



Contribution ID: 56

Type: Poster

Classifying X-ray sources with Supervised Machine Learning: Challenges and Solutions

Millions of serendipitous X-ray sources have been discovered by modern X-ray observatories like Chandra, XMM-Newton, and recently eROSITA. For the vast majority of Galactic X-ray sources the nature is unknown. We have developed a multiwavelength machine-learning (ML) classification pipeline (MUWCLASS) that uses the random forest algorithm to quickly perform classifications of a large number of sources to learn about their astrophysical nature. This approach enables quick follow-up observations of interesting sources and population studies of various kinds. MUWCLASS has been applied to Chandra Source Catalog and XMM-DR13 catalog, augmented with multiwavelength properties obtained by cross-matching to surveys performed at other wavelengths. In this talk, I will demonstrate and discuss some common obstacles encountered in supervised ML (e.g., biases between training data and unclassified data, imbalanced training data, missing values, high-dimensionality) in the context of X-ray source classification. I will also present recent developments we have implemented to address some of those issues (e.g., astrophysically-motivated oversampling, accounting for feature uncertainties and absorption/extinction biases, probabilistic cross-matching and probabilistic class inference).

Primary author: Mr YANG, Hui (The George Washington University)

Co-authors: Dr KARGALTSEV, Oleg (The George Washington University); Dr HARE, Jeremy (NASA GSFC); Mr CHEN, Steven (The George Washington University); Dr VOLKOV, Igor (Whitespace); Dr RANGELOV, Blagoy (Texas State University); Mr LIN, Yichao (The George Washington University)

Presenter: Mr YANG, Hui (The George Washington University)

Session Classification: Posters

Track Classification: Online