

ニューラルネットワークによる滞在人口の短期変動予測可能性に関する研究

早稲田大学 学生会員 ○高森 駿

早稲田大学 正会員 佐々木 邦明

1. 研究の背景と目的

人々の行動を予測するための手法として、これまでは四段階推定法や行動モデルが用いられてきた。これらは一般に、平常的な状態の予測を行う手法であり、渋滞減少や災害などの動的な現象を考慮するのは難しい。他方、IoT化の進展に伴い、多様かつ膨大なデータが利用可能となった。その中の一つが、携帯電話の通信ネットワークを活用したモバイル空間統計である。これにより、1時間単位であらゆる場所の滞在人口が概ね把握可能となった。

よって本研究では、滞在人口予測を行う際、四段階推定法等に依らない新たな手法として、ニューラルネットワーク(NN)を活用した手法の実現可能性を検討する。また、既往手法では難しいとされる動的な現象の影響をNNで考慮できるのかを検討する。

2. 研究の概要

(1)使用するデータについて

使用するデータは、2020年1月以降の毎日1時間おきに得られるモバイル空間統計とする。この期間はCOVID-19の影響により、人々の行動が日々変化していたと考えられ、この変化をNNによって捉えることが出来るのかを検証する。また、COVID-19が滞在人口に与えた影響が大きかったことが予想される新宿歌舞伎町を含む500mメッシュを対象とした。



図1 対象地域

(2)使用するNNについて

人々の行動は、通勤通学に代表されるような固定的な行動も多く、このように曜日や時間帯などによって決まる変動のパターンがあるため、滞在人口については時系列予測が適していると考えた。そこで

本研究では、時系列データを扱うことができる再帰型ニューラルネットワーク(RNN)を用いる。RNNは他のNNとは異なり、ある時刻の中間層において同じ時刻の入力層からのインプットに加えて、前の時刻の中間層からのインプットも受け取ることで、データのトレンドを学習する。本研究では、RNNの中でも記憶期間の長さの考え方を導入することにより、直近の出力だけでなく、遠い過去の出力の影響も保持することが可能なNNであるLSTMを用いる。

3. LSTMの適用

次に示す(1)(2)(3)の条件でLSTMを用いた滞在人口の予測を行った。その際、入力データは過去1週間分のデータ、すなわちウィンドウサイズを168個(24時間×7日間)とし、その直前168時点(時間)分の入力データから次の1時点の値を出力及び予測するモデルとした。ただし、(2)と(3)ではモバイル空間統計がリアルタイムで利用可能であると想定する。

(1)1～3月のデータから学習し、4月を予測

(2)1～3月のデータから学習し、4月を予測

(3)1～4月のデータから学習し、5月を予測

この結果として、各場合におけるRMSE(平均二乗誤差平方根)の値を次の表1に示す。

表1 (1)(2)(3)のRMSE

	(1)	(2)	(3)
RMSE	21923.8	3078.5	573.8

また、各日の最大値最小値を迎える時刻に着目すると、パターンの変化が誤差の原因であると考えられた。

4. 予測に関する検証

(1)入力データに変数を追加

学習、および予測を行うための入力データに滞在人口以外の項目も追加し、多変量LSTMとすることで、その予測精度の向上を試みた。滞在人口、時刻や曜日、平日休日などを加えたいくつかのパターンで予測精度を比較した用いたデータとしては、1月から4月のデータを学習に用い、5月のデータを予測した。結果として、いずれの場合も大幅な精度の

キーワード モバイル空間統計、滞在人口、ニューラルネットワーク、LSTM、COVID-19

連絡先 〒169-0051 東京都新宿区西早稲田 1-6-1 早稲田大学 TEL : 03-5286-3000 E-mail : info@sci.waseda.ac.jp

向上は見られなかった。

(2) 実用的な利用に向けて

LSTM は膨大なデータから時系列的な特徴を学習することを得意とする。それゆえに、前節までに述べた通り、時系列的特徴が変化すると精度の高い予測は困難となる。本節では、パターン等が変化した場合における LSTM を用いた滞在人口変動予測の実現手法について検討する。図 2 は、横軸の日にちの 1 時間ごとの滞在人口を LSTM で予測したときの RMSE を示したものである。その際、2020 年 1 月 1 日から予測を行う日の前日までのデータをそれぞれ学習に用いている。3 月下旬の外出自粛要請に伴い全体的に減少傾向となるが、4 月 10 日頃にパターンが変化している。実際にこのパターンの変化は図 2 にも現れており、4 月 8 日から 13 日まで RMSE の値が大きくなっているのがわかる。すなわち、RMSE 等の誤差の指標に注目し、その閾値を適切に設定することで、パターンの変化を検知することが可能と考えられる。

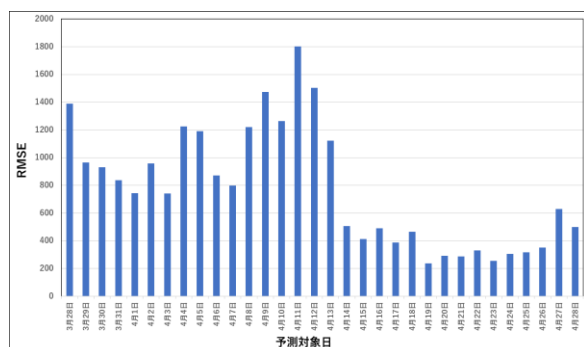


図 2 1 日ごとの RMSE

また、パターンの変化を検知した際、学習に用いるデータの期間や LSTM 内部の各種パラメータを変更して、学習および予測を行うことで、精度の高い予測の実現可能性を検討する。一例として、4 月 8 日、9 日と RMSE の値が大きくなっており、ここでパターンの変化を検知したとする。このとき、それ以降で RMSE の値が大きくなっている 4 月 10~13 日について、学習方法を変更して予測を行った。具体的には、予測対象日の前 1~3 日間分の短期間を学習に用い、またこれまでは直前 1 週間(168 時点)分のデータから予測を行っていたが、今回は直前 6 時間のデータから予測を行った。結果を表 2 に示す。結果として、パターンの変化の検知後、短期間のデータで学習を行うと、予測精度は更に低くなってしまった。この理由として、本来 LSTM をはじめとする NN は、膨大なデータからそれぞれの特徴等を学習するため、このように短期間のデータで学習を行った場合、その特徴を十分に捉えることができないからであると考えられる。

表 2 短期間学習での RMSE

学習期間	4月10日	4月11日	4月12日	4月13日
1日	4085.3	3998.1	2676.6	2176.2
2日	1663.2	1238.6	1187.4	1277.0
3日	1618.6	1160.7	1444.2	1365.7
通常	1263.8	1802.9	960.3	1121.7

5. 本研究のまとめ

LSTM をモバイル空間統計の 1 メッシュに適用し、COVID-19 の影響下での滞在人口の変動を 500m メッシュ単位で予測を行った。その結果、学習に用いるデータの期間や予測手法によってその予測の精度に大きな違いが見られた。人口総数の変化についてはある程度は柔軟に予測に反映することができるが、パターンが変化した場合は精度の高い予測が困難であった。一方で、逆に予測精度の変化からパターンの変化を検知することが可能である結果が得られた。このパターンの変化の検知が可能であることで、例えば災害時などにどこで人々の行動に直接的な影響を与えるようなことが起こっているのかを、モバイル空間統計をはじめとした観測データから導き出すことができると考えられる。今後は、パターン変化を検知した後、適切な学習方法やパターン変化を考慮した NN の構築などを通じて、精度の高い予測が行えるのかを検討する。

謝辞

本研究を行うにあたり、モバイル空間統計の貴重なデータを提供していただいた株式会社ドコモ・インサイトマーケティングに、この場を借りて深く御礼申し上げます。

<参考文献>

- 1) 渋谷剛史, 浅野周平, 十河孝介, 森本章倫, 『携帯電話基地局データを用いた立地適正化計画の評価指標に関する研究 - 宇都宮市をケーススタディとして』, 都市計画論文集, 第 53 巻, 第 3 号, pp408-415, 2008.
- 2) 佐藤大祐, 松林達史, 福田匡人, 堤田恭太, 中山彰, 戸田浩之, 『残差回帰による複合時空間予測手法』, IDIM2020(第 12 回データ工学と情報マネジメントに関するフォーラム), 16-4, day2 p74.
- 3) 斎藤康毅(2016), 『ゼロから作る Deep Learning - Python で学ぶディープラーニングの理論と実装』, 株式会社オライリー・ジャパン.
- 4) NTT docomo, 『モバイル空間統計ガイドライン』, https://www.nttdocomo.co.jp/corporate/disclosure/mobile_spatial_statistics/guideline/index.html (2020 年 11 月 20 日閲覧)