Machine Learning Autoclassification of Candidate Variable Stars in Python

Travis Stenborg

School of Computer Science, University of Sydney

SuperWASP Candidate Variables

Identification and classification of variable stars provides valuable data for understanding stellar population composition, structure and evolution. Lightcurve inspection allows detection of intrinsic pulsators, rotationally modulated variables, eclipsing binaries and other exotica. The volume of relevant lightcurve data pending analysis is considerable, so much so that some associated crowdsourced citizen science analysis efforts are projected to extend years. The efficacy of a Python-based machine learning system using multinomial logistic regression, for automating or complementing such efforts, was examined here.

Lightcurve data from the SuperWASP survey was used as a test case. Though the SuperWASP robotic telescopes were commissioned to find transiting exoplanets, the lightcurves of monitored sources also yield candidate variable stars.

Project Data Set

The machine learning system developed is a supervised learning system. As such, it uses a set of labelled lightcurves. 100 human expert evaluated lightcurves and corresponding classifications were drawn from the SuperWASP Variable Stars project on the Zooniverse online platform [1].

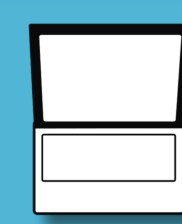
Aside from variable star classifications drawn trivially from the platform's tutorial and field guide sections, many (~60%) class exemplar and edge cases were distilled from online posts published by the Zooniverse SuperWASP project lead.

Multinomial Logistic Regression

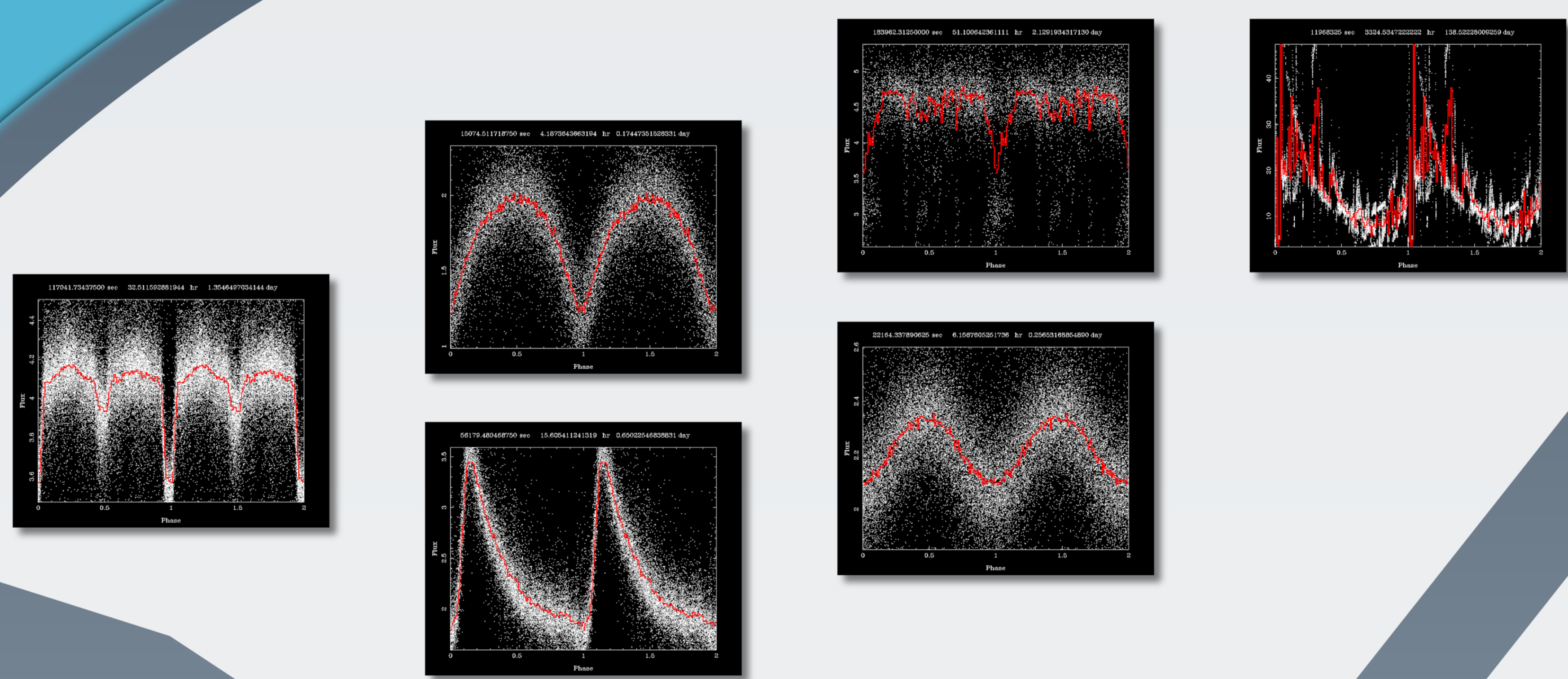
Image classification was automated here via multinomial logistic regression. Six classes were used: pulsator, rotator, EA/EB type, EW type, unknown and junk. The classifier was implemented directly in Python, not executed via calls to an external package such as sklearn.svm.

The dataset was randomly divided into (80/20)% training/test sets. The input 850×680 pixel GIF images were preprocessed to remove lightcurve titles and axes. A weights matrix was established of flattened pixels and variable star categories. Image model training via gradient descent matrix refinement with a softmax function, maximum likelihood predictor and cross-entropy loss was performed.

A regularisation parameter was used to prevent overfitting, and a stop condition applied to constrain infinitesimal model training. A learning rate hyperparameter was tuned to produce optimal testing accuracy.

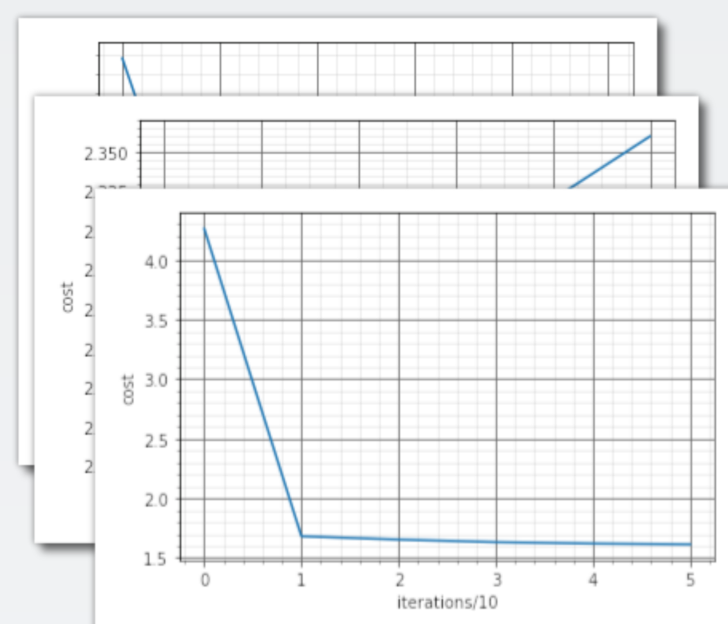


SuperWASP Lightcurves



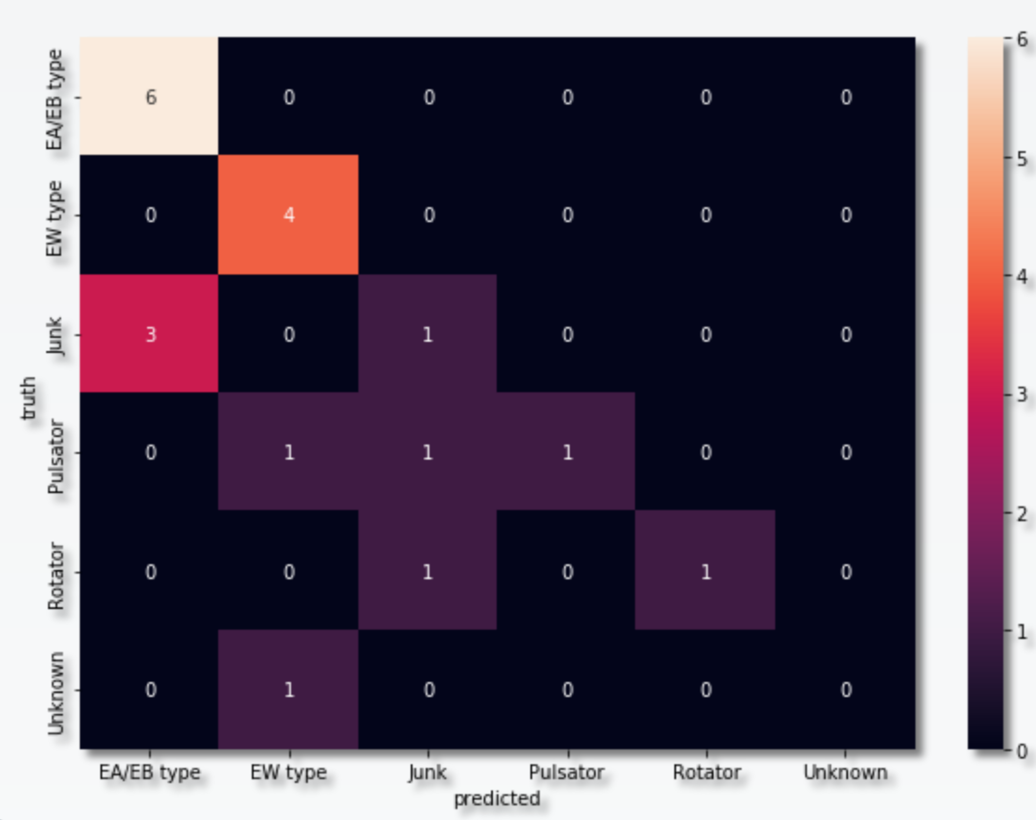
Machine Learning

The multinomial logistic classifier employed test-retest reliability, with model accuracies averaged from three machine learning runs over randomly sampled training and test sets.



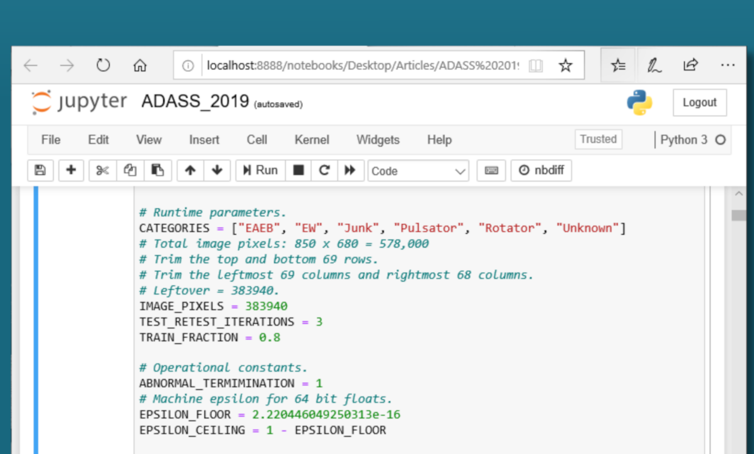
Candidate Autoclassification

The variable star candidate autoclassifier achieved average training and test set accuracies of 99% and 65% respectively. A corresponding confusion matrix is shown below right. For the 100-image input set, classification of EA/EB and EW types was perfect (less so other classes).



Python

The machine learning system was implemented in Python 3.7.3. A browser-based Jupyter notebook, running under an Anaconda Python distribution, was used as an integrated development environment. Python's status as a de facto standard language for machine learning systems will help make the system extensible and maintainable by interested third parties.



Results and Discussion

Reaching an average test accuracy of 65%, with some classes (EA/EB and EW types) receiving perfect classifications was deemed validation of proof-of-concept.

Though a significant improvement over a naive random classification case (17% accuracy), the autoclassifier isn't competitive with human evaluation. The key limitation here however isn't the autoclassifier's underlying techniques, but the training dataset. An input dataset of 100 images is very modest for a complex machine learning task. Compounding this was the fact that much of the input data set consisted of lightcurve edge cases, making it harder for the classifier to assess typical lightcurves. Repeating this project with more training data is recommended.

THE UNIVERSITY OF SYDNEY

References

[1] Norton, A. J. (2018). A Zooniverse Project to Classify Periodic Variable Stars from SuperWASP. Research Notes of the AAS, 2(4), article no. 216.

This poster uses data and images generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation.