Bachelor Project Machine Learning for the European Galactic Plane Surveys.

Main supervisor: Paul Groot
Daily supervisor: Jan van Roestel (Sally McFarlane)

The project will consist of 2 main problems. The first is to program and test methods to automatically select blue sources from UVEX data. Once the student is done with this and the best method is found, it can be applied to VPHAS+ data. The second is to use the VPHAS+ colors and OmegaWhite lightcurves, where the student will then use either supervised or unsupervised machine learning classifiers to classify the different types of objects.

The European Galactic Plane Surveys (EGAPS) are mapping a band of 5x360 degrees in the sky in multiple optical colours (u,g,r,i,Hα). OmegaWhite is a survey to obtain lightcurves on a 5-minute cadence (for a duration of 2 hours) over >100 square degrees to identify short-period compact binaries. In area it overlaps with the southern part of EGAPS, VPHAS+

To start with, the student will read up on the literature about the European Galactic Plane Surveys, (UVEX, IPHAS, and VPHAS+). The best place to start are the websites of the surveys to get some general background information. Once the student has a general idea about what data is available, he should dive in to some refereed papers about science done with EGAPS. A great place to start is the PhD thesis by Kars Verbeek, who also did some color selection of sources in early UVEX data.

Once the student has a good graps of the problem, he will continue with reading literature about Machine Learning and Data Mining methods. A number of good introductory books are available, some of which are listed below. Once the general concepts are understood, reading should focus more on methods of detecting outliers. To find out what methods are available, consultation with the Machine Learning department at the Radboud University might be possible.

When a number of potential methods have been selected, they need to be programmed and tested. A number of UVEX datasets will be available. To determine the efficiency of the methods, a training set needs to be made, so the number of false positives and false negatives can be determined. The student will test a few methods and select the method which is deemed best, given the test statistics calculated. This applies for both projects mentioned above: the outlier detection in UVEX and VPHAS+ data, as well as the combination of light curves and colours for source identification in the combined data sets of OmegaWhite and VPHAS+.

The best methods will then be applied to VPHAS+ data. For the selected candidates, for which lightcurve statistics are available, a supervised and/or unsupervised machine learning classifier will be used to group the candidates in their respective classes. The final result is a catalogue of variable , blue sources, which are classified using a computer. This catalogue can then be used by others to selected interesting targets for followup.

At the end the student will write a report about the whole project, reporting the findings about the outlier selection methods and the classification procedure.

Time schedule:

- ~14 days reading about UVEX, IPHAS, VPHAS+ and Machine learning
- ~21 days working on color-magnitude outliers

- ~21 days working on classification of VHPAS+ sources using colors and lightcurves.
- ~14 days writing the report

Literature

- White dwarfs in the European Galactic Plane Surveys (EGAPS), Morales-Rueda et al. 2006, http://arxiv.org/abs/astro-ph/0610701
- The survey websites:
  - IPHAS: http://www.iphas.org/
  - UVEX: https://www.astro.ru.nl/uvex/Uvex_home.html
  - VPHAS+: http://www.vphasplus.org/
- A first catalogue of automatically selected UV-excess sources from the UVEX survey, Verbeek 2012, http://arxiv.org/abs/1111.0440.
- Some books about machine learning and datamining, useful for some background information:
  - “Statistics, data mining, and Machine learning in Astronomy”. Željko Ivezić, Andrew J. Connolly, Jacob T. VanderPlas & Alexander Gray.
  - “Modern Statistical Methods for Astronomy: With R Applications”. Eric D. Feigelson and G. Jogesh Babu
  - "Automated Classification of variable stars: Application to the OGLE and CoRoT databases”, PhD Thesis by Jonas Debosscher