

## 地域相関を利用した状態空間モデルによる人流予測 Forecasting the people flow using state space model with regional correlation

山城 宏太<sup>†</sup> 岡崎 威生<sup>‡</sup>  
Kota Yamashiro Takeo Okazaki

### 1. はじめに

人の個人単位の予測不可能な行動に起因する人流は、現在様々な方法で研究が行われている。この流動的な人口移動を環境要因から説明するために、状態空間モデルが使用される場合がある。

本稿ではより正確な人流予測を行うために、従来の環境要因を考慮した状態空間モデルを用いた予測に加えて、予測する際に複数の地域とその地域の総人流について同時予測を行う。同時に予測を行うことによって、地域間の人の移動による時間あたりの人口への影響を体系的に表すモデルとする。これにより、単一地域で人流予測を行った場合と比べて各地域の単位時間あたり人流予測の予測精度を向上させることを目的としている。

### 2. 状態空間モデル

状態空間モデルは、物事の発生に影響を与える「状態」があることを仮定した時系列モデルの一種である。過去のデータを使用して予測するという点で同じ AR モデルや ARIMA モデルと違い、状態空間モデルでは、状態方程式と観測方程式という二つの式を使用して予測する。状態は前時点の状態を用いた予測値に過程誤差を加えた値であり、この状態に観測誤差を加えた観測値を推定する。状態として何が起因するか、過程誤差や観測誤差の大きさを設定することでモデルを推定する。状態方程式  $x_t$ 、観測方程式  $y_t$  は以下のように定式化される。

$$x_t = T_1 x_{t-1} + R_t \zeta_t, \quad \zeta_t \sim N(0, Q_t) \quad (2.1)$$

$$y_t = Z_t x_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, H_t) \quad (2.2)$$

### 3. 提案手法

#### 3.1 単一地域人流予測手法

複数地域予測の手法の前にまず単一地域の人流予測手法について説明する。予測したい地域の人流予測モデルを状態空間モデルにより定式化する。求める人流を観測方程式  $y_t$  として時間  $t$  における人口数を指揮 (3.1) に示す。周期的変動を表現する  $cycle_t$  曜日固有の特性を考慮した  $week_t$ 、気温や湿度が人流に与える影響を  $weather_t$ 、雪片  $d$  の加算によってモデルを表現する。これらの状態を逐次カルマンフィルタによって推定し、パラメータの決定には最尤推定を用いて行う。4つの環境要因については以降で詳しく解説する。

$$P_t = cycle_t + week_t + weather_t + d \quad (3.1)$$

人流データを 1 週間単位で周期を表す。周期性の要因として時間単位の人流データを 1 週間で平滑化した  $1weekma$

で扱う。周期が人流の予測にどの程度の影響を与えるかは係数  $T$  によって決定する。定式化した式を (3.2) に示す。

$$cycle_t = T \cdot 1weekma_t$$

周期性については、地域ごとで曜日による人流の大きさに地域固有の特徴が確認されている。このことから、周期成分とは別に、週に 7 日ある曜日それぞれに対して曜日特性を付与したものを曜日効果として扱う。曜日毎の人流に与える影響  $effect_{曜日,t}$  を過去の人流から推定して決定し、これを曜日効果として扱う。曜日効果を (3.3) のように定式化する。このとき、 $D_i$  をそれぞれ該当曜日からどうかについて 0 または 1 の 2 値フラグで表す。

$$week_t = \sum_{i=1}^7 D_i \cdot effect_{i,t}, \quad (0 \leq D_i \leq 1) \quad (3.3)$$

天候要因として本実験では、カテゴリカルデータである体感インデックスを用いて気温や湿度の影響を考える。体感インデックスは、人が感じた暑さや寒さの体感情報を 10 段階で表したものである。体感インデックスの詳細な内訳は Table3.1 に示す。

体感インデックス	体感情報
1	厳しい寒さ
2	寒い
3	肌寒い
4	ちょうどよい
5	暖かい
6	暑い
7	カラッと暑い
8	むし暑い
9	ジリジリ暑い
10	厳しい暑さ

表 3.1 体感情報データの内訳

この天候によって感じる感覚の情報から、これを  $heat\_index_t$  として、天候が与える人流への影響の度合いとして  $weather\_effect$  としてこれを推定する。天候要因は (3.4) のように定式化する。例えば、ちょうどよいと感じる天候の場合は、 $heat\_index_t$  の値は実験を通して、体感インデックスのそれぞれのカテゴリでどの程度の影響を受けるか 0 ~ 1 の浮動少数点数で表現する。

$$weather_t = weather\_effect * heat\_index_t, \quad 0 \leq heat\_index_t \leq 1 \quad (3.4)$$

#### 3.2 複数地域人流予測手法

地点  $n$  に対する時系列予測モデル  $P_{t,n}$  と、それら地点同士の合計を  $S_{t,n}$  として表現する。この式 (3.5), (3.6) に提示した全ての式ができる限り高い予測精度を出すように

パラメータを最尤推定により調整する。これにより、地域間の人口移動を考慮したモデルを表現する。また、評価指標として RMSE, 最大誤差を用いて単一地域のみでの予測との比較を行う。

$$P_{t,1} = cycle_{t,1} + week_{t,1} + weather_{t,1} + d_1,$$

$$\vdots$$

$$P_{t,n} = cycle_{t,n} + week_{t,n} + weather_{t,n} + d_n \quad (3.5)$$

$$S_{t,n} = \sum_{k=1}^n P_{t,k} + \epsilon \quad (3.6)$$

## 4. 人流予測適用実験

### 4.1 データセット

予測に使用する人流データとしては KDDI Location Analyzer が提供する au スマートフォンユーザの GPS 情報から得られる特定地域に 180 分以上滞在した 20 代の人口クラスを 30 分単位で取得したデータを利用した。また、天候データにはウェザーニューズから取得した体感インデックスを利用した。体感インデックスの詳細については章 3.1 の天候要因を参照。期間は 2020 年 12 月 25 日から 2021 年 1 月 31 日の 38 日間を使用した。

状態変数は全て初期値を単位行列として、観測誤差、状態誤差は初期値をどちらも零行列に設定し、最大エポック数 1000 で実験を行った。

### 4.2 複数地域を使用した予測と単一地域での予測の比較

予測結果の比較は松山地区を基準に行った。そのため、まず最初に単一地域での予測として 1 時間毎に集計した松山地区の人流と天候データをもとに、1 月 1 日から 1 月 31 日の期間について予測を行った。予測に使用した訓練期間は、初めは 12 月 25 日からの 9 日間を使用して次の 3 日間の予測を行う。それ以降は特定日の 3 日前までを訓練期間とし、同じく 3 日間の予測を行う日単位での逐次的な処理で実験した。図 4.2.1 は松山地区を同地区のみから予測を行ったときの観測値と予測値である。

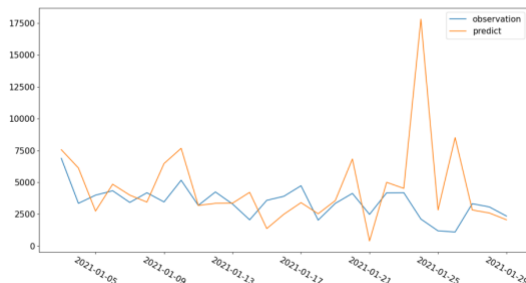


図 4.1 松山地区を単一地域から人流予測 (松山地区のみ使用)

これに対して、式 (3.5), (3.6) に導かれる予測値と観測値の誤差がそれぞれ最小化するように最適化した結果が複数地点での人流予測となる。本稿では松山地区、久茂地区の 2 つの地域から松山地区の人流予測結果を提示する。この結果を図 4.2.2 に示す。また、単一地域と複数地域からの両方

の予測値と観測値との RMSE と最大誤差を表 4.2 に示す。

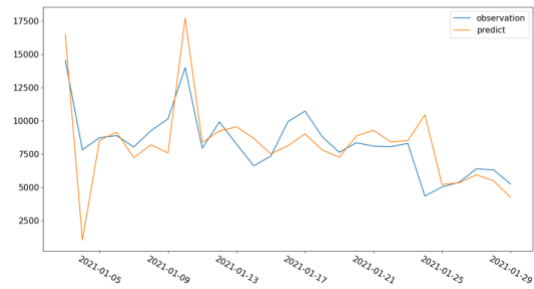


図 4.2 松山地区を複数地域から人流予測 (松山、久茂地区の 2 地域を使用)

	RMSE	最大誤差
単一地域	3637.3	15687.3
複数地域	2169.9	6772.6

表 4.2 予測結果に対する評価値の比較

単一地域で予測した場合に比べると複数地域で予測した際、表 4.2 の RMSE, 最大誤差はどちらも後者の方が良い値を示している。また、前者の予測初期は 1 日遅れで予測が行われ、中期には 1 日遅れての予測値を出しているように捉えることができる部分がある。後期は予測できているとはいえない結果になっている。後者の予測手法では、後期の予測で 1 日遅れで予測をしている部分があるが、全体的に単一地域と比較してトレンドは捉えることができている。

## 5. 今後の展望

本研究では各地点の位置情報データ、天候、周期性、曜日効果から、複数地点を同時に状態空間モデルを用いて予測することで、時間単位で集計された人流の時系列量データから複数の地点を状態空間モデルを用いて同時に予測するモデルを提案した。これにより、今まで固定地点における量として扱っていたデータに流動性を加えることで地点間の人口移動を表現し、予測精度を向上させた。今後の展望としてより多くの地点で実験を行うと同時に、地点間の指向性のあるデータを用いることによって地域間の人口移動量の正確性を高め、より正確な人流予測が可能になると考えている。

### 参考文献

- [1] 浅田 正彦, 長田 穰, 深澤 圭太, 落合 啓二, “状態空間モデルを用いた階層ベイズ推定法によるキョン(*Muntiacus reevesi*)の個体数推定”, 哺乳類科学 54:53-72(2014).
- [2] 角田 孝昭, 吉田 光男, 津川 翔, 山本 幹雄, “状態空間モデルを用いた検索トレンドとページビューからの自動車販売台数の予測”, 人工知能学会全国大会論文集 29(2015).