

Predição de explosões solares de intensidade elevada usando séries temporais e *ensembles* de árvores de decisão

Tiago Cinto, André L. S. Gradvohl, Guilherme P. Coelho e Ana E. A. da Silva
Faculdade de Tecnologia – FT, Universidade Estadual de Campinas – UNICAMP

Explosões solares são um dos fenômenos mais relevantes apresentados pelo Sol, pois tratam-se de liberações repentinas de grandes quantidades de radiação que podem afetar a atmosfera terrestre e causar danos em diversos aspectos, como a aviação, o espaço aéreo e os satélites, levando à grandes perdas econômicas. É de extrema importância, portanto, a criação de sistemas que possam prever a ocorrência das explosões. A literatura emprega diversas técnicas na predição destes fenômenos; entretanto, mais estudos são necessários em busca de se obter sistemas com melhores desempenhos (BARNES et al., 2016). Assim, este trabalho tem por objetivo propor um modelo de predição de explosões baseado em séries temporais e *ensembles* de árvores de decisão, entregando contribuições ao desempenho de estimação e ao horizonte de predição de fenômenos das classes M e X. Os dados usados no presente trabalho foram extraídos dos repositórios mantidos pelo *Space Weather Prediction Center* – SWPC. Com isso, duas bases de dados foram consideradas, a saber, *Daily Solar Data*, com registros de observações diárias de comportamentos do Sol, e *Sunspot Region Summary*, com um detalhamento maior das manchas solares registradas na primeira base – no caso, descrevendo suas classes magnéticas (taxonomia de *Mount Wilson*). Assim, foram extraídos dados do período compreendido pelos anos de 1997 à 2017, em um total de 13 atributos distintos: *radio flux 10.7cm*, *sesc sunspot number*, *sunspot area*, *x-ray background flux* e 8 atributos binários representando as classificações magnéticas das manchas, além do registro da ocorrência de explosão de classe M ou X. Os primeiros 12 atributos foram modelados segundo uma série temporal com janela deslizante de 5 dias de atraso, enquanto o décimo terceiro foi projetado como o horizonte de predição (considerou-se 24 horas de antecipação ao fenômeno). O modelo de predição desenvolvido foi uma *Gradient Boosting Machine* contendo 50 árvores de decisão internas como estimadores. Como metodologia para construção do sistema, adotou-se o cálculo univariado de F-score para a seleção dos atributos pertinentes e a técnica de SMOTE-ENN para *data augmentation*. Além disso, validação cruzada foi usada para aferir o desempenho médio de generalização do modelo sobre dados reservados tidos como de operação. Como resultados parciais obtidos, a implementação e avaliação do modelo resultou em escores de desempenho na ordem de TPR: 0,84 (taxa de acertos das classes positivas), AUC: 0,83 (área sobre a curva de operação característica do classificador) e TNR: 0,68 (taxa de acertos das classes negativas). O escore AUC indica a presença de falsos positivos entre as predições sinalizadas como explosões. Barnes et al. (2016) sugerem ser desejável apresentar falsos positivos entre as predições marcadas como explosões em detrimento de não identificar a ocorrência do evento (falso negativo). Embora TPR promissor foi alcançado, mais trabalhos são necessários em busca de patamares superiores para os demais resultados, principalmente quanto à detecção das classes negativas e a redução dos falsos alarmes.

BARNES, G. et al. Comparison of flare forecasting methods. Results from the “All-Clear” workshop. *The Astrophysical Journal*, v. 829, n. 2, p. 89, 26 set. 2016.