

M47a Gradient Boosted Trees による太陽フレア予測と時系列交差検証

西塚直人、杉浦孔明、久保勇樹、石井守 (情報通信研究機構)

我々は、太陽観測データと機械学習を用いた太陽フレア予測モデルを開発してきた。太陽全面画像をもとに、各活動領域毎の光球ベクトル磁場、彩層発光、X線活動度といった特徴量を抽出し、今後24時間以内に発生する最大クラスのフレアを予測できるモデルを開発した(日本天文学会2016年春季年会 M44a 西塚ら)。しかしながら、予測モデルの標準的な評価手法は未だ確立されていない。また、予報運用を念頭においた学習・検証データが完全独立な厳しい条件下で予測を行った場合には、我々の開発モデルでもまだ十分な精度が出ていない。

本講演では、予報精度の検証を行うため、従来のフレア予測で用いられてきたK-分割交差検証ではなく、時系列交差検証という手法を新たに導入した。どちらもある視点では合理性があるが、予報運用の目的に対しては時系列交差検証の方が適している。さらに、Gradient Boosted Treesという学習アルゴリズムを新たに使用した。ブースティングは、損失関数を最も下げるように逐次的に弱学習器(ここでは決定木)を追加していく手法で、各ステップ毎のパラメタ最適化の際に勾配を求めて学習しなおすことで、より良い予報精度を達成する。我々はそれを太陽フレア予測に応用して、時系列交差検証を行った。その結果、TSS (True Skill Statistic) で評価した場合、Xクラスフレアに対して0.6、Mクラスフレアに対して0.8まで向上することが示された。また他の5つの学習アルゴリズムとも比較を行った結果、交差検証の手法の違いで順位が全く異なってしまうこともわかった。