

モバイル空間統計データを用いた人口動態の異常検知と災害時への適用

東北大学工学部 学生会員 ○辻 葉月
 東北大学災害科学国際研究所 正会員 マス エリック
 東北大学災害科学国際研究所 正会員 越村 俊一

1. はじめに

台風や豪雨、地震などの災害によって被害が発生した際には、更なる被害の防止や迅速な復旧のために適切な対応が求められる。それは例えば避難所の開設であったり、人員の派遣であったりと、実際の形態としては様々なものが考えられるが、これらに共通して被災状況に関する情報が求められる。昨今では技術の進歩によって、様々な情報が容易かつリアルタイムに入手可能となっており、実際の緊急対応にも数多く用いられている。特に、スマートフォンを始めとする携帯端末をもとにした人口データは、近年での地域や年代を問わない幅広い普及と、これまでの人口データ(国勢調査など)と比較してデータ収集が高頻度に行われることによるリアルタイム性より、様々な解析や調査などにおいてその重要性を増してきている。

本研究は、人口データの一つであるモバイル空間統計(提供:(株)ドコモ・インサイトマーケティング)に対して異常検知を行うことにより、災害など有事の際の人口動態の特徴を分析することで、緊急対応におけるモバイル空間統計データの活用性を高めることを目的とする。

2. データ

(1) モバイル空間統計

本研究では、(株)ドコモ・インサイトマーケティングが提供するモバイル空間統計を人口データとして用いた。このデータは、各基地局エリアごとの携帯電話数を1時間ごとに集計し、ドコモの普及率を加味して人口を推計した統計データである。なお、統計処理によって、携帯端末利用者個人を特定することは不可能になっている。

使用したデータは2019年1月1日0時から2021年12月31日23時までの3年間分、メッシュサイズは500mである。

(2) 対象地域

解析の対象地域について、図-1に示す。これらの地域は、2019年から2021年にかけて災害の影響を受けたか、イ

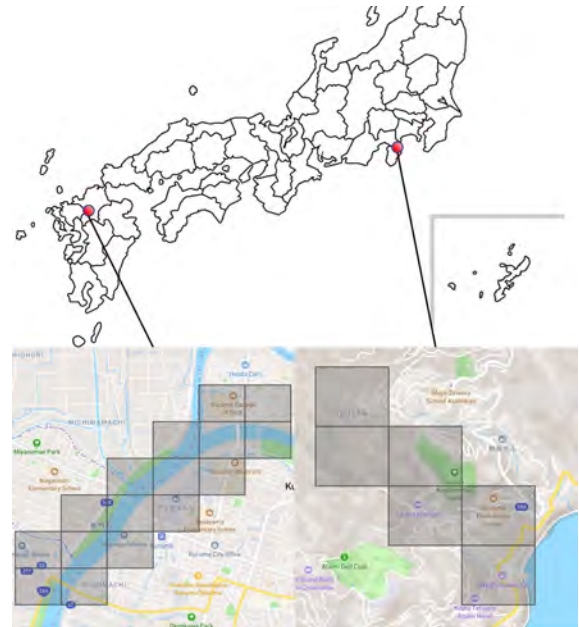


図-1 対象地域

ベントが開催されたなどの理由で、通常とは異なる人口の移動があったと見込まれる地域である。

3. 手法

(1) Matrix Profile (Yeh *et al.*, 2016)

ひとつもしくはふたつの時系列データについて、部分配列同士のユークリッド距離を用いて各部分配列の異常度を求める方法。該当する部分配列において特徴的な変化が見られる場合、Matrix Profileの値は大きくなる。

(2) Local Outlier Factor (Breuing *et al.*, 2000)

データセット内のオブジェクトについて、それ自身と近傍それぞれについて周辺の密度を求め、それを比較することでオブジェクトの異常度を求める。オブジェクト周辺が近傍と同程度の密度であるほど1に近くなり、近傍と比べて疎であるほど1より大きくなる。

(3) 解析方法

異常度の計算には8週間前までのデータを使用し、Matrix Profileの計算には期間内のすべてのデータを、Local Outlier Factorの計算には期間内の同じ曜日のデータを使用した。

4. 結果と考察

(1) 災害への適用

災害への適用の例として、熱海市伊豆山地区での解析の結果を図-2に示す。なお、図の上段が1時間ごとの人口データ、下段の青いグラフがMatrix Profile(左軸)、赤いグラフがLocal Outlier Factor(右軸)を示す。この地域では2021年7月3日に台風による土石流災害が発生し(図の黄色の破線)、大きな被害が発生した。図-2は、災害当日を含む2021年6月19日から8月14日までの期間を示す。

異常度を見ると、どちらの指標においても災害直後に高くなっていることがわかる。その後、両者とも全体的には減少しながら、災害以前と同程度の値へと戻っている。どちらの指標においても、異常度の計算においてはその地点から一定期間の過去のデータを使用している。そのため、災害直後に比べて、災害発生から少し経過した場合は、計算に使用するデータの中に含まれる災害の影響を受けたデータの割合が高くなる。これは、計算に用いるデータ中の、災害の影響を受けたデータの異常性が低くなると言い換えることができる。

(2) イベントへの適用

イベントへの適用の例として、筑後川花火大会会場に対する解析の結果を図-3に示す。筑後川花火大会は、福岡県久留米市・佐賀県鳥栖市の筑後川流域で行われる花火大会で、毎年8月に開催される。図-3は花火大会当日(2019年8月5日、図の黄色の破線)を含む2019年7月22日から8月19日の期間を示す。

異常度を見ると、どちらの指標も破線付近で著しく高くなっていることがわかる。高い値を取る範囲が異なるように見えるが、これは計算方法の違いによるものである。Matrix Profileでは1時間ごとにデータのウィンドウをずらしながら計算する。そのため、今回の花火大会のような短時間の異常であっても、異常な点を含むウィンドウ全てと対応した部分の異常度が高くなる。Local Outlier Factorでは、今回の計算方法では曜日ごとに異常性を比較するため、出てくる結果も日毎になる。そのため、異常が存在する場合には、その点を含む1日の異常度が高くなる。

(3) 課題

今回用いたどちらの指標においても、異常と通常の明確な境界は定義されていない。特にMatrix Profileでは、スケールの異なるデータを使用すると結果のスケールも

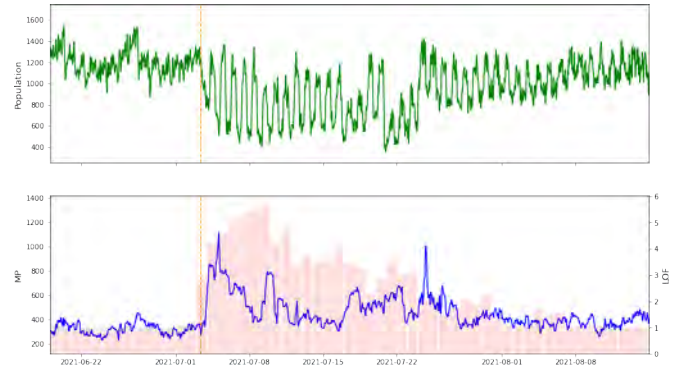


図-2 熱海市伊豆山地区の人口(上段)と異常度(下段)

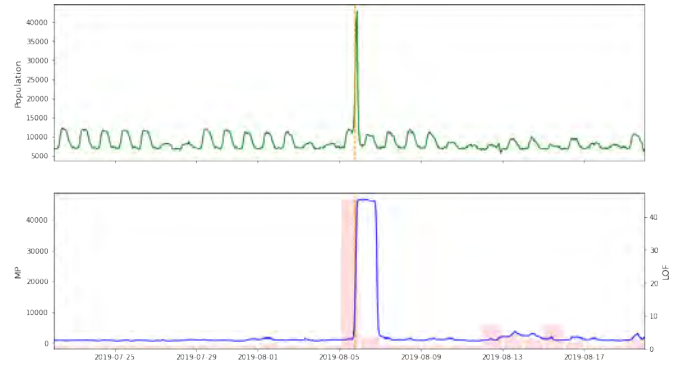


図-3 筑後川花火大会会場の人口(上段)と異常度(下段)

対応して異なるため、一定の値を閾値として利用するのは難しい。適切な閾値の決定による異常と通常の明確な判別が今後の課題と言える。

5. おわりに

本研究では、モバイル空間統計データに対してふたつの指標を用いた異常検知を行った。その結果、災害やイベントなどの異常な人口動態に対してどちらの指標も妥当な反応を示した。異常の定義と、それを判定する適切な閾値の決定が、今後の課題である。

参考文献

- Yeh, C. -C. M. and Zhu, Y. and Ulanova, L. and Begum, N. and Ding, Y. and Dau, H. A. and Silva, D. F. and Mueen, A. and Keogh, E. Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets, 2016 IEEE 16th International Conference on Data Mining (ICDM), pp. 1317-1322, 2016.
- Breunig, Markus M. and Kriegel, Hans-Peter and Ng, Raymond T. and Sander, Jörg. LOF: Identifying Density-Based Local Outliers, Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Association for Computing Machinery, pp.93-104, 2000.