

PoI 別時系列混雑度情報を用いたメッシュにおける混雑度予測の検証

山田 理^{1,2} 松田 裕貴^{1,2,3} 諏訪 博彦^{1,2} 安本 慶一^{1,2}

概要: 近年流行している COVID-19 は世界中で感染者を増やし、深刻な問題を引き起こしている。感染拡大を防ぐ手段として、混雑を避け、人との接触を抑えることが挙げられる。混雑を避けるためには、人（群衆）における行動を予測し、事前に混雑を把握する必要がある。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、これらの特徴を組み合わせ、より精度の高い予測モデルが構築されてきた。その中でも空間的特徴として、人の行動と密接に関連し、場所の特徴を表現する PoI (Point of interest) 情報を用いた手法が提案されている。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 別混雑度情報を用いることで、人数データのみで学習し予測した場合と比べて、3時間後におけるメッシュ領域の混雑度（存在人数）の予測に関して、最大50%予測誤差が減少し、最小誤差率約5%で予測可能であることを示した。また、PoI 別時系列混雑度情報を用いることで時系列を考慮しない場合と比べ、精度の向上がみられた。

1. はじめに

昨年、中国にて発生した COVID-19(以下、新型コロナウイルス)の感染拡大により、世界中の国々・地域で深刻な問題を引き起こしている。この新型コロナウイルスは、感染者の咳やくしゃみ、会話に伴う飛沫感染が主な感染拡大の原因とされている。そのため、密を避け接触回数を減らし行動することが感染リスクを抑える最も有効な手段である。日本においては感染拡大防止を図るため、緊急事態宣言により人流を抑え、3密(密集・密閉・密接)を避けた行動を呼びかけている。密を避けた行動を計画するためには、人(群衆)における行動を予測し、未来における混雑度を把握する必要がある。

近年の位置情報取得技術の発展に伴い、スマートフォン、カーナビ、位置情報ソーシャルネットワークなどから大量の GPS 軌跡データが生み出され、行動予測問題解決に利用されている。人の行動予測研究においては、個人の移動軌跡に焦点を当てた研究が盛んに行われている。しかし、個人を対象としたモデリング手法を都市全体に拡大し、各人の軌跡データから各人の予測モデルを構築・統合させる

ことは現実的ではなく、プライバシーの問題にもなりかねない。群衆における行動予測においては、機械学習・深層学習手法の発達により、短時間かつ精度の高い予測モデル構築手法が提案されてきた。予測モデル構築にあたり、空間的特徴、時間的特徴の2種類の特徴が大きく寄与することが言われている。空間的特徴とは、人々の地理的位置情報(東京、港区、渋谷など)であり、時間的特徴とは、人の移動の規則性である。この二つの特徴の中でも、空間的特徴として PoI(Point of Interest) を利用した予測モデル構築手法が予測精度を向上させている [1]。PoI とは、場所におけるカテゴリー情報であり、人の行動と非常に強い関連が強いと報告されている [2]。既存手法においては、予測対象領域における PoI の個数や PoI カテゴリー名を利用し、空間的特徴を持つ情報として入力に利用されている。しかし、店や施設カテゴリーによって混雑時間帯が大きく異なるため、PoI の時系列性も考慮して予測モデルを構築する必要がある。

我々は、人混みを避けた行動計画支援することを目的に、PoI が持つ時系列性に焦点を当て、過去におけるある地域の各メッシュごとの PoI 別時系列混雑度情報から、未来におけるあるメッシュごとの混雑度を予測してする手法を提案した [3]。提案した手法の実験においては、データの種類による違い、時間帯による違い、予測先時間による違い、過去情報量の違いの4点の項目で検証を行った。本稿にお

¹ 奈良先端科学技術大学院大学

Nara Institute of Science and Technology

² 理化学研究所 革新知能統合研究センター

RIKEN Center for Advanced Intelligence Project

³ 国立研究開発法人科学技術振興機構 さきがけ

いては、データ数を増やし、PoI 別時系列混雑度情報の有効性を新たな実験手法で検証した。検証地域においても、複数の異なる地域で実験を行った。また、前回の前処理手法における滞在・移動判定の閾値の設定・線形補間のタイミングを、より実社会の人流に即した方法に改善を図った。その結果、PoI データにおいてベースラインと比べ3時間後の予測で予測精度が最も高い結果が得られ、PoI 別時系列混雑度情報が混雑度予測において有効であることが示された。

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、9章では、本論文の結論と今後の展望について述べる。

2. 関連研究

Song らは、匿名化された携帯電話ユーザーの移動パターンから、人の移動予測の限界を調査した。その結果、移動パターンには大きな違いがあるにもかかわらず、予測可能性にはばらつきがなく、予測可能性が93%であることがわかった [4]。

行動予測の研究においては、次のタイムウィンドウの遷移先を課題として設定し、マルコフモデルや統計的手法を用いて予測モデルの構築が行われてきた。これは個人の行動予測に焦点を当てている。群衆の行動予測においては、個人の行動予測よりも複雑になるため、機械学習や深層学習手法が多く用いられる。機械学習・深層学習手法では、空間的特徴と時間的特徴という2つの特徴が利用される。空間的特徴とは、予測対象地域のPoI情報や道路情報といった地理的特徴であり、時間的特徴とは人の行動の規則性等である。これらの特徴が予測モデルに大きく寄与する。近年では、2つの特徴を効率よく抽出し、より短時間かつ高精度な予測モデル構築手法が提案されている。

以下では、本研究と関係する既存研究として、PoIと人の行動の関係、個人における行動予測、群衆における行動予測、混雑揭示サービスについて概説し、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 [2]。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 [5]。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、次のタイムウィンドウにおける目的地候補の訪問確率を推定している [6]。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 [7]。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 [8]。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 [9]。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、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「群衆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 [10]。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 [11]。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、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 情報の関係に着目し、人の軌跡データと都市の PoI データを入力とした [1]。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 混雑揭示サービス

人混みを避ける仕組みとしては、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」[12] や Yahoo「混雑度レーダー」[13] 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。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

群衆の行動予測を行うため、個人における行動予測手法を用いる場合、各個人の予測モデルを構築し結果を統合する必要があるが、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。機械学習・深層学習を用いた手法においては、空間的特徴と時間的特徴を持つ入力データから効率よく特徴を抽出するため、様々な工夫をこらし予測精度を向上させた。既存手法においては空間的特徴として 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、図書館等の公共施設においては昼間の時間帯に利用者が多く混雑するが、居酒屋などの飲食店では夜間の時間帯に混雑す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 [14]。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。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。

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)

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 2: PoI 大分類

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

表 3: PoI 小分類 (一部抜粋)

鉄道・駅関連 仕出し・弁当・宅配 電気・事務・設備機器 (卸売)	パーキング・駐輪場 デパート・スーパー・ディスカウントショップ 小売関連	衣料関連 (小売) レンタカー
--	--	--------------------

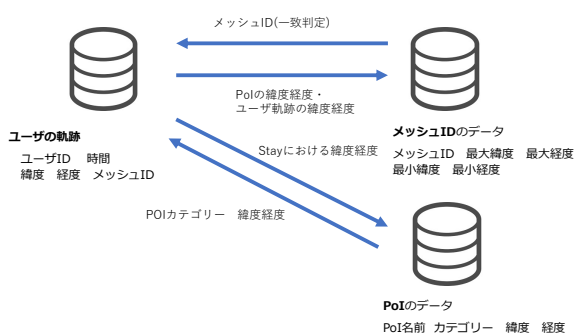


図 1: 各データの使用方法

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「POI データ」に含まれる、東京都内の 456,450 件の PoI を利用した。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 (表 3)、18 種類となる (表 2)。

4. 前処理

前処理として、GPS 軌跡データに対し、外れ値の除去、

メッシュの存在判定、移動・滞在の判定・PoI 付与、データ整形の順で行う。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。

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

子午線曲率半径を M 、卯酉線曲率半径を N 、2点の緯度の平均値を μ_y 、2点間の緯度と経度の差をそれぞれ d_y, d_x と表す。

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。GPS 精度においては誤差が半径 40m 以内の円内であるものを実験に使用している。各ユーザ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 メッシュ判定

メッシュデータセットを利用し、軌跡データの緯度経度が含まれるメッシュを判定し、そのメッシュ ID を付与する。メッシュ判定の際に所属メッシュが判明しなかったものに関しては本実験では使用しない。

4.3 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、本研究においては、データを揃えるためにデータ整形を行い、補間するため、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。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 100m 以

上の場合「移動 (move)」と判定し、100m 未満の場合、「滞在 (stay)」とした。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。その際、メッシュ判定で付与されたメッシュ ID をもとに、軌跡データが属するメッシュ内にある PoI 候補の中から、PoI の緯度軽度と、軌跡データの緯度経度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。また、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。

4.4 データ整形

スマートフォンの OS やアプリケーションの状態に応じて、異なる時間間隔で収集されるため、アップロードのタイミングがユーザによって異なり、サンプリングレートが不均一になっている。そのため、10 分間隔になるよう、後の値を参照してデータの穴埋めを行った。

5. 提案手法

提案手法においては、我々が過去に提案した手法を用いる [3].

5.1 提案手法における課題設定

本研究の課題設定は以下の通りである。

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と重回帰分析、Lightgbm, Xgboost の 4 種類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)、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, Root Mean Square Error) を使用した。予測手法における概略図を図 3 に示す。

6. 実験

本実験においては、データ数が多いメッシュ ID 「53394621」(東京都千代田区)、「53393558」(東京都品川区)、「53394525」(東京都新宿区)の 3 つの地域を対象とする。対象地域を図 6, 図 7, 図 8 に示す。使用するデータの期間は 2020/7/1~2020/7/31 (休日、祝日は除) までをデータとして使用した。この期間は緊急事態宣言解除後の人流が戻ってきたタイミングである。実験は、「PoI データの効果検証」と「過去情報量の違い」の 2 つの実験を行った。

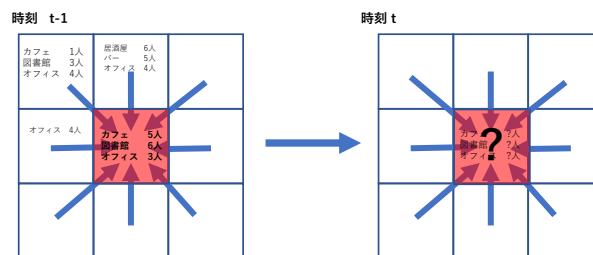


図 2: 本研究で解決したい課題の概略図

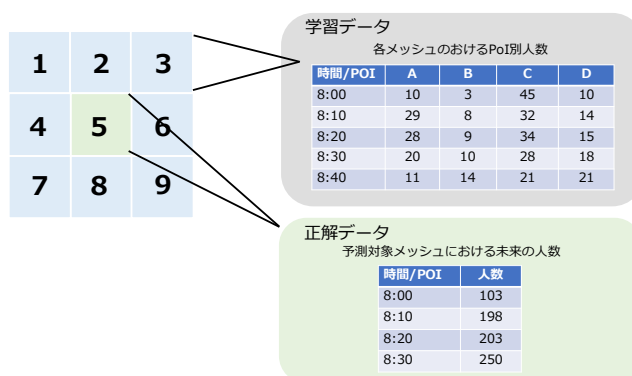


図 3: PoI 別人数情報を用いた予測手法

6.1 PoI データの効果検証

PoI データの効果検証においては以下の 3 点において検証を行う。

- PoI データが行動予測において有効性
- PoI カテゴリー数による予測精度の影響
- 予測先時間

使用するデータは、人数、滞在・移動の各人数、PoI 大分類 (18 種類) の各人数、PoI 小分類 (196 種類) の各人数の 4 つのデータセットを使用し検証を行った。本実験では 8 時から 23 時のデータを使用し、予測先時間を 10 分後、20 分後、30 分後、1 時間後、3 時間後に設定し、図 4 のようにデータフレームを生成し学習を行った。PoI 別人数データに関しては図 5 のようにデータフレーム生成した。

6.2 過去情報量の違い

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 情報の有効性を検証するとともに、何分前の情報が予測において有効であるかを検証する。今回の実験においては PoI データの効果検証実験と同様 4 種類のデータセット、8 時から 23 時のデータを使用し、予測先時間は 10 分後、30 分後、1

学習データ				正解データ
時刻	メッシュ1	メッシュ2	メッシュ9	予測対象メッシュ
8:00	6人	9人	10人	9人
8:10	3人	8人	6人	
8:20				
⋮				

図 4: データフレーム例

学習データ				正解データ
時刻	メッシュ1	メッシュ2	メッシュ9	予測対象メッシュ
8:00	飲食店 6人	運送 9人	飲食店 10人	運送 9人
8:10	5人	13人	13人	20人
8:20			25人	4人
⋮				

図 5: PoI データフレーム例

時間後, 3 時間後に設定した。過去情報量については, 「10 分前」, 「30 分前まで (10 分前, 20 分前, 30 分前)」, 「60 分前まで (10 分前, 20 分前, 30 分前, 40 分前, 50 分前, 60 分前)」の 3 種類で検証を行った。

7. 結果

7.1 PoI データの効果検証の結果

PoI データの効果検証における RMSE による評価では図 12, 図 13, 図 14, 図 18, 図 19, 図 20, 図 24, 図 25, 図 26, MAPE による評価では図 15, 図 16, 図 17, 図 21, 図 22, 図 23, 図 27, 図 28, 図 29 のような結果となった。10 分後から 1 時間後の予測ではデータの種類の異なる予測精度の有意差は見られなかったが, 3 時間後の予測においては PoI 小分類が一番良い精度を得られた。PoI 大分類と比べ PoI 小分類の方が高い精度が得られたことから, カテゴリ数が予測に影響を与え, カテゴリ数が多いほど有効であることが示された。3 時間後の LightGBM による予測においては, 人数のみと比較し PoI 小分類の予測では RMSE の値が約 50 % 減少し, MAPE においては, 誤差率が約 11 % 減少した。3 地点 (丸の内, 新宿, 品川) 全てにおいて PoI 小分類が最も高い精度が得られたことから, PoI を用いた予測は場所を問わず有効であることが示された。また, MAPE の結果より, 3 時間後の予測において PoI 小分類のデータを使用することで誤差率 10 % 以内, 最小誤差率約 5 % 以内で予測可能であることが明らかになった。ランダムフォレスト, LightGBM, XGboost の 3 種類モデルにおいて PoI 小分類が最も高い精度が得られ, 大体同様の結果となったが, PoI 小分類を用いた重回帰分析においては表 4 が示すように, 大幅に予測が外れ RMSE の値が大きくなった。

7.2 過去情報量の違いの検証結果

過去情報量の違いの検証結果は, 図 9, 図 10 に示す。過去情報量の違いにおいても PoI データの効果検証と同様, 10 分後から 1 時間後の予測ではデータの種類の異なる予測精度の有意差は見られなかったが, 3 時間後の予測では, 10 分前のデータのみ, 30 分前までのデータ, 1 時間前までのデータの全てにおいて PoI 小分類が最も高い精度を得られた。ラグ特徴量を使用した場合においても, カテゴリ数が多いほど予測精度が向上した。特に 10 分前のデータのみをラグ特徴量として追加した時, データの種類の違いが大きく見られ, 人数データのみの場合と比べ, PoI 小分類データでは RMSE の値が約 30 % 減少した。また, ラグ特徴量を使用しなかった場合 (図 12) とラグ特徴量を使用した場合 (図 9) の結果を比べた時, 3 時間後の結果においてラグ特徴量を使用した場合の方が良い精度が得られた。これにより, PoI の時系列情報が有効であることが示された。1 時間前までのデータをラグ特徴量として追加した時, PoI 小分類が最も良い精度であったが, 明確な差がある結果とはならなかった。

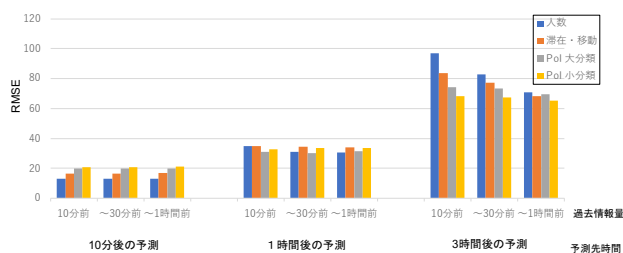


図 9: 丸の内におけるラグ特徴量を使用したランダムフォレストによる予測 (RMSE)

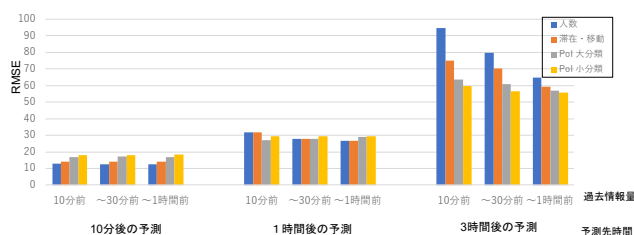


図 10: 丸の内におけるラグ特徴量を使用した LightGBM による予測 (RMSE)

8. 考察

人数のみ, 滞在・移動の各人数のデータに比べ PoI データが有効であることが示され, PoI のカテゴリ数が予測精度に影響をもたらすことがわかった。これは PoI 別人数情報が予測において有効な特徴量であることを示している。10 分後の予測から 1 時間後の予測においては, データの種類によって明らかな有意差は見られなかったが, 3 時間後

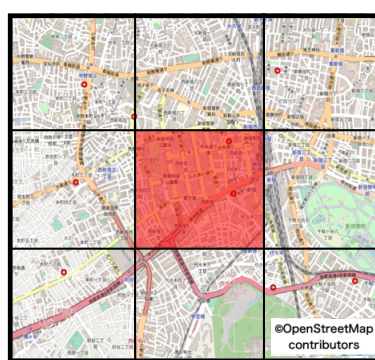
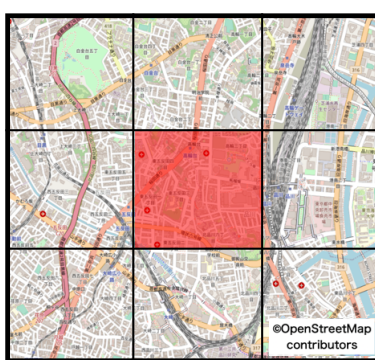
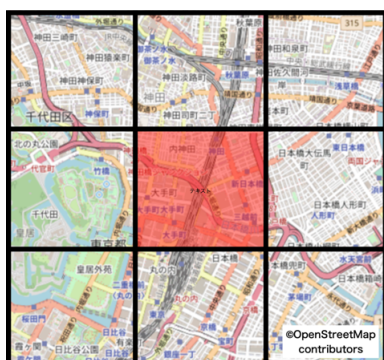


図 6: 実験における対象地域 1 (丸の内) 図 7: 実験における対象地域 2 (品川) 図 8: 実験における対象地域 3 (新宿)

表 4: 重回帰分析による予測 (RMSE)

入力データ/予測先時間	人数	滞在・移動	PoI 大分類	PoI 小分類
10 分後	14.78769189	12.74610562	11.1391015	17520612.41
20 分後	24.1749415	19.58892732	14.33588387	39737897.49
30 分後	33.26293094	26.18191574	17.41682057	7020354.825
1 時間後	93.78387714	65.8226205	41.37107343	26001898.61
3 時間後	188.4730241	149.4961189	86.43121908	103.3401887

のデータにおいては有意差が見られた。これは、図 11 で示しているように、予測先時間が近い場合、人数が大幅に変わることがないため予測しやすくなっており、3 時間後の予測においては大幅に人数が変わるため予測しにくい。そのため、人数のみといったデータでも予測先時間が近い場合は精度よく当てられるが、予測先時間が離れている場合、有効な特徴量を利用しないと予測が難しいことがわかる。ラグ特徴量を使用した実験においては、1 時間前までのデータをラグ特徴量として追加した時、有意差がみられなかった原因として、多量な特徴量が挙げられる。例えば、PoI 小分類において 1 時間前までのデータをラグ特徴量として使用した時、196 種類のカテゴリ × 6 個 (1 時間あたりのデータ数) 分が追加されるため、うまく特徴を抽出できなかった可能性がある。また、3 時間後の重回帰分析において、学習ができなかった原因も同様に特徴量の多さが挙げられる。そのため、有効な特徴量を選択し、使用することで予測の改善に繋がれると考える。

9. まとめ

本稿では、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 別混雑度情報を用いることで、人数データのみで学習し予測した場合と比べて、3 時間後におけるメッシュ領域の混雑度 (存在人数) の予測に関して、最大 50% 予測誤差が減少し、最小誤差率約 5% で予測可能であることを示した。また、PoI 別時系列混雑度情報を用いることで時系列を考慮しない場合と比べ、精度の向上がみられた。今回の実験においてはデータセットやパラ

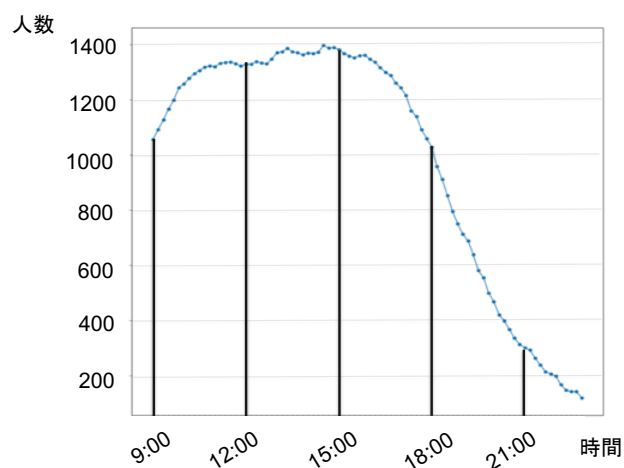


図 11: 2020/07/01 の予測対象メッシュにおける人数推移

メータを増やし、PoI 別時系列混雑度情報が行動予測において有効な特徴量であることを示した。今後は、機械学習手法だけではなく、深層学習手法による予測モデル構築に取り組みたいと考えている。また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、実験で設定した丸の内、新宿、品川周辺では電車での移動も激しいことから、予測モデル構築にあたり対象エリアの拡大が必要である。そのため、予測対象エリアを広げ、論理的距離考慮した入力メッシュの選定も予測精度向上に寄与すると考えられる。効率よく特徴量を抽出し、短時間でモデル構築が行える手法について検討していきたい。

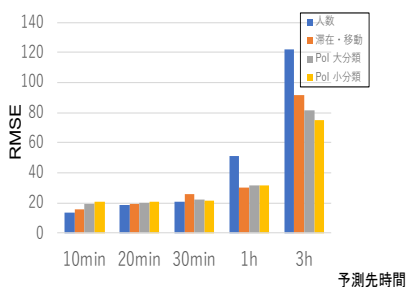


図 12: ランダムフォレストによる予測 (RMSE, 丸の内)

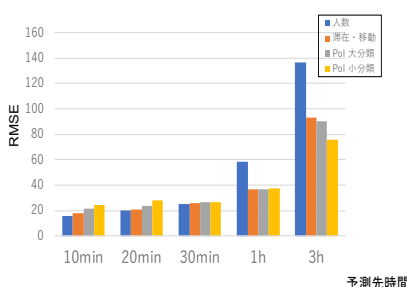


図 13: XGBoost による予測 (RMSE, 丸の内)

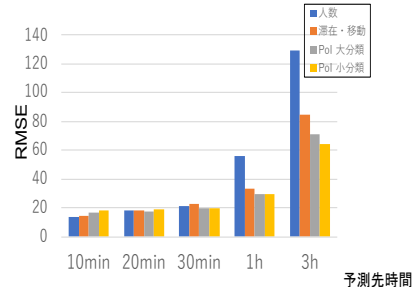


図 14: Lightgbm による予測 (RMSE, 丸の内)

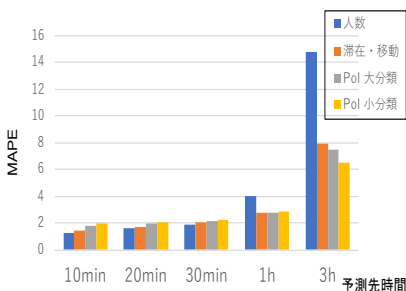


図 15: ランダムフォレストによる予測 (MAPE, 丸の内)

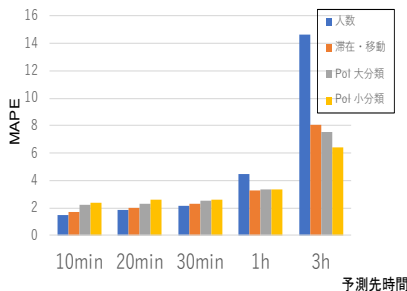


図 16: XGBoost による予測 (MAPE, 丸の内)

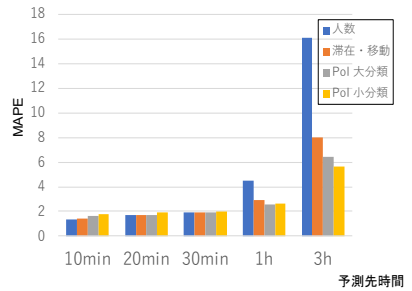


図 17: Lightgbm による予測 (MAPE, 丸の内)

参考文献

[1] 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.

[2] 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.

[3] 山田理, 松田裕貴, 諏訪博彦, 安本慶一. 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. *Multimedia, Distributed, Cooperative, and Mobile Symposium (DICOM 2021)*.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] Vacan. <https://corp.vacan.com/>. Accessed: 2021-04-29.

[13] Yahoo!map. <https://map.yahoo.co.jp/promo/>. Accessed: 2021-04-29.

[14] Agoop サービス製品. <https://www.agoop.co.jp/service/>. Accessed: 2021-04-29.

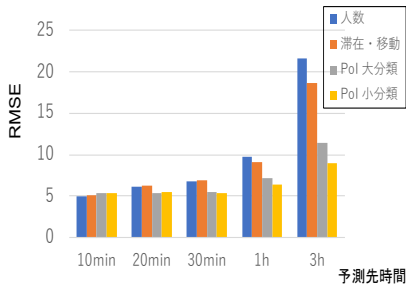


図 18: ランダムフォレストによる予測 (RMSE, 品川)

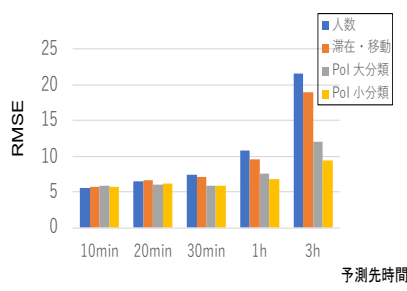


図 19: XGBoost による予測 (RMSE, 品川)

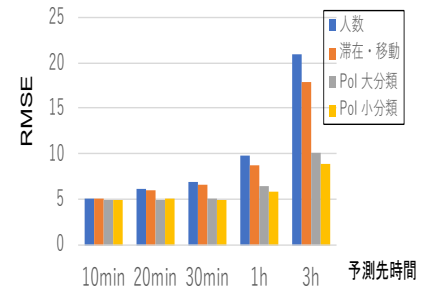


図 20: Lightgbm による予測 (RMSE, 品川)

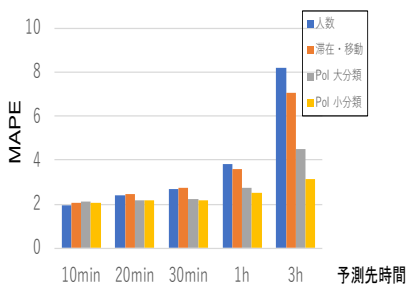


図 21: ランダムフォレストによる予測 (MAPE, 品川)

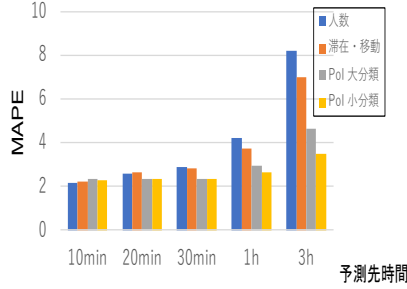


図 22: XGBoost による予測 (MAPE, 品川)

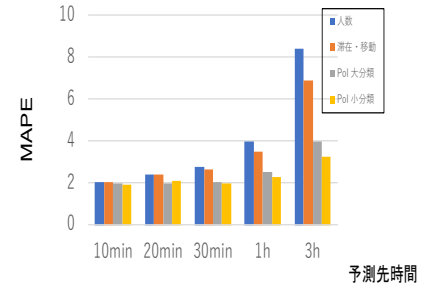


図 23: Lightgbm による予測 (MAPE, 品川)

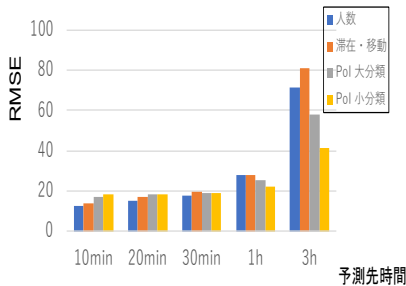


図 24: ランダムフォレストによる予測 (RMSE, 新宿)

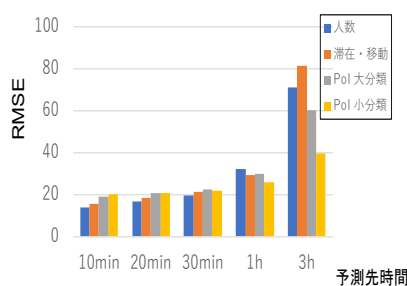


図 25: XGBoost による予測 (RMSE, 新宿)

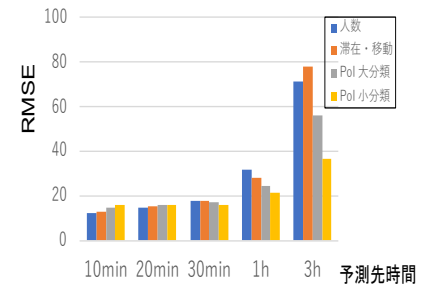


図 26: Lightgbm による予測 (RMSE, 新宿)

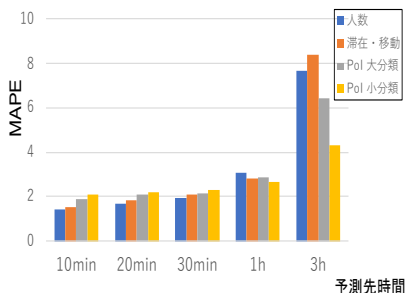


図 27: ランダムフォレストによる予測 (MAPE, 新宿)

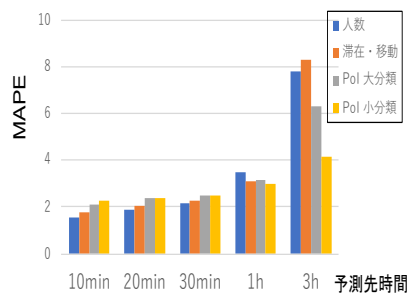


図 28: XGBoost による予測 (MAPE, 新宿)

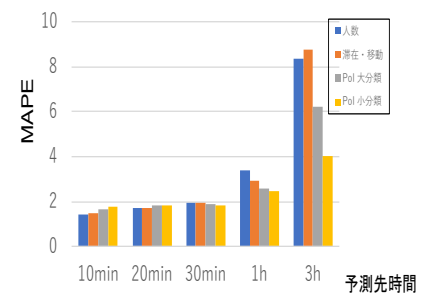


図 29: Lightgbm による予測 (MAPE, 新宿)