

# 残差回帰による複合時空間予測手法

佐藤 大祐<sup>†</sup> 松林 達史<sup>†</sup> 福田 匡人<sup>†</sup> 堤田 恭太<sup>†</sup> 中山 彰<sup>†</sup>  
戸田 浩之<sup>†</sup>

<sup>†</sup> NTT サービスエボリューション研究所 〒239-0847 神奈川県横須賀市光の丘 1-1  
E-mail: [†daisuke.satou.br@hco.ntt.co.jp](mailto:†daisuke.satou.br@hco.ntt.co.jp)

あらまし 空間的な広がりを持ち、一定の時間間隔で観測される、人数や降雨量などの時空間データの変動を予測することは重要な課題である。予測結果の活用の観点から、周期性を持つ変動の予測だけでなく、大規模イベント時におけるエリア内人数変動やゲリラ豪雨時の降雨量など、過去に経験したことない変動への対応も必要である。過去の傾向と異なる変動に対し、直近のデータを用いて予測モデルを学習し直すことが有効だが、時空間的な相関を考慮した複雑な予測モデルでは、学習に伴う計算量の大きさが問題となる。本研究では、過去に経験した変動を学習したシンプルな時系列予測モデルと、時系列予測モデルの予測残差を回帰する時空間予測モデルを組み合わせることで予測精度を保ちながら処理全体の計算量を抑えた複合時空間予測手法を提案する。本稿では人口統計データを用いて実験を行い、提案手法が処理速度および予測精度を両立させるうえで有効であることを確認した。

キーワード 時空間データ, 時系列解析, 残差回帰

## 1 はじめに

空間的な広がりを持ち、一定の時間間隔で観測される時空間データの変動予測は、降雨量などの天候予測 [1], 人の流れによる混雑予測 [2, 3] など様々な分野で行われている。

特に混雑予測の分野では、センシング技術の向上や IoT デバイスの急速な普及に伴い、混雑状況の計測技術が数多く提案されており [4–6], 多数のセンサからリアルタイムにデータを取得し、即座に利用することも可能になりつつあり、次々と移り変わる状況に対応した予測を行うことが重要である。

従来、時系列データの予測技術には自己回帰モデル (AR) と移動平均 (MA) を用いた自己回帰移動平均モデル (ARIMA) [8] が広く用いられており、過去のパターンから平均的に予測を行う技術が広く用いられている。また、過去のパターンでも季節性を考慮した自己回帰移動平均モデル (SARIMA) [9] や、複数変数間における時間変動を考慮したベクトル自己回帰モデル (VAR) [10] などとも利用されているが、いずれも時系列的なパターンのみに注目し、空間的な近接性などの考慮がされていない。

一方、空間的な近接性を考慮し、空間的自己相関を考慮した方法にクリギング法 [11, 12] がある。クリギング法は未観測地点の観測値を、周囲の観測値から推定する手法であり、時空間拡張した時空間通常クリギング法 [11] を用いることで、未観測である未来時刻の観測値を推定することが可能である。クリギング法には、蓄積されたデータがなくても直近のデータのみから予測を行うことができるという利点があるものの、時間的、空間的な近接性を考慮するため、急激な変化にほとんど追従できないという問題があった。

近年では、深層学習ベースによる時空間データの予測手法が

多数提案されている。Shi ら [1] は時系列データに対して広く使われている深層学習ベースの予測モデルである LSTM [16] を、周辺のデータを畳み込むように拡張した Convolutional LSTM を提案している。その後、ネットワーク構造を複雑に拡張し、誤差伝搬のパスをさらに深くすることで、より予測性能を高めた手法も複数提案されている [17–19]。これらの深層学習ベースの手法は過去の大量の蓄積データを学習することで、高い精度での予測が期待できるが、学習データと大きく傾向が異なる変動への対応は期待できない。

上田 [7] は、イベント会場に急激に人が集まるような、通常時と異なる突発的な変動を含む時空間データに対しても予測を行っている。予測は、時空間的な近接性を考慮した観測値分布の空間モデルを作成し、未来時刻の空間モデルパラメータを時系列予測することで、非定常な時空間データの変動を精度よくとらえる事が可能である。この予測手法ではリアルタイムに得られる最新の時空間データを組み込むために、モデルを逐次更新しながら予測を行う。しかし、観測時刻数や観測地点数が多くなることで計算負荷が大きくなるため、同時に蓄積したデータから周期性などをモデリングしようとする逐次更新することができない。

本研究では、空間上に設置されたセンサから一定時間間隔で取得する (タイムスタンプ, 座標, 観測値) からなる過去の時空間観測データから、未来時刻の観測値を予測する問題に取り組む。これまでの時空間予測手法では、過去の蓄積データの変動傾向をとらえたモデリングと、過去に経験のない変動が起きた際のモデルの更新を、高い計算負荷をもちながら両立することは困難であった。従来法の課題を解決するために、過去の変動を学習した単一地点ごとのシンプルな時系列予測モデルと、時系列予測モデルの予測残差を回帰する時空間予測モデルを組み合わせることで予測精度を保ちながら処理全体の計算量を抑え

た複合時空間予測手法を提案する。本稿では人口統計データを用いて実験を行い、提案手法の性能の検証を行った。

## 2 従来手法

イベント会場に急激に人が集まるような、通常時と異なる突発的な時空間データの変動を予測するための従来手法として、上田 [7] において提案されている時空間予測モデルの概要について説明する。従来手法では、過去の傾向と異なる変動に対応するため、直近に得られたデータのみをもちいて時空間的な相関を学習し、予測モデルを逐次作り直す。時刻及び座標にひもづく時空間観測値データを入力とし、後述の処理を行うことで、予測先時刻における任意地点の予測値を行う。

まず、クリギング法による空間内挿推定処理を行う。特に入力となる時空間データが空間的に疎である場合に、空間モデルの精度を上げるためにデータ点を増やす。

次に、カーネル回帰による空間回帰処理を行い、観測値を位置座標  $x$  の関数として空間的にモデル化する。各時刻  $t$  の空間モデル  $f_t(x)$  は重み付きのカーネル関数の線形和とし、以下の式で表される。

$$f_t(x) = \sum_{j=1}^M w_{jt} K(x; \theta_j) \quad (1)$$

カーネル関数は Radial Basis Function(RBF) カーネル:

$$K(x; \theta_j) = \frac{1}{2\beta_j^2} \exp(-\|x - \mu_j\|^2) \quad (2)$$

を用いる。ここで、 $\theta_j = (\mu_j, \beta_j)$  はカーネルパラメータで、 $\mu_j$  は位置パラメータ、 $\beta_j$  はカーネル幅パラメータであり、 $M$  はカーネル数を表す。カーネルパラメータ  $\theta_j$  は各タイムステップ  $t$  で共通とし、重みパラメータ  $w_{jt}$  によって  $t$  での分布のバリエーションを表現する。

上記で求めた空間モデルの重みパラメータ  $w_{jt}$  の時系列データから、予測先時刻  $t_f$  の重みパラメータ  $\hat{w}_{jt_f}$  の予測を行う。予測のためのモデルとして、過去  $\tau$  時刻までの全カーネル関数の重みパラメータの線形和で近似したものを用いる。

$$\hat{w}_{jt_f} = \sum_{k=1}^M \sum_{l=1}^{\tau} \phi_{jkl} w_{k(t_0-l+1)} + \epsilon_j, j = 1, \dots, m \quad (3)$$

ここで、 $\phi_{jkl}$  は式 (3) における重み係数であり、 $\epsilon_j$  は誤差項である。 $\phi_{jkl}$  は縮小ランク回帰 [14] により低ランク分解した上で近似式の自乗誤差が最小となるものを求める。予測した時刻  $t_f$  における空間モデルの重みパラメータ  $\hat{w}_{jt_f}$  を式 (1) に適用することで、未来時刻の観測値分布を推定する。

## 3 提案手法

提案手法では、計算コストの低いシンプルな時系列予測モデルと時空間予測モデルを組み合わせる。提案手法の処理全体の概要を図 1 に示す。まず、蓄積された時空間データから、地点ごとに過去の時系列変動パターンを抽出してそれぞれ時系列予測モデルの学習を行う。次に、学習した時系列予測

### Algorithm 1 時系列予測モデル

**Input:**

$\mathbf{x}_l = \{\mathbf{x}_{t_0-\tau_l}, \dots, \mathbf{x}_{t_0}\}$  // 入力観測値  
 $a(t) \in \{0, \dots, 23\}$  // 補助情報

**Procedure:**

```

1: 時系列パターンを格納する PatternList を初期化
2: 残差ベクトルを格納する  $\mathbf{e}$  を初期化
3: for  $i = 1, \dots, I$  do // 地点ごとに処理
4:   for  $t = t_0 - \tau_l + T, \dots, t_0$  do
5:      $\mathbf{p}_{it} = \{x_{i(t-T+1)}, \dots, x_{it}\}$  // 時系列パターン
     // PatternList に時系列パターンを蓄積
6:      $PatternList[i][a(t - \tau_f)]$  add  $\mathbf{p}_{it}$ 
     // 蓄積パターンから最も  $\mathbf{p}_{it}$  に類似するパターンを探す
7:      $\hat{\mathbf{p}}_{it} = Nearest(PatternList[i][a(t)], \mathbf{p}_{it})$ 
8:      $\hat{x}_{it_f}^{temporal} = \hat{\mathbf{p}}_{it}[\tau_p + \tau_f]$  // 類似パターンからの予測
9:      $e_{it} = x_{it} - \hat{x}_{it_f}^{temporal}$  // 時刻  $t$  における残差
10:   end for
11:   if  $PatternList[i]$  のサイズが大きい then
12:     for  $a = 0, \dots, 23$  do
       // クラスタリングによる PatternList の圧縮
13:        $PatternList[i][a] = Clustering(PatternList[i][a])$ 
14:     end for
15:   end if
16: end for
17: return  $\hat{\mathbf{x}}_{t_f}^{temporal}, \mathbf{e}, PatternList$ 

```

モデルによる予測を行い、当てはめた際の観測データとの残差を抽出する。時空間モデルでは、大きな残差データのみを入力とし、予測先時刻の残差の予測を行う。以下、2つの予測モデルでの予測方法と、その合成方法について説明する。

### 3.1 時系列予測モデル

Algorithm.1 に時系列予測モデルの具体的な処理フローの例を示す。観測データは、観測地点  $i$ 、一定の時間間隔のタイムステップを  $t$  とした観測値  $x_{it}$  が欠損なく格納されているものとする。現在時刻を  $t_0$ 、学習期間を  $\tau_l$  とし、時刻  $t_0 - \tau_l$  から  $t_0$  までの観測値  $\mathbf{x}_l = \{\mathbf{x}_{t_0-\tau_l}, \dots, \mathbf{x}_{t_0}\}$  を学習のための入力データとする。また、 $t$  にひもづき、時系列変動パターンと関連する補助情報を  $a(t)$  とし、これも入力データに含まれるとする。補助情報は、例えば曜日や時刻などの観測時刻のタイムスタンプから得られる情報であり、複数の情報から成る場合もある。ここでは補助情報を 0 から 23 までの離散値となる時刻情報  $a(t) \in \{0, \dots, 23\}$  として説明する。

変動パターンの学習は、学習用の観測データから地点  $i$  ごとに、パターン長  $T$  の時系列パターンを切り出す。ここで  $T$  は、予測時に説明変数として用いる説明変数期間  $\tau_p$  と、予測期間  $\tau_f$  を合わせた長さとし、学習期間内の各時刻  $(t - T + 1)$  から  $t$  を時系列パターン  $\mathbf{p}_{it}$  とする (Algorithm.1:5 行目) 図 2 に、学習データからの時系列パターンを切り出すイメージと、各時間変数との対応を示す。

次に、切り出した時系列パターンを、各パターンの補助情報  $a(t - \tau_f)$  ごとにグルーピングしたうえで、各補助情報グループ

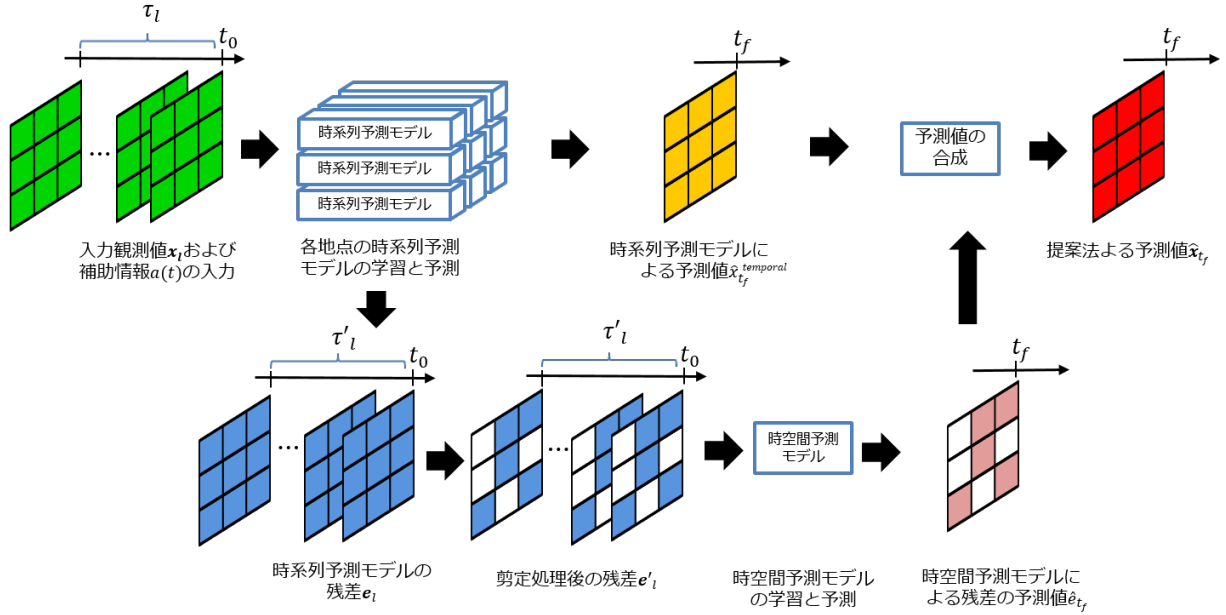


図 1 提案手法の処理の流れ

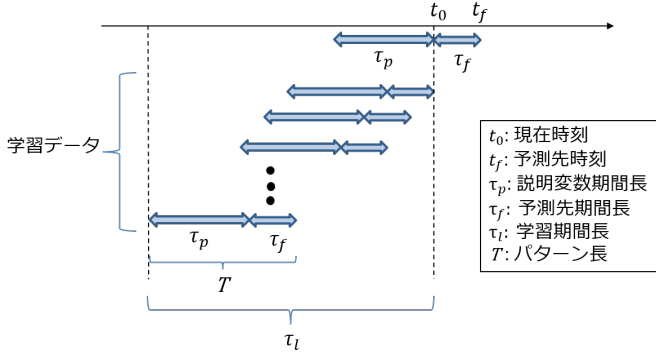


図 2 時系列予測モデルの学習

ごとに格納する (Algorithm.1:6 行目). これは、予測を行う際に、予測実施時刻の補助情報と同じ時系列パターンを利用するために行う。

予測は、予測実施時刻の補助情報と同じ蓄積時系列パターンのうち、直近の説明期間  $\tau_p$  の観測値と最も類似するパターンの  $\tau_f$  部分を推定値として行う (Algorithm.1:7,8 行目). 類似パターンは、観測値とのパターンの距離  $d$  が最も小さくなるように選択し、

$$\text{Nearest}(\mathbf{U}, \mathbf{p}) = \arg \min_n d(\mathbf{u}_n, \mathbf{p}) \quad (4)$$

として定義する。ここで、表記を簡単にするために、選択対象となるパターン集合を  $\mathbf{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_N\}$ ,  $N$  を蓄積パターン数、直近の観測値を  $\mathbf{p}$  としている。類似の基準となる距離  $d$  は、期間  $\tau_p$  における二乗誤差

$$d(\mathbf{u}_n, \mathbf{p}) = \sum_j (x_{(t-\tau_p+j)} - u_{nj})^2, j = 1, \dots, \tau_p \quad (5)$$

とする。このとき、比較する際の添え字が2つの時系列パターンで異なるのは、 $\mathbf{u}_n$  はパターン長が  $T$ ,  $\mathbf{p}$  は  $\tau_p$  となるため

ある、

このように、時系列予測モデルでは学習期間内に、時系列パターンの蓄積と予測を古い時刻から順に行う。この処理の中で、予測と観測の誤差である予測残差を求めておく (Algorithm.1:9 行目)

時系列パターンの蓄積を終えた際に、蓄積したデータのサイズが大きくなりすぎた場合、各グループの時系列パターンの集合に対しクラスタリング手法を適用し、類似するパターンを1つの変動パターンとしてまとめることで学習データのサイズを削減する。本研究では、k-means [15, 20] をクラスタリング手法として用いる。クラスタリング手法によって出力された各クラスターの平均値ベクトルを蓄積したデータと置き換え、新たに予測のための時系列パターンとする。

最終的に、時系列予測モデルの処理の結果として、蓄積した時系列パターンと、学習期間内での予測残差ベクトル  $\mathbf{e}$ 、未観測時刻  $t_f$  の予測値  $\hat{\mathbf{x}}_{t_f}^{\text{temporal}}$  を出力する。

### 3.2 時空間予測モデルによる残差回帰

時系列予測モデルは、学習データに含まれる変動パターンと観測される直近の変動パターンが類似する場合のみ、精度よく予測することができる。そこで、学習データに含まれない変動パターンに対しては、直近の時空間データの解析に基づく予測手法と組み合わせることで対応する。具体的には、時系予測モデルの残差の変動を入力として時空間予測モデルの学習を行い、時空間予測モデルによる残差の予測値を時系列予測モデルの予測値と合わせることで最終的な予測を行う。Algorithm.2 に、時空間予測モデルによる残差回帰の処理フローを示す。

残差回帰処理では、現在時刻を  $t_0$ 、時空間予測モデルの学習期間長を  $\tau'_l$  とし、時刻  $t_0 - \tau'_l$  から  $t_0$  までの残差  $\mathbf{e}_t = \{e_{t_0-\tau'_l}, \dots, e_{t_0}\}$  を入力データとする。このとき、時空間予測モデルでは直近のデータのモデリングを行う目的か

## Algorithm 2 時空間予測モデルによる残差回帰

Input:

$e_l = \{e_{t_0-\tau'_l}, \dots, e_{t_0}\}$  //時系列予測モデルの残差

Procudure:

//剪定閾値  $k$  より残差の小さな地点データの剪定

```
1: for  $i = 1, \dots, I$  do
2:   if  $e_{li}$  の最大値  $< k$  then
3:      $e_{li}$  を削除
4:   end if
5: end for
6: 剪定処理後の残差データを  $e'_l$  とする
7:  $\hat{e}'_{t_f} =$  時空間予測モデル ( $e'_l$ )
8: return  $\hat{e}'_{t_f}$ 
```

ら、 $\tau'_l$  は時系列予測モデルの学習期間  $\tau_l$  よりも短い期間を設定する。

入力された全観測地点の残差をすべてモデリングすると、地点数が大きい場合に処理が大きくなってしまう。そこで、残差が小さい地点は過去の変動傾向と似た変動であることから、予測対象時刻  $t_f$  においても時系列予測モデルの当てはまりがよく、予測残差が小さいことが想定されるため、入力残差が大きく、過去の変動と異なる変動が起きている地点のみ時空間予測モデルで扱うことで、予測精度を維持しながら処理が大きくなることを防ぐ。具体的には、各観測地点の残差の最大値  $\max(e_{li})$  が剪定閾値  $k$  を下回るかどうかを基準として、扱うデータ量を減らすためのデータ剪定処理を行う (Algorithm.2:1~5 行目)。

剪定処理によりデータ数を削減した時空間残差データ  $e'_l$  を入力とし、過去の変動傾向と異なる地点の予測先時刻  $t_f$  の残差変動を予測する (Algorithm.2:6,7 行目)。ここで、時空間予測モデルは直近の限られた変動パターンから、時空間的な相関をモデリングする必要がある、本研究では上田 [7] による時空間予測モデルを用いる。この手法では観測分布の空間モデルを作成するため、剪定した観測地点の予測値を出力することもできるが、残差が小さくなるという想定から、本研究では剪定した地点の予測は行わない。

### 3.3 予測値の合成

3.1 節で導出した過去の時系列変動パターンによる予測値と、3.2 節で導出した時空間予測モデルによる残差の予測値合成し、最終的な予測値  $\hat{x}_{t_f}$  として出力する。このとき、残差の予測値  $\hat{e}'_{t_f}$  は、残差の大きい地点の予測値のみであるため、剪定した地点の残差の予測値をすべて 0 としたものを  $\hat{e}_{t_f}$  として合成する。

$$\hat{x}_{t_f} = \hat{x}_{t_f}^{\text{temporal}} + \alpha \hat{e}_{t_f} \quad (6)$$

ここで、 $\alpha$  は予測残差の採用率であり、 $0 \leq \alpha \leq 1$  の値を設定する。予測先期間長  $\tau_f$  が長くなるほど残差の予測は困難になるため、 $\alpha$  に小さな値を設定することで、過去の変動傾向に大きく依存した予測を行い、逆に  $\tau_f$  が短い場合には  $\alpha$  に大きな値を設定することで、より積極的に過去の変動と異なる変動を

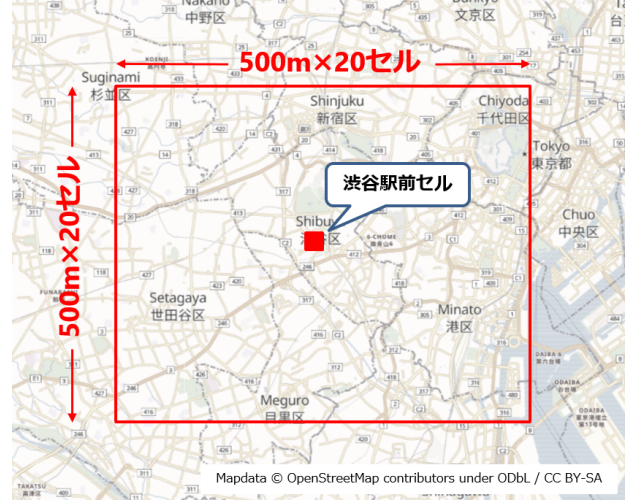


図3 実験データセットの対象エリア。渋谷駅周辺を中心とした約10km四方の領域であり、データの単位領域である500mセルが400個分の大きさである。

予測結果に含めることが可能となる。

## 4 実験

提案手法の有効性を評価するため、人口統計データを用いて実験を行った。実験条件および実験結果について述べる。

### 4.1 実験条件

#### 4.1.1 データセット

本実験では、携帯電話のネットワークのしくみから推計される人口統計情報であるモバイル空間統計<sup>1</sup>のデータを用いる。人口データは、1辺の長さが約500mのセル内の人数推計値であり、1タイムステップは1時間である。各セルは2分の1地域メッシュ<sup>2</sup>の範囲と同等である。手法を適用する対象エリアは、渋谷駅を中心とする約10km四方の範囲とする。これは500mセルのデータが縦20×横20に敷き詰められた、計400セルの大きさのデータとなる。実験データセットの対象エリアを図3に示す。

#### 4.1.2 手法設定

補助情報として、タイムステップ  $t$  における時刻情報を用いる。ここで、時刻情報は日付を含まないタイムスタンプであり、0時から23時までの24種類の観測時刻から成る。説明変数期間  $\tau_p = 24$  とし、1ステップ先の予測を行う。 $\tau_f = 1$  より  $T = 25$  である。時系列変動パターンの事前学習期間を2018/8/1~2018/10/28の89日間(2136ステップ)とした。事前学習により、400地点、24時刻ごとに89パターンの時系列パターンを蓄積し、さらに、クラスタ数を50としたk-meansによるクラスタリングを行った。

時空間予測モデルのハイパーパラメータを決定するため、テスト期間の1週間前にあたる2018/10/22~10/24の3日間(72ステップ)のデータを使ってグリッドサーチを行った。時空間

1: モバイル空間統計 <https://mobaku.jp/>

2: 総務省統計局 地域メッシュ統計 <http://www.stat.go.jp/data/mesh/>

予測モデルのみでの予測を行い、400 セルの予測誤差が最も小さいパラメータを選択した。また、本実験では空間的に密な人口統計データを用いるため、時空間予測モデルにおける空間内挿処理は省略して実行した。残差データに対する剪定閾値  $k$  は 1000 とし、予測値合成の採用率  $\alpha$  は 1 とした。

本実験でのテスト期間は 2018/10/29~10/31 の 3 日間 (72 ステップ) とした。テスト中もタイムステップ毎に時系列変動パターンの学習をオンラインで行った。ただし、テスト中にクラスタリング処理は行っていない。

#### 4.1.3 比較手法

提案法の予測精度を評価するため、複数の予測手法との比較を行った。比較手法には、基本的なモデルとしてコピーモデルと平均モデルを、深層学習ベースの時系列予測モデルとして広く使われている LSTM と convolutional LSTM を、そして提案手法に組み込まれている時系列予測モデルと時空間予測モデルを選択した。実験における各比較手法の設定について述べる。  
**コピーモデル (Copy)** 予測対象となる地点・時刻における、1 時刻前の値を予測値とした。

**平均モデル (Avg.)** 予測対象となる地点・時刻・曜日における、過去 4 週間の平均値を予測値とした。

**LSTM** 時系列データに対して広く使われている深層学習ベースの予測モデル [16]。各地点ごとにモデルの作成と学習を行い、それぞれのモデルで予測を行った。モデルの構成はレイヤ数を 1、隠れ層のユニット数を 128 とした。2018/8/1~2018/10/28 の 89 日間をモデルの学習期間とした。

**convolutional LSTM (ConvLSTM)** 時空間データに対して周囲のデータを畳み込んで LSTM を行う、深層学習ベースの予測モデル [1]。予測対象となる 400 地点を 1 つのモデルで扱うよう学習し、予測を行った。モデルの構成はレイヤ数を 4、すべてのレイヤでフィルタ数を 128、カーネルサイズを  $5 \times 5$  とした。2018/8/1~2018/10/28 の 89 日間をモデルの学習期間とした。

**時系列予測モデル** 提案手法で用いられる時系列変動パターンのみによる予測。提案法と同様の実験設定でモデルの作成と学習を行った。

**時空間予測モデル** 提案手法で用いた時空間予測モデル [7] のみによる予測。提案法では入出力を時系列予測モデルの予測残差としたが、ここでは時空間データを入出力としてモデルの作成と学習を行った。

#### 4.1.4 評価指標

評価指標として二乗平均平方根誤差 (Root Mean Squared Error, RMSE)、平均絶対誤差 (Mean Absolute Error, MAE)、平均絶対誤差率 (Mean Absolute Percentage Error, MAPE) を用いる。RMSE, MAE および MAPE はそれぞれ以下の式で求める。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i \in I} \sum_{t \in T} (x_{it} - \hat{x}_{it})^2} \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i \in I} \sum_{t \in T} |\hat{x}_{it} - x_{it}| \quad (8)$$

$$MAPE = \frac{1}{n} \sum_{i \in I} \sum_{t \in T} \left| \frac{\hat{x}_{it} - x_{it}}{x_{it}} \right| \quad (9)$$

ここで、 $I$  は評価対象地点数、 $T$  は評価対象時刻数を表し、 $n$  は  $I \times T$  で求められる評価対象となるデータ数である。 $\hat{x}_{it}$ ,  $x_{it}$  はそれぞれ、地点  $i$ 、時刻  $t$  の予測値と正解となる観測値である。

## 4.2 実験結果

実験で得られた各モデルの予測精度を比較した結果を表 1 にまとめる。予測精度は RMSE, MAE および MAPE の三つの評価指標を記載しており、ともに小さな値であるほど精度が良いことを示している。さらに、テストデータ全体の評価値とは別に、過去の傾向とは異なる変動に対する予測性能を評価するために、テストデータに含まれるイベント時のデータに対する評価も行った。ここでのイベントデータは、2018/10/31 16:00~23:00 の渋谷駅前セルのデータであり、ハロウィンの夜に渋谷駅周辺に急激に人々が集まった際のデータである。

まずテストデータ全体に対する予測精度を見ると、コピーモデルが悪いことから、テストデータは 1 タイムステップの間にある程度の人数変動をしているデータであることがわかる。また、平均値モデルが良いことから、人数変動は強い周期性をもつことがわかる。LSTM および ConvLSTM は、学習により変動の周期性をとらえることができており、コピーモデルや平均値モデルよりよい結果であった。これらの予測モデルのうち、MAE 及び MAPE においては、過去の変動をより直接的に活用する時系列予測モデルが最も良い結果であった。これは、祝日や天候などの周期性がくずれる外部要因の影響を受けた変動が学習データおよびテストデータに存在し外部要因の情報を含まない直近の時系列データのみを入力とした場合には、時系列予測モデルが最もうまく過去のデータを活用できていることを示している。時空間予測モデルは直近のデータのみを学習するモデルであるため、周期性をうまくモデルに組み込むことができず、ほかの予測モデルと比べると誤差が大きくなっている。提案法は、MAE および MAPE に関しては時系列予測モデルよりわずかに悪いが、RMSE に関しては最もよい精度となった。これは、時系列予測モデルと比較して、提案法は誤差の分散が小さく、大きく予測を外さないことを示している。

一方でイベントデータに対する予測誤差を見ると、テストデータ全体に対する結果と比較して、平均値モデルや時系列予測モデルがは悪いことから、学習データと大きく変動傾向の異なるデータであることがわかる。イベントデータに対する時系列予測モデルの悪さは、先ほど述べたように、テストデータ全体における RMSE が悪くなる大きな要因となっている。LSTM や ConvLSTM などの予測モデルのほうが良く、直近のデータのみを利用する時空間予測モデルはより良い。予測モデルのうち、提案法が最も良く、時系列予測モデルと時空間予測モデル単体の結果から大幅に精度が改善している。これは、提案法の枠組みの中で、時空間データの変動を通常時の変動と、通常時と異なるイベントに関する変動にそれぞれ分けてモデリングできており、これにより性能が大きく向上していると考えられる。

図 4 に、イベントデータを含む渋谷駅前セルでの提案法の予



表 1 各モデルの予測精度の比較結果. 全テストデータ (期間: 2018/10/29 0:00~10/31 23:00, 対象エリア: 500m セル 20 × 20) と, イベントデータ (期間: 2018/10/31 16:00~23:00, 対象エリア: 渋谷駅前セル) それぞれの評価を行った. 評価指標には RMSE, MAE 及び MAPE を用い, 各評価指標で最も良い精度を太字で示す.

予測モデル	全テストデータ			イベントデータ		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Copy	1,224	513	0.079	7,481	6,284	0.136
Avg.	766	265	0.038	19,488	16,076	0.316
LSTM	457	185	0.033	8,791	6,870	0.133
ConvLSTM	329	166	0.034	10,861	8,280	0.152
時系列予測モデル	327	<b>130</b>	<b>0.025</b>	13,118	11,101	0.218
時空間予測モデル	1,436	646	0.104	7,244	6,338	0.137
提案法	<b>285</b>	139	0.028	<b>2,937</b>	<b>2,214</b>	<b>0.043</b>

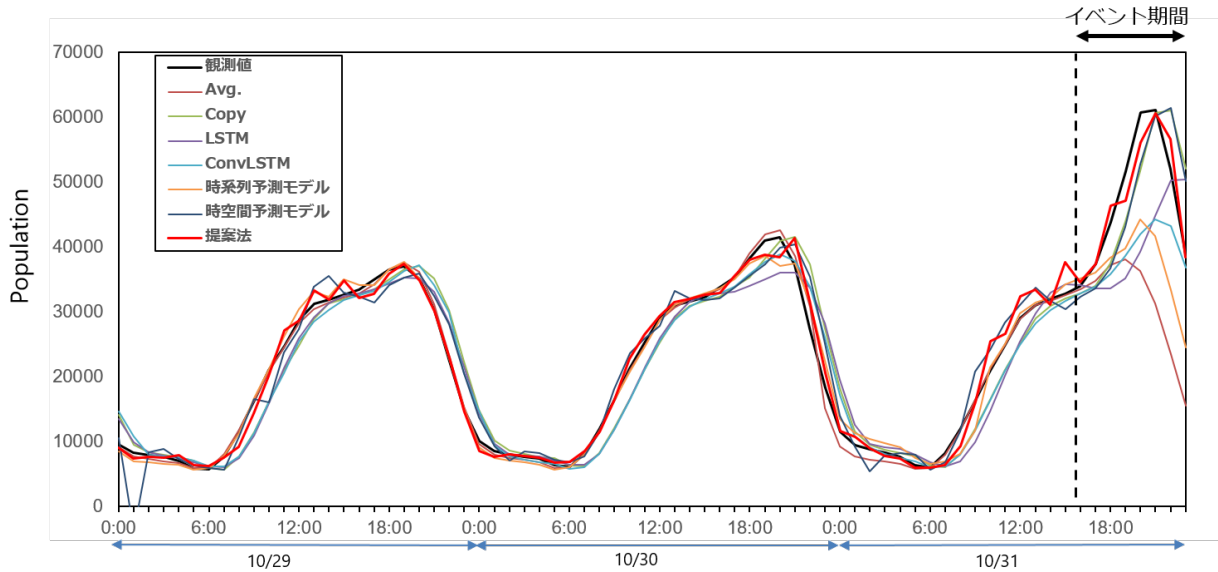


図 4 渋谷駅前セルにおける観測値と各手法の予測結果. 当該セルにおける 10/31 の 16:00~23:00 の期間の観測値がイベントデータである.

測結果のグラフを示す. 図中の観測値と平均値モデルの結果から, 当該セルにおけるイベント期間以前の変動については過去の傾向とほぼ同じ変動であり, イベント期間中は過去の傾向と大きく異なることがわかる. 時系列予測モデルも, イベント前の期間について精度よく予測ができており, イベント期間についても過去の類似するパターンを探索して利用することで, 平均値モデルよりも人数の増加に対応することができているが, 過去のデータに含まれない変動パターンを予測することはできていない. 提案法の結果を見ると, 時系列予測モデルが良い時間帯には, 残差の予測を行わないため, ほぼ同じ予測結果となっており, 悪い時間帯には残差の予測が行われるため二つのモデルの予測結果は異なる. イベント前の期間においては, 提案法は残差の変動を過敏にとらえることによって悪くなっていることもあるが, イベント期間においては時系列予測モデルの残差を埋めて観測値の急激な変動を精度よく予測することができていることがわかる.

図 5 に, イベントデータ予測時における, 剪定閾値  $k$  に様々な値を設定した際の 1 回の平均処理時間と誤差 (RMSE) を示す. 処理時間には, 新規に取得した処理開始時刻データの追加

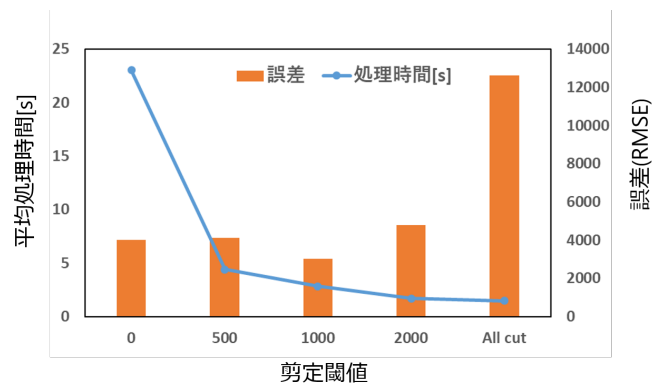


図 5 提案法でのイベントデータ予測時における, 様々な剪定閾値での平均処理時間と誤差 (RMSE).

によるモデルの更新処理と, 更新されたモデルによる 1 回の予測処理の時間を含む. 剪定閾値を 0 とした場合, 剪定処理によって削除されるデータ数は 0 であり, すべての残差データが時空間予測モデルで処理される. 時空間予測モデルでは入力となる 400 地点すべての残差データに対してモデリングを行い,

結果として誤差は小さくなるものの、非常に多くの処理時間がかかってしまっている。一方で、剪定閾値に非常に大きな値を設定することで、すべての残差データを削除した場合が図中「All cut」の結果である。この場合、時空間予測モデルによる残差回帰処理が行われず、時系列予測モデルによる予測処理のみが行われる。したがって、イベントデータに対する予測誤差が非常に大きく、処理時間は短い。今回の実験では、閾値  $k$  が 1000 であるとき最も誤差が小さいという結果であった。閾値が 0 の場合よりも誤差が小さくなった理由としては、残差の予測において余分な小さな変動のノイズが除去されたことによるものと考えられる。この時、剪定処理によって平均で 374 地点のデータが削除され、時空間予測モデルの入力地点数は 26 地点となり、剪定処理を行わない場合と比べて処理時間が 10 分の 1 程度であった。

以上の結果から、提案手法は計算量を抑えたうえで、周期性を含む過去に経験した変動の予測だけでなく、過去に経験したことのない変動についても精度よく予測を行うことが可能であることが確認できた。

## 5 おわりに

本研究では、過去に経験した変動を学習した、時系列での相関のみから予測を行うシンプルな時系列予測モデルと、直近の時空間データからこれまでに経験のない変動を予測する時空間予測モデルを組み合わせることで、予測精度を保ちながら処理全体の計算量を抑えた時空間予測手法の提案を行った。人口統計データを用いた評価実験を行い、提案法は適切な閾値で選定処理を行うことで、予測精度を保ちつつ、時空間モデルに比べて処理時間を大幅に短縮できていることを確認した。さらに、通常時よりも急激に人が集まるイベント時のデータに対し、時系列予測モデルや時空間予測モデル単体の予測精度を大きく上回る性能を示すことを確認した。今後は、より様々な状況を含む大規模な人口統計データや、人口統計データ以外の時空間データに対して手法を適用し、有効性の評価を行う必要があると考えている。

## 文 献

- [1] S. Xingjian, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting." *Neural Information Processing Systems*, 2015.
- [2] 竹内孝, 鹿島久嗣, 上田修功, "自己回帰テンソル分解による時空間データ予測," 2018 年度人工知能学会全国大会 (第 32 回), 2018.
- [3] 下坂 正倫, 和田 英之, 坪内 孝太, 築地 毅, "モバイル端末位置履歴を用いた都市動態予測のための核ノルム最小双線形ボアソン回帰" 2016 年度人工知能学会全国大会 (第 30 回), 2016.
- [4] 森本哲郎, 白浜勝太, 上善恒雄. "Wi-Fi パケットセンサを用いた人流・交通流解析の手法" 第 14 回情報科学技術フォーラム (FIT2015) 一般講演論文集, vol.14, no.4, pp.505-511, 2015
- [5] 長尾晃貴, 柳澤大地, 西成活裕. "ウェーブレット変換と機械学習を用いた雑踏密度の推定" 第 23 回 交通流のシミュレーションシンポジウム論文集, pp. 87-90, 2017
- [6] 西村友洋, 樋口雄大, 山口弘純, 東野輝夫. "スマートフォンを活用した屋内環境における混雑センシング" 情報処理学会論文

- 誌, Vol.55, No.12, pp2511-2523, 2014
- [7] 上田修功. "時空間予測技術に基づく先行的集団最良誘導", 応用統計学 (招待論文), Vol. 45, No.3, pp.89-104, 2016.
- [8] Contreras, Javier, et al. "ARIMA models to predict next-day electricity prices." *IEEE transactions on power systems* 18.3 (2003): 1014-1020.
- [9] Singh, Rayman Preet, Peter Xiang Gao, and Daniel J. Lizotte. "On hourly home peak load prediction." 2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm). IEEE, 2012.
- [10] Stock, James H., and Mark W. Watson. "Vector autoregressions." *Journal of Economic perspectives* 15.4 (2001): 101-115.
- [11] Royle, A. G., F. L. Clausen, and P. Frederiksen. "Practical Universal Kriging and Automatic Contouring." *Geoprocessing 1*: 377-394. 1981.
- [12] Hengl, Tomislav, Gerard BM Heuvelink, and Alfred Stein. "A generic framework for spatial prediction of soil variables based on regression-kriging." *Geoderma* 120.1-2 (2004): 75-93.
- [13] Cressie, N. and Christopher, W. K. "Statistics for Spatio-Temporal Data" Wiley. 2011.
- [14] Reinsel, G.C. and Velu, R.P. "Multivariate Reduced rank regression, Lecture Notes in Statistics". New York: Springer. 1998.
- [15] David Arthur, "k-means++: The advantages of careful seeding", *Proc. of the eighteenth annual ACM-SIAM symposium on Discrete algorithm*, 1027-1035, 2007.
- [16] S.Hochreiter and J.Schmidhuber, "Long short-term memory". *Neural Computation*, vol.9, no.8, pp.1735-1780, 1997.
- [17] Yunbo Wang, Mingsheng Long, Jianmin Wang, Zhifeng Gao, and S Yu Philip, "Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms", NIPS, 2017.
- [18] Yunbo Wang, Zhifeng Gao, Mingsheng Long, Jianmin Wang, and Philip S Yu. Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. ICML, 2018b.
- [19] Hehe Fan, Linchao Zhu, and Yi Yang. "Cubic LSTMs for Video Prediction." AAAI, 2019.
- [20] Lloyd, Stuart P., "Least squares quantization in PCM", *IEEE Transactions on Information Theory*, 28 (2): 129-137 (1982)