

深層学習・ビッグデータを用いた人獣共通感染症パンデミックの要因解明 および時空間分布予測

安納住子*、平川翼**、安本晋也**、杉田暁**、李明安***、佐々木善信****、大吉慶****

*上智大学、**中部大学、***国立台湾海洋大学、****宇宙航空研究開発機構

1. はじめに

デング熱や新型コロナウイルス感染症(以下COVID-19)のように動物から人へ伝播可能な人獣共通感染症は増加傾向にあり、対策が必要とされている。しかし、開発途上国では財政負担が困難であることや、安全性・有効性の懸念により未だ治療薬やワクチンがないことから、効率的な介入と予防効果の高い感染症対策の確立が急務となっている。

本研究では、ビッグデータおよび深層学習を応用し、368 の町からなる台湾全島におけるデング熱の発生分布および 189 の国・地域における COVID-19 のパンデミック(世界的大流行)に起因する問題複合体の解題と、予防策としてこれらの発生分布および流行を予測するモデルを開発することを目的とする。

2. 方法

ネットワークモデル：デング熱

台湾のデング熱発生分布を時空間で予測するため、U-Net を用いる。U-Net ベースの encoder-decoder モデルは、encoder 部で画像に対し convolution と downsampling を繰り返し適用し高次の特徴抽出を行い、decoder 部で convolution と upsampling を繰り返し適用して特徴マップの解像度を向上させ、目的の領域抽出を行う。また、同一解像度の特徴マップを encoder から decoder に受け渡すことにより正確な定位を可能にする(1)(図1)。

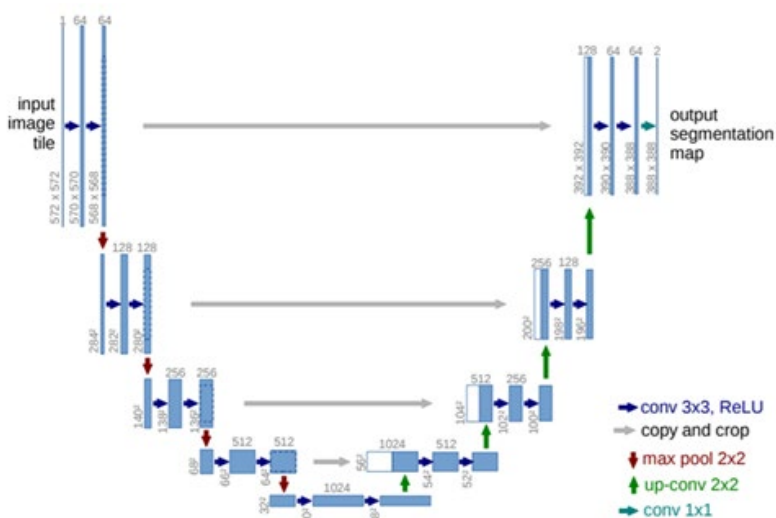


図1. U-Net アーキテクチャ

ネットワークモデル：COVID-19

国・地域別の COVID-19 発生分布を時空間で予測するため、グラフ畳み込みネットワーク(Graph Convolutional Network: 以下GCN)を用いる。グラフ構造データ解析の基本的タスクには、ノード分類、リンク予測、グラフ分類などがある。GCN はグラフデータに対して畳み込み(Convolution)を行う手法であり、レコメンデーションシステム、化合物の物性推定、交通・物流予測などの課題を解決するのに適している(2, 3)。

データセット：デング熱

U-Net に入力するデータは、次のように収集および加工して用いる。2002/4/16 から 2020/9/30 までに観測された日毎のリモートセンシングデータ：米国地球観測衛星(NOAA-12~16/AVHRR) | 海面水温度(4)、衛星全球降水マップ(GSMaP) | 降雨量(5)、米国地球観測衛星(Terra・Aqua/MODIS) | 地表面日射量(光合成有効放射量)(6)は宇宙航空研究開発機構(JAXA)およびアメリカ航空宇宙局(NASA)が公開している HP からそれぞれ取得した。同様に、2002/4/16 から 2020/9/30 までに収集された日毎・座標値(患者が居住する町の中心座標)を含むデング熱のサーベイランスデータは台湾 CDC の HP から取得した(7)。

取得したリモートセンシングデータの欠損値を補間、[0, 1]に正規化、可視化を行い、上記の日数に相当す

る枚数の画像を作成した。本稿では、U-Net へ入力する画像を次の 5 通り：1)海面水温度+降雨量+地表面日射量、2)海面水温度、3)降雨量、4)地表面日射量、5)降雨量+地表面日射量とした。入力画像に対する正解画像は、デング熱のサーベイランスデータを用い、ピクセルにデング熱発生有無を示す 0 または 1 のラベルを付与して作成した。なお、対象期間を通して 1 日毎のデング熱の症例数がほとんどないため、2 週間単位に丸めて実験に使用した。

データセット：COVID-19

GCN に入力するデータは、次のように収集および加工して用いる。COVID-19 流行に関与している空港の位置および空路のデータは OAG 社から提供されたものを用いた(8)。また、鉄道の路線と各国の国境データは Natural Earth のウェブページから取得した(9)。最後に、「地理的に国境が隣接している場合にはその 2 国間には道路が通っている」という前提を元に、各国間の道路による接続性のデータを作成した。さらに、2020/1/1 から 2021/2/11 まで収集された日毎の COVID-19 のサーベイランスデータも Our World in Data の HP から取得した(10)。

取得したデータを用いて有向・無向グラフを作成し、各ノード/エッジに属性(特徴量)を付与し、GCN へ入力するデータセットを作成した。

実験：デング熱

5 通りの画像を入力し、各ピクセルにおけるデング熱発生有無を予測する実験を行った。作成された 2002/4/16 から 2020/9/30 までのデータのうち、2002/4/16 から 2017/12/31 までを学習データ、2018/1/1 から 2018/12/31 までを検証データ、2019/1/1 から 2020/9/30 までを評価データとし、U-Net の実装には PyTorch を用いた。学習パラメータは、エポック数を 2000、バッチサイズを 8、最適化関数を Stochastic Gradient Descent (SGD) with Momentum、学習率を 1×10^{-2} 、モーメンタムを 0.9 とした。評価指標として、2 週間単位で丸めたデング熱発生分布の正解画像と U-Net が出力した画像との間の Overall Accuracy、Mean Accuracy、Freq W Acc、Mean Intersection over Union (IoU)、Dice 係数を用いた。

実験：COVID-19

グラフ構造の時空間データセット構築後、空港・空路・鉄道・陸路データと GCN を用い、地域間の関係を考慮した世界の国・地域における COVID-19 の流行予測を行う。

3. 結果

抽出結果：デング熱

海面水温度、降雨量、地表面日射量を組み合わせた 5 通り：1)海面水温度+降雨量+地表面日射量、2)海面水温度、3)降雨量、4)地表面日射量、5)降雨量+地表面日射量、それぞれを U-Net へ入力・出力した画像および正解画像を図 3a-3f に示す。平均 Dice 係数はどの組み合わせでも低かった。強いて言えば、地表面日射量においてやや良い精度が出ていた。



図 3a



図 3b



図 3c

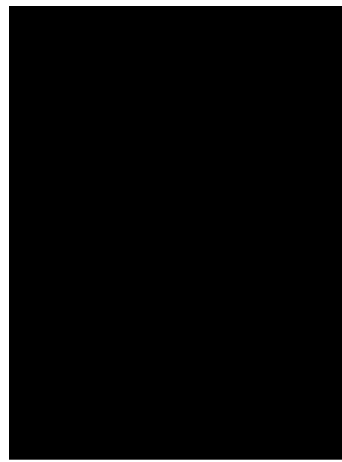


図 3d

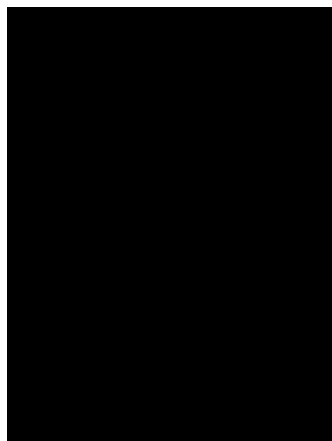


図 3e

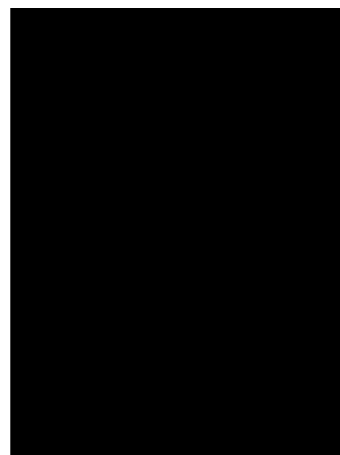


図 3f

図 3. 抽出結果の抜粋。同日のデング熱発生有無を白黒の 2 値で可視化している。2 週間単位で丸めたデング熱発生分布の正解画像(図 3a) 海面水温度+降水量+地表面日射量を入力し U-Net が出力した画像(図 3b) 海面水温度を入力し U-Net が出力した画像(図 3c) 降水量を入力し U-Net が出力した画像(図 3d) 地表面日射量を入力し U-Net が出力した画像(図 3e) 降水量+地表面日射量を入力し U-Net が出力した画像(図 3f)

4. 考察

考察：デング熱

今回の実験において、精度評価がどの組み合わせでも低くなった理由としては、海面水温度のデータが陸域に欠損していたことや、抽出対象が小さく学習時に情報が失われたため(11)と考えられる。

5. まとめ

COVID-19 の感染流行予測に向けて、グラフ構造の時空間データセットを構築中である。構築後、GNN を用いた実験に加えて、静的・動的的外部要因を考慮した属性拡張時空間グラフ畳み込みモデル (AST-GCN) を用いた実験も行い、精度の高い COVID-19 感染流行予測モデルの開発を目指す。

6. 謝辞

本研究は中部大学問題複合体を対象とするデジタルアース共同利用・共同研究 IDEAS202008 の助成を受けたものです。

参考文献・データ

- (1) Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention-MICCAI, Springer, pp. 234-241.
- (2) 立花誠人, 村田剛志 (2019) 構造特徴とグラフ畳み込みを用いたネットワークの反教師あり学習. 人工知能学会論文誌, 34 巻 5 号 B, 1-8.
- (3) Zhang S, Tong H, Xu J, Maciejewski R (2019) Graph convolutional networks: a comprehensive review. Comput Soc Netw 6, 11. <https://doi.org/10.1186/s40649-019-0069-y>.
- (4) Simons RA.: ERDDAP, <https://coastwatch.pfeg.noaa.gov/erddap>.
- (5) JAXA Global Rainfall Watch, <http://sharaku.eorc.jaxa.jp/GSMaP/index.htm>.
- (6) JAXA Satellite Monitoring for Environmental Studies, <http://kuroshio.eorc.jaxa.jp/JASMES/index.html>.
- (7) Taiwan CDC, <https://data.gov.tw/dataset/21025>.
- (8) ORG, <https://www.oag.com/ja/airline-schedules-data>
- (9) Natural Earth, <https://www.naturalearthdata.com/>
- (10) Our World in Data, <https://ourworldindata.org/covid-cases>
- (11) Yan J, Wang H, Yan M, Diao W, Sun X, Li H. IoU-Adaptive Deformable R-CNN: Make Full Use of IoU for Multi-Class Object Detection in Remote Sensing Imagery. Remote Sensing. 2019; 11(3):286. <https://doi.org/10.3390/rs11030286>