

機械学習を用いた降雨の直前予測

北海道大学 大学院理学院 宇宙理学専攻

野田 明羅

(学籍番号：20213064)

2022年12月6日

概 要

ゲリラ豪雨や台風に伴う降雨帯、線状降水帯といった極端降雨気象の発生や降水量を正確に予測することは難しい。これらの降雨領域の狭さや発達時間の速さが、予測の困難さの原因になっている。近年、計算機の処理速度の向上やビッグデータの登場の恩恵を受けて、機械学習技術の発展が著しい。加えて、気象学分野においても、機械学習を用いた手法が盛んに研究されている。Aifang et al. [2020] は、従来の予測手法であるオプティカルフロー法と機械学習を用いた予測手法の精度を比較し、結果的に機械学習を用いた手法の方がよい精度で得られたということが示された。Cul-Min et al. [2020] は、従来の予測手法に機械学習を組み合わせたことで、予測精度が向上したことが、結果で示された。しかし、依然として課題は残っている。上記に挙げた先行研究でも、予測時間が延びるにつれ予測値と実測値との誤差が多くなることや、上手に予測ができない場合があったと述べられている。我々の研究グループは、2017 年から P-POTEKA と呼ばれる自動気象観測装置の導入を進めてきた。豪雨やそれに伴う洪水の被害が多発しているマニラ（フィリピン）の首都圏に、現在 35 個の P-POTEKA が設置されていて、降水量や気温、気圧、湿度、風速、風向、太陽の放射照度を 1 分毎に観測している。これにより、極端降雨現象を捉えるのに適した、時間的に、そして特に空間的な解像度の良いデータを得ることができるようになった。P-POTEKA のデータに加え、通常クリギング手法と呼ばれるデータの空間的な自己相関から未知の地点におけるデータを近似計算する内挿手法を用いて、データを内挿することで、降雨の時系列画像データを作成した。2020/04 ~ 2020/11 の期間で、約 1 万 5 千枚の画像を作成した。本研究で用いる機械学習モデルは、時系列画像データを学習させるのに適した ConvLSTM (Convolutional Long-Short Term Memory) と呼ばれるモデルである。このモデルは、画像認識などの画像学習で目覚ましい成果を上げている CNN (Convolutional Neural Network) と、時系列データを学習するのに適した LSTM (Long-Short Term Memory) を組み合わせたモデルで、Xingjian et al. [2020] が 2015 年に提唱した機械学習モデルである。上記の降雨の時系列画像データを ConvLSTM (Convolutional Long-Short Term Memory) と呼ばれる機械学習モデルに学習させることで、一時間後までの降水量と降水分布の短時間予測を行った。予測画像と正解画像の対応するピクセルの RGB 値の平均平方二乗誤差 (RMSE) を、未学習の降雨 (40 パターン) に対して計算して、本研究の機械学習手法の性能の評価を行った。まず、今回用意できたデータにおける、機械学習モデルの最適な学習回数を調べた。機械学習における学習回数とは、同じデータを繰り返し学習させる回数である。基本的に、学習回数が多いほど、モデルの性能は上がっていくが、多すぎると過学習を起し、学習 2 データに対しては予測

が上手にできるが、未知のデータに対しては予測が上手にできなくなる場合がある。本研究では、200、300、400、500、1000 回の 5 種類の学習回数の結果を比較した。200～500 回までは、平均平方二乗誤差が小さくなっていったが、1000 回になると逆に大きくなり、モデルは過学習を起こした。次に、画像の配色を変えることでモデルの性能に影響があるかどうかを確認した。3 つの配色パターンで、同じデータを学習させて、平均平方二乗誤差を計算した。結果的に、配色の変化は、モデルの性能を変化させることが分かった。最適な配色があるのか、本研究の結果からは判断できないため、今後の研究でより多様なパターンを試す必要がある。さらに、データの学習のさせ方を変化させることで、モデルの性能が変化するかどうか確認した。1 時間離れた画像を予測する学習方法と、10 分離れた画像を予測する学習方法の 2 パターンを比較した。結果、後者の方が、モデルの性能が良いことがわかった。これは、後者の学習方法の方が、より時間解像度が高いので、降雨の変化が上手に学習させやすいためと考えられる。また、より詳細にモデルの性能を評価するために後者のモデルの、3 つの降雨に対するケーススタディを確認した。予測画像の、平均平方二乗誤差の時間変化を見ると、前半 30 分までの誤差は比較的小さいが、後半 30 分の誤差は大きくなり、予測を大きく外す場合が多くみられた。結果として、実用に足るような高い精度での予測は実現できなかった。考えられる要因としては、まず学習データの量が十分でないことが挙げられる。一般的な、画像の機械学習では 6 万枚前後の画像データが用いられることが多い。さらに、降雨のデータのみを学習させているので、本研究の機械学習モデルは予測時のインプット画像の降雨トレンドに、強く影響されやすいことも大きな要因の一つである。すなわち、予測時に雨が強まっていると、予測も雨は強くなり、逆に弱まっている場合、予測も弱まる。一方で、強まっていた雨が急速に止む場合や、雨が降っていなかったのに、急に降り出し強まる場合、予測が上手にできない。機械学習モデルに気温や気圧、風などの情報を、降雨データと合わせて学習させることで、この弱点を克服できる可能性がある。今後、使用可能なデータは時間とともに増加する。さらに、使用するモデルのパラメータの最適化や、学習させるデータの種類を増やすことでさらなる予測精度の向上に向け、研究していく予定である。

1 章のタイトル

この文書はサンプルです .

1.1 節のタイトル

この文書はサンプルです .

1.1.1 項のタイトル

この文書はサンプルです .

2 sample

Hello World!

参考文献

- [1] Aifang Su, Han Li, Liman Cui and Yungang Chen, “ A Convection Nowcasting Method Based on Machine Learning, ” Hindawi, China, 2020.