MACHINE LEARNING
FOR ASTROPHYSICS
CATANIA, 30 MAY - 1 JUNE, 2022

Contribution ID: **95**   Type: **Poster Presentation**

Time series anomaly detection in a cataclysmic variable using Support Vector Machine

Tuesday, 31 May 2022 14:36 (3 minutes)

Cataclysmic variable stars (CVs) often show multiple frequency components with a quasi-periodic occurrence. Those can be very subtle and their confidence using standard statistical methods is often of a less significance, e.g., falls under the 1-σ interval. In our study we aim to use a Support Vector Machine (SVM) to train a model to detect those components with a plausible confidence. We used the lightcurve of MV Lyrae, which is a very bright member of the CVs family. Those are known to have a specific pattern in the variability of their brightness. We used a 272 day long light curve obtained at a cadence of $\approx$ 59 s from the Kepler satellite archive. Using Lomb-Scargle (1982) algorithm we created periodograms and averaged all its values within the bins to get the power density spectrum (PDS). The searched characteristic frequency component is located near log(f/Hz) $\approx$ -3.4. We simulated the data using the Timmer & König (1995) method based on the PDS with the frequency component. The quasi-periodic frequency in the simulated data had a low confidence level (1-σ) which was based on the same confidence level as in the observed PDS. The simulated data had a quasi-periodic peak near log(f/Hz) $\approx$ -3.4. The dataset was divided in two categories, with the presence of the quasi-periodic oscillation and without it (removing it from the simulated PDS with the oscillation). These data, as a time series, were used to train a supervised SVM model. We used a Gaussian kernel to find the support vectors and a hyperplane. Optimal parameters of the regularization parameter C as well as the kernel coefficient (γ) were found using a cross-validated grid-search. We trained several models using different sizes of the training dataset. The model was tested using simulated data as well as the PDS from the observation of our studied CV. The classification accuracy reflects the confidence of the manifesting quasi-periodic frequency.

Main Topic

Supervised/Unsupervised/Semi-supervised Learning

Secondary Topic

Time series analysis, transients

Participation mode

In person

Primary author: BENKA, Denis (Advanced Technologies Research Institute)

Co-authors: Dr DOBROTKA, Andrej (Advanced Technologies Research Insitute); Dr STRÉMY, Maximilián (Advanced Technologies Research Institute)

Presenter: BENKA, Denis (Advanced Technologies Research Institute)

Session Classification: Poster Session Day 2