

深層学習を用いた太陽フレアの予測と実利用化に向けて Solar Flare Prediction using Deep Neural Networks and Practical Application

西塚 直人
Nishizuka Naoto

情報通信研究機構

National Institute of Information and Communications Technology (NICT)

地球周辺の宇宙プラズマや太陽プラズマは、複数の人工衛星搭載の観測機器によって、様々な物理量がほぼリアルタイムで取得可能な時代になってきた。太陽フレアや太陽風によるX線放射、太陽高エネルギー粒子、そして地磁気擾乱(宇宙嵐)が地球インフラに影響を与えることがあり、これらの観測データは基礎科学の解明とともに、宇宙環境の監視や宇宙天気予報にも利用されている。

太陽フレアの予測は、宇宙天気予報にとって最も重要な課題の一つであり、今まで多くの研究者が発見機構の解明と、予測精度の向上に挑戦してきた。従来、(i)人手による経験則に基づいた予測、(ii)統計的手法による予測、そして(iii)物理モデルに基づいた数値シミュレーションによる予測といった3つの予測研究が行われてきた。そして近年、第4のアプローチとして、ビッグデータ化した宇宙観測データを用いた(iv)機械学習による予測手法が開発されつつある。この第4の手法は現在最もホットな研究課題の1つであり、既にいくつかのモデルが予測精度の向上に成功している。最前線では、深層学習モデルの太陽フレア予測への応用が進みつつある。深層学習は機械学習アルゴリズムの1つで、一般には現状最大の予測精度を達成することができるモデルである。その中でも、畳み込みニューラルネットは画像から自動的に特徴量を抽出することができ、深層学習の社会応用を加速させた。その一方で、予測理由を人が理解できるように説明できないという課題も併せもっている。

本講演では、まず我々が開発した深層学習を用いた太陽フレア予測モデル、Deep Flare Net (略称 DeFN)について紹介する。このモデルは、太陽面上の各黒点領域で、今後24時間以内に起こる太陽フレアの発生確率を求めることができ、最も起こり得るような太陽フレアの規模を予測することができる。太陽フレアの規模とは、地震のマグニチュードのようにエネルギーが1桁変わるとにクラス分けされ、最大規模のものをXクラス、中規模のものをMクラス、小規模のものをCクラスと呼んでいる。NASAの太陽観測衛星SDOによって2010-2015年の間に観測された約30万枚の太陽画像をもとに、我々はまず黒点領域を検出し、次に各領域に対して黒点面積や磁場強度、ローレンツ力、軟X線強度、フレア発生履歴といった79個の物理的な特徴量を計算した。そ

してフレアの起こった領域には、X・M・Cクラスのラベルを添加し、教師データを作成した。特徴量としては、Nishizuka et al.(2017)で用いた物理特徴量に加え、より運用的な予測を行うため、1千万度以上の高温コロナから放射される131Å極端紫外線強度と発光面積、さらに131Åと軟X線放射の1、2時間前データの履歴を特徴量として追加した。さらに、運用的な予測評価を行うために、我々はデータベースを時系列に分割して学習用とテスト用のデータセットを作成した。ここでは2010-2014年を学習用、2015年をテスト用とした。本深層学習モデル(DeFNモデル)はバッチ正規化やスキップ接続といった手法を組み込むことによって、深い多層ニューラルネットワーク構造を達成した。

予測モデルの学習では、True Skill Statistics (TSS)というスキルスコアを最大化するよう最適化した。その結果、従来人手による予測では3~5割程度だった予測精度を、8割程度まで向上させることに成功した。より具体的には、Mクラス以上に対してTSS=0.80、Cクラス以上に対してTSS=0.63を達成した。ここで、通常の深層学習モデルでは予測過程はブラックボックスであるのだが、本DeFNモデルでは物理的な特徴量を手動で計算しており、どの特徴量が予測に効果的であるかを解析することができるのが特長である。

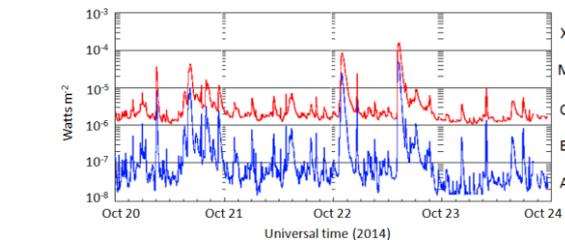
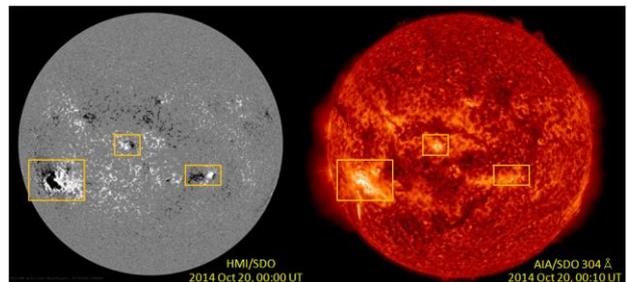


図1:(上)太陽観測画像(SDO衛星/NASA)と(下)軟X線光度曲線(GOES衛星/NOAA:ピーク時がフレア)

以上の我々のDeFNモデルの予測精度向上にとって、多波長観測データの利用、新たな特徴量の追加、予測アルゴリズムの性能向上の3点が重要である。まず1つのプロトタイプモデルを作成し、予測精度を評価検証しながら、データの改善、特徴量の追加変更、より良い予測アルゴリズムの選択を考える。さらにこの過程の繰り返しで、より優れた予測モデルを生成する。深層学習はそれ自身、学習の中で最善の予測モデルを作成する。また特徴量の重要度ランキングを調査することで、網羅的にどの特徴量が予測に有効であるか明らかになり、改善策を考えることが可能である。これらは、従来の仮説から理論モデルを考えて数値シミュレーションを行ってみたり、逆に太陽フレアが起こった際の観測データを詳細に調べてみたりするイベント解析とは異なる手法である。このデータ駆動型のモデル開発によって、太陽フレア予測モデルの精度向上が可能になってきたと言える。

ここでデータ駆動型の研究開発において、いくつかの循環ができていくのに気づく。まず精度向上には、観測家、データ解析家、そして機械学習(モデル)専門家の循環がある。また、研究開発初期は研究者間にて予測モデルの実行・評価検証が行われるが、次第に予報運用者間で予測・評価・改善が行われ、循環の輪が大きくなる。さらに社会実装されることで、予報ユーザー(事業者、例えば衛星通信会社、航空会社、電力会社など)レベルで予測・評価検証・改善が行われる。研究者、予報運用者、ユーザー(事業者)の立場によって目的や需要が異なり、モデルの最適化手法も変わってくると予想される。現在、我々の予測モデルは予報運用へ移行中であるが、社会実装していく中で、こうした循環を上手く取り組むことが重要になるかもしれない。

引用文献

- (1) Nishizuka, N. et al. 2017, *Astrophysical Journal*, 835, 156
- (2) Nishizuka, N. et al. 2018, *Astrophysical Journal*, 858, 113