに分類され、業種大分類においては、17 種類のカテゴリーに分類される。また、移動中のデータを「move」として PoI のカテゴリーに加えたため、それぞれの分類は 196 種類、18 種類となる (表 2)。

## 4. 前処理

前処理として、位置時系列データに対し、外れ値の除去、欠損値の補間、所属するメッシュの判定、移動・滞在の判定と滞在 PoI の決定を行う。以降、それぞれ詳細を述べる。本提案手法では、2 点間の緯度経度から距離を計算し、PoI 情報の付与、外れ値の計算を行う。ここでは、ユークリッド距離ではなく、地球を楕円体と考え計算を行うヒュベニの公式を利用した。以下の数式で表される。

$$d = \sqrt{(d_y M)^2 + (d_x N \cos \mu_y)^2} \quad (1)$$

$d_y$  子午線曲率半径,  $N$  卯酉線曲率半径,  $\mu_y$  二点の緯度の平均値,  $d_y, d_x$  2 点間の緯度と経度の差である。

### 4.1 外れ値の除去

軌跡データに含まれる GPS 精度を利用し、GPS 誤差が大きいデータを除去した。スマートフォン等のデバイスによる GPS 精度は、衛星の電波の遮断や、反射が起こる場所 (室内やビル街など) で悪くなる傾向があり、緯度経度が取りうる範囲の円の半径が大きすぎるデータはノイズとした。また、各ユーザーの時系列データをもとに、前後のデータの緯度経度からヒュベニの公式で距離を算出し、電車の速度以上 (分速 2km) で移動しているデータは除去した。

### 4.2 データ補間

今回は、10 分間隔のデータで実験を行なう。一部タイムウィンドウにおいてデータが欠けていたため、欠けている軌跡データに対し、データ補間を行なった。データの補間方法としては、前後の値から線形補間を行い、データを作成した。この際、生データに対し 1 分間のダウンサンプリング後、線形補間を行い、その後 10 分間のアップサンプリングによってデータフレームを作成した。アップサンプリングでは、平均値を使用した。

### 4.3 メッシュ判定

メッシュデータセットから、メッシュ ID の付与を行う。メッシュ ID データに含まれる最大緯度、最大経度、最小緯度、最小経度から成るメッシュに対し、対象となる緯度経度が含まれているかどうかを判断する。このとき、100m 間隔のメッシュ ID を付与した。

### 4.4 滞在判定・PoI 判定

滞在判定においては、一定の間幅のタイムウィンドウに、一定数以上の軌跡データがある場合、meanshift 法を適用し滞在判定を行う方法などが用いられる。しかし、今回は補間によってデータ数を増やしているため、タイムウィンドウあたりのデータ数の条件は満たすものとした。補間し、10 分ごとのアップサンプリングしたデータに対し、前後の緯度経度のデータからヒュベニの公式を用いて、距離を算出した。算出した距離が 800m 以上を 2 回続けた場合、「移動 (move)」と判定し、それ以外の場合、「滞在 (stay)」とした。800m という閾値は、徒歩における速度が 80m/分のため、10 分間隔のタイムウィンドウにおいて 2 地点間の距離が 800m 以上のときに移動と設定した。滞在 (stay) と判定されたデータに対しては、PoI データセットを用いて PoI を付与した。PoI データセットに含まれる

表 2 PoI 大分類

交通・運輸・倉庫	自動車・オートバイ・自転車・ドライブ	販売・卸
飲食店	スポーツ・趣味娯楽・レジャー	医療・薬・保険衛生
官公庁・団体・福祉	旅行・観光・温泉・旅館・ホテル	学校・趣味教室・図書館
その他サービス	不動産・賃貸・展示場	金融・保険・証券
製造・加工	建設・工事	農林・水産・鉱業
電気・ガス・通信・放送・新聞	出版・印刷	move

緯度経度情報と、軌跡データより得られた緯度経度情報をヒュベニの公式で距離を算出し、一番近い距離の PoI 情報を使用した。このとき、軌跡データと使用する PoI が同一メッシュ内のものでなければならない。そこでメッシュ判定により付与されたメッシュ ID をもとに、軌跡データの緯度経度によるメッシュ ID と PoI データによるメッシュ ID を照らし合わせ、整合性を確認した。このとき、一致しなかった軌跡データに関しては破棄した。また、続けて滞在 (stay) と判定されたデータに対しては、あらためて PoI 判定を行うのではなく、一つ前の滞在 (stay) 判定された PoI 情報を利用した。また、メッシュ ID に関しても、一つ前のメッシュ ID を利用した。移動 (move) と判定されたデータに対しては、PoI カテゴリ「move」を付与した。移動 (move) と判定されたすべてのデータに対して、メッシュ判定を行った。

#### 4.5 予測対象メッシュの決定

本研究においては、Agoop のメッシュ ID をもとに予測を行う。予測対象地域の設定においては、対象となる時間帯において、予測対象となるメッシュの周りにメッシュが存在し、データ量が一番多いメッシュを対象地域とする。

### 5. 提案手法

#### 5.1 提案手法における課題設定

既存の手法においては PoI を空間的特徴とし、PoI の時系列性を考慮していない。よって、PoI ごとの時系列混雑度を考慮し予測を行う。本研究では、以下の課題を設定し実験を行い、評価する。課題の概略を図 2 に示す。

課題：メッシュに区切られた対象地域において、過去の各メッシュの PoI 別人数から、未来における予測対象メッシュの混雑度を推定する

#### 5.2 PoI 別人数データを使用した予測手法

予測対象メッシュと周囲のメッシュ 8 個の PoI 別人数データを利用し、ランダムフォレスト (RF) と重回帰分析 (MR) によって予測モデルの構築を行う。各メッシュにおける PoI 別人数を入力とし、予測対象メッシュの未来における人数を正解データとする。予測モデルの検証に関しては交差検証をおこない、検証手法は k-分割交差検証 (kfold) と leave-one-out (LOO) の二つで評価を行なった。

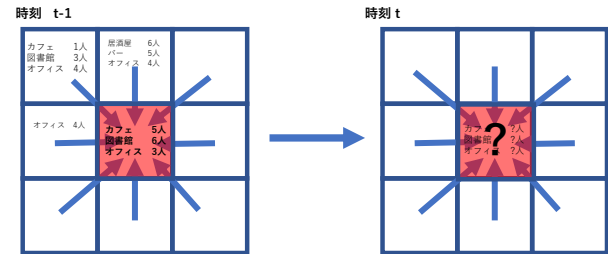


図 2 本研究で解決したい課題の概略図

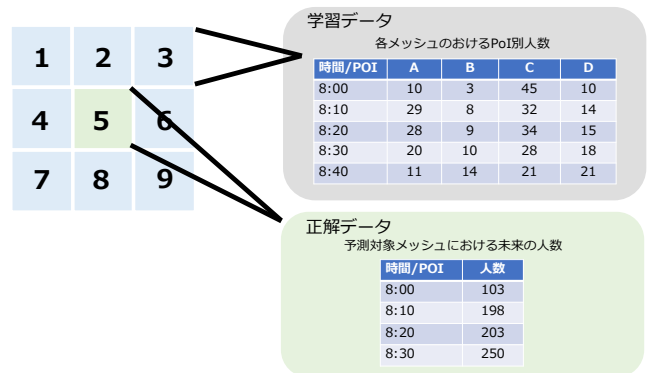


図 3 PoI 別人数情報を用いた予測手法

予測手法における概略図を図 3 に示す。

### 6. 実験

データ量が一番多いメッシュ ID 「53394621」(東京都千代田区) を予測対象メッシュとして設定する。対象地域を図 4 に示す。2020/7/1~2020/7/9 (土日である 2020/7/4、2020/7/5 は除く) までをデータとして使用し、以下の 4 種類のパラメータを変更し実験を行なった。

- 入力データの粒度
- 予測時間帯
- 予測先時間
- PoI の時系列情報

#### 6.1 入力データの粒度の違い

PoI のカテゴリ粒度の違いにおける予測精度の影響について実験を行なった。ベースラインとして、人数情報、滞在 (stay)・移動 (move) を用意し、PoI カテゴリ大分類 18 種類 (17 種類 + move) と中分類 196 種類 (195 種



表 3 入力データの粒度による予測誤差 (RMSE, 単位:人)

MODEL	人数		滞在・移動		PoI 大分類		PoI 小分類	
	kfold	LOO	kfold	LOO	kfold	LOO	kfold	LOO
MR	14.03	11.02	12.81	<b>9.86</b>	33.12	35.06	16.15	13.10
RF	16.64	13.88	19.24	14.88	15.58	<b>11.18</b>	18.7	13.0

表 4 入力データの粒度による予測誤差 (MAPE, 単位:%)

MODEL	人数		滞在・移動		PoI 大分類		PoI 小分類	
	kfold	LOO	kfold	LOO	kfold	LOO	kfold	LOO
MR	1.22	1.23	1.13	<b>1.11</b>	2.86	3.92	1.42	1.53
RF	1.48	1.53	1.62	1.63	1.38	<b>1.31</b>	1.69	1.54

類+ move) の4つで比較を行なった。朝における時間帯で検証をおこなった。

## 6.2 予測時間帯

人の行動には時間的特徴が大きく影響する。例えば、朝の時間帯において通勤・通学の移動が多く、昼の時間帯においては仕事や学校、ショッピング等で滞在することが多くなることが容易に想像つく。そこで、時間帯の違いによる PoI 別人数データが及ぼす予測精度の影響について実験を行なった。今回の実験では、朝の時間帯を8時から12時、昼の時間帯を13時から17時、夜の時間帯を18時から22時と設定し、PoI の分類は大分類、予測先時間(正解データ)は10分後を使用した。各時間帯において学習を行い、各時間帯において予測を行なった。各時間帯におけるデータセット数は表 11 に示す。

## 6.3 予測先時間

本研究が提案する混雑度予測手法において、正解ラベルに対象メッシュの未来における人数をあたえる。この実験において、現在の PoI 情報をもとに未来における混雑度の予測限界を検証する予測先時間を10分後、20分後、30分後に設定し、正解ラベルとして使用した。PoI の分類は大分類を使用し、朝の時間帯を使用した。

## 6.4 PoI の時系列情報

PoI の時系列情報を考慮するため、ラグ特徴量を使用した学習データを作成した。ラグ特徴量とは、機械学習のモデル構築手法において、時系列的特徴を考慮するために、過去のデータを入力に加えたものである。過去の PoI 情報がどれだけ有効であるかを測るとともに、何分前の情報が予測において重要であるかを検証する。ラグ特徴量を用いることで、本研究の目的である PoI の時系列要素が考慮される。今回の実験においては、PoI の分類は大分類、朝の時間に設定し、「10分前」、「10分前と20分前」、「10分前と20分前と30分前」の3種類で実験を行なった。

## 7. 結果

予測モデルの評価指標は、平均絶対誤差率 (MAPE, Mean Absolute Percentage Error) と平均平方二乗誤差 (RMSE,

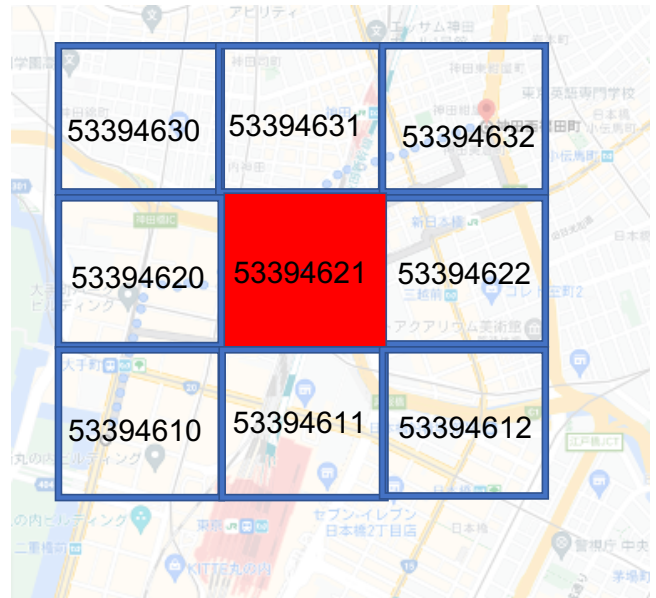


図 4 実験における対象地域

表 5 予測時間帯による予測誤差 (RMSE, 単位:人)

MODEL	朝		昼		夜	
	kfold	LOO	kfold	LOO	kfold	LOO
MR	33.12	35.06	22.23	<b>17.37</b>	23.72	25.72
RF	15.58	11.19	13.08	<b>9.26</b>	17.34	14.52

表 6 予測時間帯による予測誤差 (MAPE, 単位:%)

MODEL	朝		昼		夜	
	kfold	LOO	kfold	LOO	kfold	LOO
MR	2.86	3.92	1.54	<b>1.51</b>	4.09	5.75
RF	1.38	1.31	0.87	<b>0.81</b>	3.36	3.38

表 7 予測先時間による予測精度 (RMSE)

MODEL	10分後		20分後		30分後	
	kfold	LOO	kfold	LOO	kfold	LOO
MR	33.12	35.06	30.41	<b>18.81</b>	35.39	46.10
RF	15.58	<b>11.19</b>	16.45	11.75	16.56	12.05

表 8 予測先時間による予測精度 (MAPE)

MODEL	10分後		20分後		30分後	
	kfold	LOO	kfold	LOO	kfold	LOO
MR	2.86	3.92	2.55	<b>1.98</b>	2.84	4.61
RF	1.39	1.31	1.38	1.30	1.34	<b>1.27</b>

Root Mean Square Error) を使用した。

入力データの違いにおいて、4種類のデータを入力したときの結果を表 3, 表 4 に示す。滞在・移動情報を用いた重回帰分析が最も低い誤差であったが、ランダムフォレストによるモデル構築においては、PoI 大分類によるデータがベースラインよりも低い誤差であった。朝、昼、夜の3つの時間帯において、対象時間帯の PoI 別人数データを用いてモデル構築、予測を行なった場合の結果を表 5, 表 6 に示す。どちらの機械学習手法においても昼の時間帯が最も高い精度が得られた。予測先時間の違いにおける結果を表 7, 表 8 に示す。重回帰分析においては、RMSE の指

表 9 PoI の時系列情報による予測誤差 (RMSE, 単位: 人)

入力データ	なし		10 分前		10 分前~20 分前		10 分前~30 分前	
	kfold	LOO	kfold	LOO	kfold	LOO	kfold	LOO
MR	33.12	35.06	19.54	11.74	16.90	<b>11.47</b>	17.32	12.98
RF	15.58	11.19	15.19	10.64	13.95	<b>10.56</b>	15.94	11.66

表 10 PoI の時系列情報による予測誤差 (MAPE, 単位: %)

入力データ	なし		10 分前		10 分前~20 分前		10 分前~30 分前	
	kfold	LOO	kfold	LOO	kfold	LOO	kfold	LOO
MR	2.86	3.92	1.64	1.39	1.32	1.49	1.54	1.52
RF	1.39	1.31	1.35	1.23	1.23	1.22	1.33	1.37

表 11 朝・昼・夜におけるデータセット数

時間帯	データセット数
朝	20000645
昼	22977821
夜	17499918

標においては 10 分後が最も高い精度であり, MAPE においては 30 分後の予測が高い精度であった. このことから, 現在の PoI 別人数データの予測先限界は 30 分後ではないことが示された. また, 表 8 より, PoI 情報を考慮することで 30 分先の時間帯においても, 最小予測誤差率 1.27% で予測可能であった. 予測先時間において, 10 分後の予測よりも 20 分後の予測の方が精度が高かったことから, 人の短時間的行動において, 20 分前の行動情報が有用であることを示した PoI の時系列情報による予測結果を表 9, 表 10 に示し, 予測値と真値のグラフを図 5 に示す. 10 分前+20 分前のデータが最も高い精度を得たことから, 20 分前のデータの有用性を改めて示した. 重回帰分析においては, 時系列情報が予測精度に大きく寄与することが示された. また, 図 5 より, 予測値が真値に追いつき予測できていることがわかる.

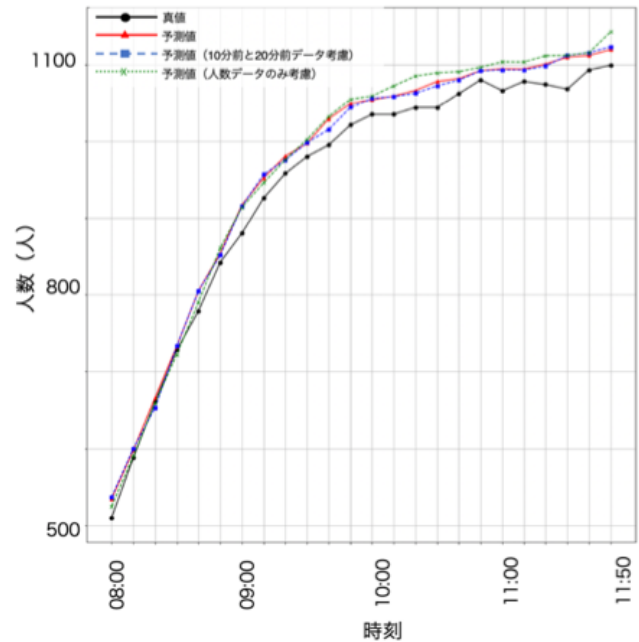


図 5 PoI の時系列情報による予測値と真値

## 8. 考察

### 8.1 時系列 PoI データの予測可能性

ベースラインとの比較において, PoI 別混雑度情報が行動予測において重要な特徴量であることが示された. 入力データにの情報の粒度によって予測精度が大きく変わることがわかる. また, ランダムフォレストの PoI の小分類において, 大分類よりも精度が下がった. これは, カテゴリーの種類が多く, 各 PoI がもつ人数情報が少なくなってしまうが故に予測誤差が大きく出てしまったと考えられる. 予測時間帯の比較においては昼がもっとも高い精度が出たが, 表 11 に示すように, データセット量の多さに起因すると考えられる.

### 8.2 展望

今回の実験においては統計手法である重回帰分析と機械学習であるランダムフォレストの 2 つを用いて予測を行

なった. 行動予測においては, より時系列を考慮する RNN や LSTM といった手法が多く用いられる. 今回のデータセットは 1 週間分のデータであったが, 数ヶ月のデータセットにおいて時系列手法による PoI 別混雑度情報が有効であるかを検証したいと考えている, また, Sadari らの実験において, 人の行動の周期性が実証された, この特徴を考慮した上で, 群衆における PoI の行動周期性についても検証したいと考えている. 今回の実験においては, 予測対象地域を予測対象メッシュと周りのメッシュ 8 つを学習データとして使用したが, 人の長距離的移動についても考慮する必要がある. 実験で設定した大手町周辺においては, 電車による移動も多いことから, 3km メッシュ四方のみの予測モデル構築は不十分であると考え. 予測対象メッシュの拡大や, メッシュごとにモデルを構築し, そのモデルの情報を交換することでより高精度なモデル手法を検討したいと考えている.

## 9. おわりに

本研究では、PoI における時系列データに着目し、PoI 別混雑度情報を用いた予測モデル構築について提案した。4 つの実験の結果、PoI 別時系列混雑度データを用いることで、用いない場合と比べて、最大 23 % 予測誤差が減少することが示された。また、PoI 別混雑度情報と予測先時間、予測時間帯において関連性がみられ、現在の PoI 別人数情報から 30 分先の混雑度を最小予測誤差率 1.27% で予測可能であることを示した。これにより、PoI 別混雑度情報が行動予測において大変有用であることが示された。今後は、この結果からさまざまな既存の行動予測における手法に適用させ、混雑を避け行動するための仕組み作りに寄与したいと考えている。

### 参考文献

- [1] World Health Organization. 新型コロナウイルス感染症 (covid-19) who 公式情報特設ページ. [https://extranet.who.int/kobe\\_centre/ja/covid](https://extranet.who.int/kobe_centre/ja/covid). Accessed: 2021-04-29.
- [2] 厚生労働省. 新たな生活様式. [https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000121431\\_newlifestyle.html](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000121431_newlifestyle.html). Accessed: 2020-07-14.
- [3] Vacan. <https://corp.vacan.com/>. Accessed: 2021-04-29.
- [4] Yahoo!map. <https://map.yahoo.co.jp/promo/>. Accessed: 2021-04-29.
- [5] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, Vol. 327, No. 5968, pp. 1018–1021, 2010.
- [6] Wei Zeng, Chi-Wing Fu, Stefan Müller Arisona, Simon Schubiger, Remo Burkhard, and Kwan-Liu Ma. Visualizing the relationship between human mobility and points of interest. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 18, No. 8, pp. 2271–2284, 2017.
- [7] Jing Yuan, Yu Zheng, and Xing Xie. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 186–194, 2012.
- [8] Yuki Endo, Kyosuke Nishida, Hiroyuki Toda, and Hiroshi Sawada. Predicting destinations from partial trajectories using recurrent neural network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 160–172. Springer, 2017.
- [9] Dejiang Kong and Fei Wu. Hst-lstm: A hierarchical spatial-temporal long-short term memory network for location prediction. In *IJCAI*, Vol. 18, pp. 2341–2347, 2018.
- [10] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. Predicting the next location: A recurrent model with spatial and temporal contexts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30, 2016.
- [11] Amin Sadri, Flora D Salim, Yongli Ren, Wei Shao, John C Krumm, and Cecilia Mascolo. What will you do for the rest of the day? 2018.
- [12] 群集誘導のための人流予測技術. <https://www.ntt.co.jp/journal/1806/files/JN20180638.pdf>. Accessed: 2021-04-29.
- [13] Tianzi Zang, Yanmin Zhu, Yanan Xu, and Jiadi Yu. Jointly modeling spatio-temporal dependencies and daily flow correlations for crowd flow prediction. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, Vol. 15, No. 4, pp. 1–20, 2021.
- [14] Renhe Jiang, Xuan Song, Zipei Fan, Tianqi Xia, Zhaonan Wang, Qunjun Chen, Zekun Cai, and Ryosuke Shibasaki. Transfer urban human mobility via poi embedding over multiple cities. *ACM Transactions on Data Science*, Vol. 2, No. 1, pp. 1–26, 2021.
- [15] Agoop サービス製品. <https://www.agoop.co.jp/service/>. Accessed: 2021-04-29.